chapter 4

How Should Fluency Be Assessed?

Since fluency is one critical aspect of proficient reading, it is important that schools have procedures in place to track fluency development. Fluency assessment, particularly the monitoring of development in beginning readers, allows instruction to be modified when fluency difficulties first appear.
Far too many readers in the upper grades struggle with fluency simply because little attention was paid to their fluency problems in the early grades. Too many older readers struggle with fluency and with reading generally because early reading instruction involved too many of the features discussed in Chapter 3. As illustrated so powerfully by Torgeson and Hudson (2006) in their review of prevention and intervention studies, early attention to fluency difficulties produces far more positive outcomes than do interventions that begin after fluency problems have been around for a while.

When readers fail to develop fluent reading abilities early on, they struggle with developing other reading skills and abilities year after year. By the time they complete third grade, their achieving peers have read millions more words, and the word reading deficit undermines all efforts to develop normal reading abilities. Therefore, it is important to monitor fluency early and to continue to monitor the fluency of any reader who is exhibiting fluency difficulties. It is also important to monitor the volume of reading that struggling readers do. The amount of teacher-directed reading may need to be increased for these students, especially if they are either unable or unwilling to read independently.

At this point I must note, however, that assessing and monitoring fluency will not solve fluency difficulties. Once fluency assessments have identified the readers having difficulty, appropriate instructional interventions are essential if fluency problems are to be overcome. This may involve altering the nature of classroom reading lessons, providing powerful reading interventions including special fluency instruction, or both. But, as I noted earlier, far too many fluency difficulties arise as a result of the particular characteristics of many classroom reading lessons. If monitoring the development of fluency in primary-grade classrooms leads to appropriate changes in classroom reading lessons, then fewer students will struggle with fluency and with learning to read.

Fluency: The Great Debate

Fluency has been hijacked (Mathson, Solic, & Allington, 2006). In many schools across the nation, reading speed and accuracy is being regularly assessed but fluency development is not. As noted in Chapter 1, fluency is
the ability to read in phrases with expression and comprehension. This was the meaning as developed in automaticity theory (LaBerge & Samuels, 1974), the theoretical foundation for the importance of fluency in reading development. Nonetheless, the most popular assessment tool that schools use to evaluate and monitor fluency development is the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2002).

Why DIBELS Is Not a Fluency Assessment

The DIBELS assessments (there are several subtests) do assess rate and accuracy of student responses on measures of various subskills. Each of those DIBELS subtests incorporates the word fluency in the title—for instance, “Letter Naming Fluency” and “Non-Word Reading Fluency.” In all cases, however, the fluency tag is simply not valid (Samuels, 2006). Speed of response is something different from fluency. Accuracy of response is something different from fluency. Certainly the DIBELS tests measure something, but it is not fluency that is being measured. Thus, the DIBELS assessment cannot be considered a reliable and valid assessment of fluency as required under the No Child Left Behind Act. (I will note that there are other similar systems, though less popular, that have most of the same limitations—the AIMSweb assessments, for instance.)

As NRP member Samuels (2007) argues: “One criticism I have of the DIBELS tests is that, despite their labels, they are not valid tests of the construct of fluency as it is widely understood and defined. They only assess accuracy and speed. . . . By attaching the term fluency to their tests, they create the false assumption that that is what their tests measure.” Samuels is only one of the researchers who have criticized DIBELS for various inadequacies.

Michael Pressley, the former editor of the Journal of Educational Psychology, the premier experimental research journal, and his colleagues (Pressley, Hilden, & Shankland, 2005) conducted a multifaceted study of the adequacy of the DIBELS Oral Reading Fluency and Retelling Fluency subtests. They concluded their research report with the following: “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” (p. 1). The authors noted that one potential danger of the DIBELS was that
readers who can read accurately can meet DIBELS rate, accuracy, and comprehension benchmarks for progress even though they may not comprehend much of what they have read. That is because the DIBELS retelling test did not reliably estimate understanding of the material that had been read. Students could obtain a high DIBELS retelling score without recalling much of the information in the passages they had read. And although the DIBELS oral reading scores did correlate with standardized reading achievement test scores, the correlation was weak to modest. In other words, reading fast and accurately did not predict reading proficiency as typically assessed with any high level of accuracy. Markel and Deno (1997) and Altwerger, Jordan, and Shelton (2008) report similar problems with using rate and accuracy assessments to predict reading comprehension.

Likewise, the research of Schilling, Carlisle, Scott, and Zeng (2007) indicated that although DIBELS benchmark scores were reasonably accurate (75%) in identifying readers below the 25th percentile, over one-third of the students who met or exceeded the DIBELS oral reading fluency benchmarks in second and third grade failed to achieve normal reading levels on the state reading assessments. These students had been denied added instructional support based on the “adequacy” of their DIBELS performances. The DIBELS assessments indicated they were achieving readers, but the reading achievement of these students, their comprehension especially, was actually lagging behind. Such unreliable assessments prove extremely costly in states and school districts where failing to achieve the state reading standard leads to retention in grade.

