

Behind Test Scores: What Struggling Readers *Really* Need

Sheila W. Valencia and Marsha Riddle Buly

Every year thousands of U.S. students take standardized tests and state reading tests, and every year thousands fail them. With the implementation of the No Child Left Behind legislation (www.ed.gov/nclb/landing.jhtml), which mandates testing all children from grades 3 to 8 every year, these numbers will grow exponentially, and alarming numbers of schools and students will be targeted for “improvement.” Whether you believe this increased focus on testing is good news or bad, if you are an educator, you are undoubtedly concerned about the children who struggle every day with reading and the implications of their test failure.

Although legislators, administrators, parents, and educators have been warned repeatedly not to rely on a single measure to make important instructional decisions (Elmore, 2002; Linn, n.d.; Shepard, 2000), scores from state tests still seem to drive the search for programs and approaches that will help students learn and meet state standards. The popular press, educational publications, teacher workshops, and state and school district policies are filled with attempts to find solutions for poor test performance. For example, some schools have eliminated sustained silent reading in favor of more time for explicit instruction (Edmondson & Shannon, 2002; Riddle Buly & Valencia, 2002), others are buying special programs or mandating specific interventions (Goodnough, 2001; Helfand, 2002), and some states and districts are requiring teachers to have particular instructional emphases

(McNeil, 2000; Paterson, 2000; Riddle Buly & Valencia, 2002). Furthermore, it is common to find teachers spending enormous amounts of time preparing students for these high-stakes tests (Olson, 2001), even though a narrow focus on preparing students for specific tests does not translate into real learning (Klein, Hamilton, McCaffrey, & Stecher, 2000; Linn, 2000). But, if we are really going to help students, we need to understand the underlying reasons for their test failure. Simply knowing which children have failed state tests is a bit like knowing that you have a fever when you are feeling ill but having no idea of the cause or cure. A test score, like a fever, is a symptom that demands more specific analysis of the problem. In this case, what is required is a more in-depth analysis of the strengths and needs of students who fail to meet standards and instructional plans that will meet their needs.

In this article, we draw from the results of an empirical study of students who failed a typical fourth-grade state reading assessment (see Riddle Buly & Valencia, 2002, for a full description of the study). Specifically, we describe the patterns of performance that distinguish different groups of students who failed to meet standards. We also provide suggestions for what classroom teachers need to know and how they might help these children succeed.

Study Context

Our research was conducted in a typical northwestern U.S. school district of 18,000

students located adjacent to the largest urban district in the state. At the time of our study, 43% were students of color and 47% received free or reduced-price lunch. Over the past several years, approximately 50% of students had failed the state fourth-grade reading test that, like many other standards-based state assessments, consisted of several extended narrative and expository reading selections accompanied by a combination of multiple-choice and open-ended comprehension questions. For the purposes of this study, during September of fifth grade we randomly selected 108 students who had scored below standard on the state test given at the end of fourth grade. These 108 students constituted approximately 10% of failing students in the district. None of them was receiving supplemental special education or English as a Second Language (ESL) services. We wanted to understand the “garden variety” (Stanovich, 1988) test failure—those students typically found in the regular classroom who are experiencing reading difficulty but have not been identified as needing special services or intensive interventions. Classroom teachers, not reading specialists or special education

teachers, are solely responsible for the reading instruction of these children and, ultimately, for their achievement.

Data Collection and Assessment Tools

Our approach was to conduct individual reading assessments, working one-on-one with the children for approximately two hours over several days to gather information about their reading abilities. We administered a series of assessments that targeted key components of reading ability identified by experts: word identification, meaning (comprehension and vocabulary), and fluency (rate and expression) (Lipson & Wixson, 2003; National Institute of Child Health and Human Development, 2000; Snow, Burns, & Griffin, 1998). Table 1 presents the measures we used and the areas in which each provided information.

To measure word identification, we used two tests from the 1989 Woodcock-Johnson Psycho-Educational Battery–Revised (WJ–R) that assessed students’ reading of single and

Assessment	Word Identification	Meaning	Fluency
Woodcock-Johnson–Revised			
Letter-word identification	X		
Word attack	X		
Qualitative Reading Inventory–II			
Reading accuracy	X		
Reading acceptability	X		
Rate			X
Expression			X
Comprehension		X	
Peabody Picture Vocabulary Test–Revised			
Vocabulary meaning		X	
State fourth-grade passages			
Reading accuracy	X		
Reading acceptability	X		
Rate			X
Expression			X

multisyllabic words, both real and pseudo-words. We also scored oral reading errors students made on narrative and expository graded passages from the 1995 Qualitative Reading Inventory–II (QRI–II) and from the state test. We calculated total accuracy (percentage of words read correctly) and acceptability (counting only those errors that changed the meaning of the text). Students also responded orally to comprehension questions that accompanied the QRI–II passages, providing a measure of their comprehension that was not confounded by writing ability. To assess receptive vocabulary, we used the 1981 Peabody Picture Vocabulary Test–Revised (PPVT–R), which requires students to listen and point to a picture that corresponds to a word (scores of 85 or higher are judged to be average or above average). As with the comprehension questions, the vocabulary measure does not confound understanding with students’ ability to write responses. Finally, in the area of fluency, we assessed rate of reading and expression (Samuels, 2002). We timed the readings of all passages (i.e., QRI–II and state test selections) to get a reading rate and used a 4-point rubric developed for the Oral Reading Study of the fourth-grade National Assessment of Educational Progress (NAEP) (Pinnell, Pikulski, Wixson, Campbell, Gough, & Beatty, 1995) to assess phrasing and expression (1–2 is judged to be nonfluent; 3–4 is judged to be fluent).