In a similar vein, Samuels (2006) rejected the claims that DIBELS did a good job of predicting reading achievement as measured on traditional standardized assessments. He noted, for instance, that the studies reporting a strong relationship between DIBELS oral reading rate and accuracy scores and reading comprehension were fundamentally flawed—flawed because the estimates of comprehension were not done on the DIBELS materials that the students read but on other tests given at a later time. Thus, on the DIBELS assessment, students could read fast and not worry about comprehension. However, with no reliable test of comprehension after the fast DIBELS reading, there are no data to indicate whether fast reading is linked to better understanding.
The studies reviewed by Walczyk and Griffith-Ross (2007) suggest otherwise. They report that the first strategic adaptation that good readers use when they confront comprehension difficulties is to slow down their rate of reading. The next most common adaptation is to reread the confusing sentence. If readers do either of these while taking the DIBELS tests, their scores will plummet. In other words, DIBELS scoring guidelines penalize readers for engaging in the very fix-up strategies that good readers use.

Finally, in an analysis of research that appears on the DIBELS website, Wilde (2006) noted that although all the numbers look impressive, there are real problems for making instructional decisions with DIBELS data:

If this DIBELS test were used to decide which students should receive extra instruction, schools would have to determine whether to help just the “high-risk” students or the “some-risk” ones as well. There were 361 students who didn’t meet the state benchmark; only half of them (178) would have gotten extra help if it only went to those in the “high-risk” category. If help were given to the “high-risk” and “some-risk” categories, it would catch most of those likely to not meet benchmark (308), but extra help would also be given to 230 students who didn’t need it. (p. 67)

Independent researchers examining the reliability of DIBELS for making judgments about reading development have demonstrated the inadequacy of this test. What DIBELS does do reasonably well is identify the readers making very little progress toward reading fluency. As for higher-achieving readers, DIBELS misidentifies far too many children as making adequate progress who are not and identifies too many children as at-risk when they are not.

Mathson, Solis, and Allington (2006) reported a small study of the reliability of DIBELS with third-grade readers. The data in Table 4.1 are drawn from that study and illustrate how widely variable performances on the DIBELS Oral Reading Fluency subtest can be. These data illustrate the DIBELS words correct per minute (wcpm) scores of 10 students in the analysis. Each student read three DIBELS passages aloud on the same day, passages of equivalent difficulty according to the DIBELS manual. Examine the “variability” column that shows the difference between the fastest and slowest reading each student did on the day they were tested.
Altwerger, Jordan, and Shelton (2008) report findings almost identical to our study and they further found that standardized assessments of word reading did not predict DIBELS fluency scores well. The fastest readers on the DIBELS ORF assessment had word recognition scores between the 17th and 99th percentiles on the Woodcock-Johnson reading assessment. How can any test that produces this much variation in performance be considered reliable? It can only if you discard two-thirds of the data, which is precisely what the DIBELS manual directs teachers to do. In other words, one way to attempt to make the DIBELS tests look more reliable than they are involves tossing out the fastest and slowest performances. But which of the three rates is really the right one—the one that best portrays students’ typical speed and accuracy performances?

The authors (Mathson, Solic, & Allington, 2006) also computed the correlation between the three DIBELS scores each student achieved. Basically, these readers fell into three distinct groups as a result of this analysis. For one group ($n = 11$) the reliability of the DIBELS reading rate and accuracy was

<table>
<thead>
<tr>
<th>Student #</th>
<th>Variability in wcpm</th>
</tr>
</thead>
<tbody>
<tr>
<td>S4</td>
<td>47</td>
</tr>
<tr>
<td>S13</td>
<td>28</td>
</tr>
<tr>
<td>S14</td>
<td>32</td>
</tr>
<tr>
<td>S25</td>
<td>36</td>
</tr>
<tr>
<td>S30</td>
<td>49</td>
</tr>
<tr>
<td>S31</td>
<td>56</td>
</tr>
<tr>
<td>S32</td>
<td>33</td>
</tr>
<tr>
<td>S34</td>
<td>33</td>
</tr>
<tr>
<td>S35</td>
<td>39</td>
</tr>
<tr>
<td>S38</td>
<td>29</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Passages numbered</th>
<th>Student #</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S4</td>
<td>118</td>
<td>81</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>S13</td>
<td>66</td>
<td>38</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>S14</td>
<td>92</td>
<td>60</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>S25</td>
<td>90</td>
<td>73</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>S30</td>
<td>94</td>
<td>79</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>S31</td>
<td>165</td>
<td>128</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td>S32</td>
<td>97</td>
<td>64</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>S34</td>
<td>80</td>
<td>66</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>S35</td>
<td>74</td>
<td>44</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>S38</td>
<td>81</td>
<td>52</td>
<td>58</td>
</tr>
</tbody>
</table>

Table 4.1 Variability in DIBELS Oral Reading Fluency performances
below $r = .25$. This means that their performances varied widely from passage to passage. Another group ($n = 16$) had scores that had correlations between .25 and .50. For these readers, DIBELS reading rate and accuracy again varied quite a bit. The third group ($n = 12$) had correlations above .50, still not very high but higher than the other groups.

What should you make of this analysis? The authors’ interpretation is that DIBELS does not measure reading rate and accuracy with any reliability. The variation from one passage to another is simply so large that even using just the middle range score is likely to misinform teachers about students’ development.

Now all of this may seem a bit picky and a lot technical. But given the time, cost, and energy it takes to administer the DIBELS assessments regularly, I think we should expect a better test. If the DIBELS developers and promoters could provide evidence of what has been dubbed “consequential validity,” then I would be willing to modify my stance. But there are no such data. Basically, a study of the consequential validity of DIBELS would require a demonstration that when schools use the DIBELS assessments, instruction improves as well as achievement. But to identify DIBELS as the reason scores improved would require a randomized experiment where a number of schools used the same core reading program, and in some of those schools, but not all, DIBELS was randomly assigned to also be used to make instructional decisions. If the reading achievement of the schools using DIBELS then improved more than the scores in the schools using the same core program but not using DIBELS, then I might be satisfied that DIBELS is not simply a time-consuming fraud.

I won’t hold my breath until the developers conduct such a study, but I will continue to point out that there is no evidence that using DIBELS produces better teaching or learning.

**Why Is DIBELS Used in So Many Schools?**

How is it that DIBELS has become the most widely used assessment in schools intending to monitor fluency development? The short answer to this question is: powerful political connections and a hint of entrepreneurial corruption. As noted by the Office of the Inspector General (2006) of the U.S. Department of Education and in the Congressional hearings on the Reading Fluency: The Great Debate.

Fluency: The Great Debate
First component of the NCLB (Glenn, 2007), various federally linked consultants promoted the use of DIBELS to state education agencies, and a contested review of the adequacy of various reading assessments, including DIBELS, was purposely positioned on the Reading First website to make it appear it had some official endorsement by the USDE. It did not.

In the end, the majority of states included DIBELS as the primary tool for monitoring fluency development and reading progress. Officials from several states testified that they were told by the then director of the federal Reading First office that the state’s Reading First application would not be approved unless DIBELS was the preferred assessment tool (Brownstein & Hicks, 2006). After the state education agency identified DIBELS as the preferred monitoring assessment, school districts participating in the Reading First program then were required to use DIBELS. Thus, DIBELS became a commercial success not because of its unique technical qualities but because of entrepreneurial manipulation. The millions of dollars in earnings that DIBELS delivers have the developers and publishers smiling all the way to the bank.

Before moving on, I must also note that DIBELS is now being used in many schools that do not participate in the Reading First program. These schools elected to use DIBELS. Why? I cannot explain these decisions and will also suggest that given the research available, school personnel would be hard pressed to defend their decisions on scientific grounds. What I think the widespread nonmandated use of DIBELS suggests is that educational faddism is still alive and prospering. Most worrisome is that in too many school systems DIBELS is being used as the progress monitoring tool in response to intervention (RTI) initiatives. However, the research on DIBELS is such that anyone using DIBELS to make important decisions about students’ development of reading proficiencies is guilty of educational malpractice.
If Not DIBELS, What?

The truth of the matter is that if schools want to monitor the rate and accuracy of students’ oral reading, there are several well-known techniques for doing so. In fact, DIBELS was an attempt to use a somewhat similar technique but to generate earnings as well. Both the words correct per minute technique (Deno, 1985) and the running records technique (Johnston, 2000) provide rate and accuracy data just as DIBELS does.

The advantage of both of these well-known techniques, besides the no-cost feature, is that both assess student progress on the curriculum materials they use every day—in other words, they are curriculum-based data. If teachers want information on whether instruction needs to be modified, then data gathered from student responses while reading the texts that are being used for instruction are far more useful than scores from random and specially written passages such as those found in the DIBELS assessment.

For instance, I’ve noted that one critical feature of reading lessons is ensuring that students have books of appropriate difficulty in their hands. Using curriculum-based measures of rate, accuracy, fluency, and comprehension can tell you whether you are using the right texts. Data from random texts cannot provide that critical bit of information.

If, for some reason, you wanted to assess the acquisition of certain decoding skills using a sample of non-words, then using non-words that are constructed of the phonic elements that have been taught will provide better information than a test that includes non-words with elements that have not yet been taught (which is the case with the DIBELS non-word fluency subtest and many other decoding tests).

Like Sandra Wilde (2006), I suppose I could see the DIBELS assessment being used as an initial screening device—as a crude but quick tool to identify which students seem to be in trouble with reading acquisition. But given how crudely it measures subskills, I wonder whether schools (and children) might not be better served by using either of the curriculum-based assessment techniques. In the end, the evidence for the power of curriculum-based measurements to improve instruction is too compelling to ignore. The NRP (2000) agreed and noted that a number of procedures might be used to assess and monitor oral reading fluency (pp. 3–9). These included running records,
words correct per minute calculations, miscue analysis, informal reading inventories, standardized oral reading assessments, and pausing indices. I will discuss the first two procedures next.

Curriculum-Based Measurement of Fluency

Curriculum-based measurement offers a number of advantages that make these sorts of assessments instructionally useful. One key finding is that when schools adopt curriculum-based measurement techniques, both teaching and learning improves (see Deno & Marston, 2006, for a review).

As you listen to readers who struggle with fluency you should be asking yourself this question: Why is this reader having fluency problems? Is it because:

___ He doesn’t recognize many high-frequency words at a glance?
___ She decodes lots of words slowly, letter by letter?
___ He seems to be missing decoding skills appropriate for his reading level?
___ She waits for adult help rather than attempt unknown words independently?
___ He doesn’t know the meanings of many words he is trying to read?
___ She doesn’t pay attention to punctuation?
___ He is not self-monitoring while reading?
___ She is trying to read fast and not attending to understanding?
___ He seems anxious and uncomfortable reading aloud?
___ She is remarkably better in accuracy, rate, and fluency if I allow her to read the text silently before she reads aloud?
Teachers can use these measures to monitor growth and progress of students in the curricular materials students use every day. When gathering such data repeatedly, teachers use the data for individuals and track progress toward peer-referenced or other normative standards. I’ve mentioned two well-known examples, running records and words correct per minute. In the following section each of these techniques will be developed in greater detail. In addition, since neither technique originally evaluated fluency, I provide guidelines for complementing the rate and accuracy data with a rating of fluency.

Running Records

Peter Johnston (2000) has written a small book that details the development of running records and the research on this technique that originated with Marie Clay (1985). Johnston includes detailed guidelines on how data are collected using the running record technique and provides an audiotape for developing proficiency using running records to inform instruction. If you are unfamiliar with the running record technique for recording oral reading performances but plan to use this approach, I’d suggest getting a copy of Johnston’s book even though I outline the procedure here.

It is important to understand that in this section I am not providing the full running record process. Instead, I am recommending that the technique used in recording reading behaviors be implemented in assessing and monitoring oral reading performance. As a matter of fact, I am not even providing all the coding features used in the running records process, because you can use the technique to collect useful evidence even without the full complement of codes.

Using the running record technique to gather data on student performance in the curriculum materials used in the reading instruction requires virtually nothing but a pencil and a clean sheet of lined notebook paper. The advantage of this is that running records can be collected with very little preparation. In fact, if you always have a notebook of lined paper, a pencil, and timer handy when you work with students, there would be no preparation required. Thus, unlike techniques that require that you photocopy the text the student will read to you or require that you locate the packet of assessment materials that will be used, the running record can be done literally spontaneously.
Following the basic procedures for gathering running record data, you would select a text that is used in the daily reading lessons but one the student has not yet read. Typically, you would have the reader begin at the beginning of the selection and read through it. If the selection is long, you might decide to have the student read only some portion of the text.

When the reader begins, you note the time or start a stopwatch. When the reader has finished, you note the time or turn off the stopwatch and record the time spent reading. As the reader reads aloud, you simply make a check mark on the lined sheet of paper for every word pronounced correctly. For words that are mispronounced, you write a dash. On the blank sheet of lined paper, then, you would have a record that looked like this for a five-word sentence where one word was mispronounced: ✔✔✔ — ✔. If the reader skipped over a word, you write an O (for omitted) and the record would look like this: ✔✔ O ✔✔. It is important to remember that if a misread word is self-corrected—read correctly—then you count that word as correctly identified. The feature box below provides the other symbols you will need to create the running record of oral reading behaviors.

After the child has completed the oral reading assessment, you need to take a minute (literally) and calculate accuracy and rate of reading and fluency. To figure the accuracy level, divide the number of words read incorrectly by the total number of words read. Example: 2 incorrect out of 22: 2/22 = 0.09 = 9%

### Commonly Used Running Record Symbols

- **Correct response** ✔
- **Misread word** —
- **Omission** O
- **Insertion** ^
- **Teacher prompt** T

Example:
- ✔✔✔ ✔✔ — ✔
- ✔✔ O ✔✔
- ✔✔ ✔✔ ✔ ✔
- ✔✔ ✔✔ ✔ ✔
- ✔✔ ✔✔ ✔ ✔
or omitted by the total number of words read (for example, 5 words read inaccurately in a 121-word text results in a 4 percent error rate, or a 96 percent accuracy level). To calculate reading rate, simply divide the total time spent reading by the total number of words read (if the reader took 2 minutes and 12 seconds to read a 121-word passage, the reading rate would be 55 words per minute). One advantage of measuring reading in one-minute segments is that it makes calculating reading rate much easier.

Recording the reader’s level of fluency would be the next step (details on fluency ratings appear later in this chapter). The full running record procedure also includes evaluating the nature of the errors readers make to draw inferences about the strategies being used and not being used. But you will need to study Johnston’s book to add that level of analysis to your toolbox.

You can use both oral reading accuracy and rate data to monitor a reader’s progress. Additionally, the accuracy data can be used to determine whether the classroom text that the reader read from is of an appropriate level of difficulty for instructional use.

Words Correct per Minute (wcpm)

Shinn (1989) provides a detailed account of how to gather and use wcpm data to monitor oral reading development, and interested readers should study his text to develop a fuller understanding of the method. The traditional wcpm technique has readers read aloud for one minute from materials used in their reading lessons. Typically, the teacher makes a photocopy of the text to be read and then, as the child reads, simply crosses out any word read incorrectly. When the reading is completed, you simply count the number of correctly read words to calculate wcpm data. This number is recorded and typically entered on a graph that is used to track changes in rate and accuracy of oral reading. By tracking a reader’s wcpm, you can observe whether progress is being made in terms of faster and more accurate reading of instructional texts. As with the running record technique, you can also evaluate the appropriateness of the difficulty of a text for instructional purposes. An example of a sequence of graphed wcpm data is displayed in Figure 4.1.
Figure 4.1  Weekly oral reading wcpm data over a 12-week intervention
Judging the Adequacy of a Student’s Oral Reading Rate

Graphing wcpm or rate and accuracy data is a good start, but how do you interpret such data? If you gather repeated samples of oral reading over time (say, one semester), you will probably be able to see whether improvement is being made. I say “probably” because to assess progress on wcpm, you would need to ensure the texts used in the several samples were of similar difficulty. It may be naïve to think that samples from the same core reading program are of equal difficulty. However, using one-minute oral reading samples means that you could use a single story from later in the reader to gather all of your samples. Alternatively, you could use a single book from the classroom library to gather the growth samples.

If one goal is using oral reading wcpm to evaluate the appropriateness of the text for instructional purposes, then the data should be gathered on those texts. Likewise, if you are interested in evaluating a student’s success in the curriculum materials you are using, you would also gather the wcpm data from the curriculum materials in use.

The best data on average grade-level oral reading wcpm levels can be found in Table 4.2. These data come from a study of over 200,000 students in 23 states. The table displays the average wcpm at three times each year (no data for fall of grade 1 are provided because many students have not begun to read at that point). The average wcpm data can help you see how any particular student in your class compares to most other students at his or her grade level at different times during the school year. But these are “average” wcpm data, with half of the students reading faster and slower than the numbers displayed. If you are interested in information on oral reading rates that fall in the average range (as opposed to exactly average) go to www.prel.org/products/re_/assessingfluency.pdf to find an excellent brief written by Tim Rasinski that includes those data.

Note that wcpm data for the spring of grades 6, 7, and 8 do not vary. In other words, the typical student shows no growth over the three years of middle school! It isn’t that 150 wcpm is the uppermost level that can be achieved, because the top 10 percent of grade 5, as well as grade 6, 7, and 8, students read at the 200 wcpm level or faster. These wcpm data indicate
what lots of other data on middle school reading achievement indicate—at the present time, middle school instruction does not foster much reading growth.

Additional Considerations

I worry that when teachers collect oral reading data cold, those situations where the reader has never seen the text before, teachers might get data that do not truly reflect what the reader would do normally when reading. Oral reading is different from silent reading, if only in the public performance aspect. It has been my experience that some, maybe many, readers experience a bit of anxiety when asked to read aloud, especially in settings where their peers can hear them. But even if anxiety isn’t particularly high, I still think asking readers to read aloud may alter their use of reading strategies.

For instance, researchers know that proficient readers slow down their reading when they encounter difficulty. Proficient readers also reread sections that are confusing. But these powerful and productive fix-up strategies may be suppressed when students read aloud. Likewise, sounding out an unknown word is another useful strategy that proficient readers use but perhaps use less

---

**Table 4.2** Average wcpm oral reading performances by grade level and time of year

<table>
<thead>
<tr>
<th>Grade</th>
<th>Fall</th>
<th>Winter</th>
<th>Spring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 1</td>
<td>23</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>Grade 2</td>
<td>51</td>
<td>72</td>
<td>89</td>
</tr>
<tr>
<td>Grade 3</td>
<td>71</td>
<td>92</td>
<td>107</td>
</tr>
<tr>
<td>Grade 4</td>
<td>94</td>
<td>112</td>
<td>123</td>
</tr>
<tr>
<td>Grade 5</td>
<td>110</td>
<td>127</td>
<td>139</td>
</tr>
<tr>
<td>Grade 6</td>
<td>127</td>
<td>140</td>
<td>150</td>
</tr>
<tr>
<td>Grade 7</td>
<td>128</td>
<td>136</td>
<td>150</td>
</tr>
<tr>
<td>Grade 8</td>
<td>133</td>
<td>146</td>
<td>151</td>
</tr>
</tbody>
</table>

often when reading aloud because it sounds bad and slows down reading. In some readers’ minds, it may be better just to guess and go on reading aloud. Finally, the very presence of you, the teacher, may alter what the reader does. In independent silent reading you are not available nor are you monitoring the reading. So maybe in this case the reader will elect just to skip over an unknown word the first time she or he encounters it (“I think I’ll skip over it because my reading experience tells me that if this is an important word, I’ll likely see it again in the story”). But when the reader is reading aloud to you, the reader may try to get you to support his or her efforts. This is precisely what McGill-Franzen and McDermott (1978) reported. The very presence of the teacher altered the reading strategies the reader used. And once assistance was given, the reader more frequently requested assistance. Why not take advantage of having the teacher right there beside you?

So, ask yourself, “Does this reader’s oral reading seem to reflect what I know about him or her as a reader?” If you have questions about a student’s cold oral reading data, I have two alternatives to suggest.

### Reading Cold versus Reading Previously Read Text

First, it might be useful to collect running records data on curriculum-based passages read cold and again on curriculum materials the reader has had a chance to read silently before reading aloud. It seems to me that if you see a dramatic difference between the two reading samples, you need to reflect on just what you want to learn about the reader’s development.

For instance, as commonly seems to occur, an oral reading of a text after having first read it silently produces a different reading record. Often there are fewer errors, hesitations, and self-corrections, and better fluency when the text has been previously read silently. So what might this tell you about the reader? I think the oral rereading provides the best case scenario. It demonstrates that the reader has successfully employed a variety of strategies to read the text accurately and fluently. What I worry about is that if you collect only oral reading data cold, you might underestimate the use of powerful strategies that the reader normally uses.

And what if the silent prereading produces no observable changes in rate, accuracy, or fluency? In this case it seems that what you see when the reader
reads a text cold is what you would see if we could somehow observe his or her silent reading processes. But when large differences are observed in rate, accuracy, and fluency when the text is read cold or read after silent reading, I think you should consider that the cold oral reading performance is at odds with the typical way the reader reads.

Over the years I've encountered some children who have quite discrepant scores when I compare cold reading versus rereading of a text. Some readers seem to find the cold oral reading of a text intimidating and so rush through it, just to be over with it. Having the opportunity to read the selection silently prior to the oral rendition works to ameliorate the intimidation factor and produces evidence of better reading abilities than one would have expected from the cold oral reading alone.

**Allowing a Running Start Before Gathering Oral Reading Data**

I will suggest another alternative that also seems potentially useful in improving the reliability of cold oral reading records. This involves allowing the reader to read a bit of the text aloud before beginning to collect the oral reading record data. If you are using the core reading program anthology or a leveled book to gather the curriculum-based oral reading data, you simply allow the reader to read the first few paragraphs (or pages) before you begin to gather your one-minute sample.

This modification allows the reader to become familiar with the story and perhaps get into the flow of reading aloud. This alternative does not go against the traditional grain of oral reading assessment procedures but it does provide the reader with a type of support that is unavailable in the typical oral reading assessment.

**Audiotape Recording the Oral Reading Session**

Years ago I compared teachers’ scoring of oral reading behavior with audiotaped recordings of those same reading performances. What I found was that, generally speaking, teachers were typically not very accurate at scoring oral
reading while the reader was reading. This continues to worry me today. So, following Johnston’s (1991) advice, I will suggest that as you learn to use the running records or wcpm techniques, you initially tape record the student’s reading and check on the accuracy of your coding. Do the same with ratings of fluency. Once you are satisfied that you are recording reading behaviors accurately, you can stop taping. You might want to recheck your accuracy every once in awhile (say twice a year), but taping creates more preparation and uses twice as much time as recording oral reading live.

Assessing the Size of the At-a-Glance Vocabulary

One final assessment that may be useful is examining how many words the struggling reader knows at a glance. There are two basic strategies: One is a contextualized assessment done while the student is reading aloud from a text and the other is a test of recognition in isolation. For the contextualized assessment, you simply examine your running record or wcpm data. Note how many high-frequency words were misread or omitted. This analysis lets you see whether virtually all high-frequency words were read accurately.

You might also want to assess the recognition of the high-frequency words in isolation (and maybe even compare the reader’s accuracy in isolation to his or her accuracy in context). To do this you need to assess in a way that does not allow the student unlimited time to examine the word and respond. The simplest technique I’ve found for this is to write the words to be assessed on 3” × 5” cards (or type or print them using a 16- to 20-point font size). The most common source for high-frequency words is the Dolch list of 220 common words (see Appendix B) but there are other lists that you could use. To ensure that this assessment is curriculum-based, check that any word you assess is found in the reading materials being used.

Once you have a list of the words you want to check to see if they are recognized at a glance and have printed those words on cards, you are ready to begin the assessment. Shuffle the word cards and then place a blank 3” × 5” card as the first card. You will use this blank card to “flash” each word to the student. Tell the student you are going to show him a word for just a second and then uncover the word. He should try to say the word as soon as he
knows it. You then proceed through the deck of word cards, one at a time. You simply slide the blank card up and then down, exposing the word on the second card very briefly. You might begin sliding the card up and down more slowly then pick up the speed, thus shortening the exposure. You can simply place any card with a word the student does not pronounce correctly in one pile and the words correctly identified in another. Later you can complete a record form indicating which words seem to be consistently and quickly recognized after a brief exposure and which still present difficulties.

Assessing the recognition of function words in isolation (on flashcards or word lists) makes the task more difficult than assessing recognition in the context of oral reading of a text. An alternative, then, is to identify the high-frequency function words in the passages a reader read during the wcpm assessment. In other words, analyze the accuracy of identification of function words in those passages as an optional or additional assessment. You could simply list the function words and note whether each was correctly identified when reading the wcpm passages aloud. I’ve found that some struggling readers seem to achieve higher recognition scores when reading a passage than when these words are assessed in isolation.

It is important that readers have a large store of words they recognize at a glance. It is also important that the number of at-a-glance words keeps growing if fluent reading is the goal. At every grade level the best readers add more words to their at-a-glance word bank so that by adulthood the vast majority of words a person encounters are read at a glance with little cognitive effort. This is the key to reading fluently with expression and comprehension.

Assessing Comprehension

Remember that the definition of fluency includes the ability to demonstrate an understanding of the text that was read. So how do you best assess student understanding after reading? I suggest that a global comprehension rating of a retelling of the text content is perhaps the most reliable and most efficient strategy for evaluating student comprehension after reading. This is an individual assessment, but because fluency measures are also individual assessments this technique seems appropriate. Retelling assessments involve you asking the student to recall everything she or he remembers about the text
that she or he just read. You then compare the text content to what the student told you and then rate comprehension as good, fair, or poor. Figure 4.2 provides guidelines for each of these ratings.

If you want to make this evaluation more formalized, you can use the scoring strategy that Pressley, Hilden, and Shankland (2005) used. This strategy is similar to the comprehension assessment found in the Durrell Analysis of Reading Difficulties (Durrell, 1955). Here are the guidelines for scoring retellings using an “idea unit” analysis.

Idea unit analysis has been long used by researchers as a measure of comprehension, although it is primarily a measure of recall. The basic notion is to examine the retelling to see how many of the ideas in the passage were included in the retelling. Idea units can be understood as phrases in the passage. So, to develop a retelling scoring guide you would first break the text down into phrases. Next, print the phrases in order as illustrated on page 70. As the student does the retelling, simply put a check mark next to each idea/phrase that is recalled. After the student finishes the retelling, ask, “Is there anything more that you remember?” If more information is now recalled, place check marks next to those phrases. Now count the number of recalled idea units and compare that to the total number. You could convert this into a percentage and use that as your score for the retell.

Although this percentage could be used as the score, it does not exactly represent a “comprehension” score as one normally uses it. That’s because to earn a 100 percent score the reader would have to retell every bit of the passage, including even small details. But in normal reading, people often forget details but recall the gist of texts they’ve read. So I suggest that you use the percentage and graph it and use it to monitor comprehension but do not get worried if the percentage hovers around 60 percent rather than 100 percent.

Figure 4.2  Retelling global comprehension scoring guidelines

| Good:    | Recalls most of the important ideas in the text that was read. |
| Fair:    | Recalls some of the ideas but omits some details as well.       |
| Poor:    | Recalls few ideas from the text.                                |
Evaluating Fluency and Its Development

In this final section I offer a powerful procedure for monitoring oral reading fluency. Fluency evaluations should always be part and parcel of every evaluation of oral reading proficiencies.

There are several fluency rating schemes including a couple I’ve developed (Allington, 2006b; Allington & Brown, 1979), but I think the preferred fluency rating scale is the one used in the fluency evaluations (Daane et al., 2005; Pinnell et al., 1995) as part of the National Assessment of Educational Progress (NAEP). Both research teams used a four-point fluency scale (see Figure 4.3) to portray the status of fluency development among fourth-grade students. In both studies, fluency ratings were strongly linked to reading comprehension performance. They also found that approximately 60 percent of the students achieved Level 3 or 4 on the fluency scale. These students read with 96 percent accuracy on average. The average reading rate was 119 words per minute. Almost all students reading at Levels 1 and 2 on the scale


This passage has been broken down into idea/phrase units for scoring a retelling.

I couldn’t do my homework.
I was thinking about what to say /
to Michael tomorrow.
There had to be a million mean things /
to say to him.
But I was so mad /
I couldn’t think of any.

fell below the basic reading proficiency standard on the NAEP and read more slowly with an average of 94 percent accuracy.

To monitor fluency development, you simply include a numerical rating of fluency every time you gather data using either the running record or wcpm technique. In addition to charting rate and accuracy or wcpm performances of each reader, you also chart the fluency rating (see Figure 4.4). Appendix A includes several forms useful for recording these data. Forms for individuals and a class record are necessary for monitoring.

As you evaluate student oral reading fluency, remember that it is quite common that first-graders are still stuck to the print. Thus, on cold readings of text you should expect a Level 1 or 2 rating. It is typically in second grade when readers begin to get unstuck. But honestly, it is not uncommon for many readers to be in the upper elementary grades before they can read a text cold at Level 3 or 4. Having younger students practice a text (you model fluent reading and they reread the text several times attempting to emulate

---

**Figure 4.3 NAEP four-point fluency scale**

| Level 4: | Reads primarily in larger, meaningful phrase groups. Although some regressions, repetitions, and deviations from the text may be present, those do not appear to detract from the overall structure of the story. Preservation of the author’s syntax is consistent. Some or most of the story is read with expressive interpretation. |
| Level 3: | Reads primarily in three- or four-word phrase groups. Some smaller groupings may be present. However, the majority of phrasing seems appropriate and preserves the syntax of the author. Little or no expressive interpretation is present. |
| Level 2: | Reads primarily in two-word phrases with some three- and four-word groupings. Some word-by-word reading may be present. Word groupings may seem awkward and unrelated to larger context of sentence or passage. |
| Level 1: | Reads primarily word-by-word. Occasionally two- or three-word phrases may occur, but these are infrequent and/or do not preserve meaningful syntax. |

Source: Pinnell et al., 1995.
your performance) until they are achieving Level 3 and 4 ratings is also useful because it helps them see, hear, and feel what fluent reading is like. My point is that it is unreasonable to expect high levels of fluency when reading texts cold until students develop some advanced reading proficiencies.

Finally, it is important that when working with struggling readers you select rate, accuracy, wcpm, and fluency targets that match their reading level rather than their grade level. A critical first step in accelerating reading development is to help the reader exhibit performances that mirror good readers at his or her reading level. In other words, you work to turn the struggling reader into a good reader at the level he or she is currently at before moving the child into more difficult materials and expecting mastery of more complex skills and strategies.

Technology Applications for Recording Running Record and wcpm Data

Wireless Generation (www.wirelessgeneration.com/products.php?prod=mClass:Reading) produces the handheld personal digital assistants (PDAs) that many schools use for gathering and graphing DIBELS data and sells software for digital recording of running record and wcpm data. This reading record software built into the PDA allows you to simply tap words misread,
Monitoring Students’ Reading Development

When it comes to monitoring the development of student reading proficiencies we have two different but related schemes for doing so that have experimental research supporting their use. Ross (2004) compared the quality of reading instruction offered and student achievement for two groups of teachers. One group had been trained in the running record data-gathering scheme and the other group of teachers received no such training. The teachers who used the running records technique provided higher-quality lessons, and student achievement was significantly higher than the achievement of students in the control classroom where the technique was not used. Fuchs, Deno, and Mirkin (1984) report on a similar study with similar outcomes but here the one group of teachers received training in using the wcpm technique for monitoring reading development. Again, both teaching and learning were better in the classrooms of teachers who used that technique.

In both studies some teachers learned a strategy for monitoring student performance in the curriculum materials used during their reading lessons. Both focused teachers’ attention on reading accuracy and reading rate. Just a small amount of training on either technique improved the reading instruction offered and this led to improved reading achievement.

omitted, self-corrected, repeated, and so on. As soon as the student is finished, the software calculates accuracy, wcpm, and self-correction rate and enters the data in both a student record and a class record file. It also updates the graphs showing previous performances. For those teachers who know how to complete a full analysis of the running records data, the software allows the analysis of the errors and then charts that performance as well.

The Wireless Generation software also currently includes an electronic file of over 3,000 leveled books from a dozen or so educational publishers (e.g., National Geographic, Newbridge, Heinemann, Mondo, Wright Group, and others). Teachers can call up any of these texts for scoring a student’s read-aloud. However, none of the publishers of core reading programs currently allows Wireless Generation to display the stories included in their
readers. Thus, the use of this assessment tool is limited to schools that use a leveled books approach rather than a core reading program, since you need to track progress in the curriculum materials students are using. A final limitation of this product is that it confuses wcpm with a fluency assessment. No fluency scale is included on the software.

Other technology-based tools for developing curriculum-based measurement graphs and class profiles of oral reading performance data are available at no cost at www.interventioncentral.com.

Summary

Gathering data on student’s oral reading performances is one powerful way to monitor progress and to access data that are useful for adapting reading lessons. Perhaps because there has been a widespread mislabeling of reading rate and accuracy assessments as “fluency” assessments, much of the research and development has been conducted on gathering and using rate and accuracy records to monitor progress and inform instruction. Both the running records technique and the wcpm procedure provide oral reading rate and accuracy data but neither have traditionally provided data on fluency.

Techniques for gathering rate, accuracy, and at-a-glance word data have been detailed in this chapter along with a simple four-point fluency scale used in the National Assessment of Educational Progress. Although other fluency scales are available, the NAEP fluency scale is relatively straightforward and requires little training to use.

Monitoring fluency development is important, but if the assessments find that some readers seem to be in trouble, then instruction must be adapted to address those difficulties. Only when the assessment data produce positive changes in reading instruction can one say that the time spent in assessment was worthwhile. The next chapter discusses how instruction might be adapted and modified when not every student is developing as a proficient reader.