Findings

Scores from all the assessments for each student fell into three statistically distinct and educationally familiar categories: word identification (word reading in isolation and context), meaning (comprehension and vocabulary), and fluency (rate and expression). When we examined the average scores for all 108 students in the sample, students appeared to be substantially below grade level in all three areas. However, when we analyzed the data using a cluster analysis (Aldenderfer & Blashfield, 1984), looking for groups of students who had similar patterns across all three factors, we found six distinct profiles of students who failed the test. Most striking is that the majority of students were not weak in all three areas; they were actually strong in some and weak in others. Table 2 indicates the percentage of students in each group and their relative strength (+) or weakness (–) in word identification, meaning, and fluency.

The Profiles

We illuminate each profile by describing a prototypical student from each cluster (see Figure) and specific suggested instructional targets for each (all names are pseudonyms). Although the instructional strategies we recommend have not been implemented with these particular children, we base our recommendations on

Table 2
Cluster Analysis

Cluster	Sample Percentage	English Language Learner Percentage	Low Socio-economic Status Percentage	Word Identification	Meaning	Fluency
1—Automatic word caller	18	63	89	++	–	++
2—Struggling word caller	15	56	81	–	–	++
3—Word stumblers	17	16	42	–	+	–
4—Slow comprehenders	24	19	54	+	++	–
5—Slow word caller	17	56	67	+	–	–
6—Disabled readers	9	20	80	--	--	--

Prototypical Students From Each Cluster

Cluster 1—Automatic Word Callers (18%)

Word Identification	Meaning	Fluency
++	-	++

Tomas

Word identification = ninth grade (WJ-R)
 > fourth grade (QRI-II)
 = 98% (state passages)
 Comprehension = second/third grade
 Vocabulary = 108
 Expression = 3
 Rate = 155 words per minute
 Writing = proficient

Cluster 2—Struggling Word Callers (15%)

Word Identification	Meaning	Fluency
-	-	++

Makara

Word identification = fourth grade (WJ-R)
 < second grade (QRI-II)
 = 75% (state passages)
 Comprehension = < second grade
 Vocabulary = 58
 Expression = 2.5
 Rate = 117 words per minute
 Writing = below proficient

Cluster 3—Word Stumblers (17%)

Word Identification	Meaning	Fluency
-	+	-

Sandy

Word identification = second grade (WJ-R)
 = second-grade accuracy/third-grade
 acceptability (QRI-II)
 = 80% accuracy/99% acceptability
 (state passages)
 Comprehension = fourth grade
 Vocabulary = 135
 Expression = 1.5
 Rate = 77 words per minute
 Writing = proficient

Cluster 4—Slow Comprehenders (24%)

Word Identification	Meaning	Fluency
+	++	-

Martin

Word identification = sixth grade (WJ-R)
 > fourth grade (QRI-II)
 = 100% (state passages)
 Comprehension = > fourth grade
 Vocabulary = 103
 Expression = 2.5
 Rate = 61 words per minute
 Writing = proficient

Cluster 5—Slow Word Callers (17%)

Word Identification	Meaning	Fluency
+	-	-

Andrew

Word identification = seventh grade (WJ-R)
 > fourth grade (QRI-II)
 = 98% (state passages)
 Comprehension = second grade
 Vocabulary = 74
 Expression = 1.5
 Rate = 62 words per minute
 Writing = not proficient

Cluster 6—Disabled Readers (9%)

Word Identification	Meaning	Fluency
--	--	--

Jesse

Word identification = first grade (WJ-R)
 < first grade (QRI-II)
 < 50% (state passages)
 Comprehension = < first grade
 Vocabulary = 105
 Writing = not proficient

our review of research-based practices (e.g., Allington, 2001; Allington & Johnston, 2001; Lipson & Wixson, 2003; National Institute of Child Health and Human Development, 2000), our interpretation of the profiles, and our experiences teaching struggling readers. We conclude with several general implications for school and classroom instruction.

Cluster 1—Automatic Word Callers

We call these students Automatic Word Callers because they can decode words quickly and accurately, but they fail to read for meaning. The majority of students in this cluster qualify for free or reduced-price lunch, and they are English-language learners who no longer receive special support. Tomas is a typical student in this cluster.

Tomas has excellent word identification skills. He scored at ninth-grade level when reading real words and pseudowords (i.e., phonetically regular nonsense words such as *fof*) on the WJ-R tests, and at the independent level for word identification on the QRI-II and state fourth-grade passages. However, when asked about what he read, Tomas had difficulty, placing his comprehension at the second-grade level. Although Tomas's first language is not English, his score of 108 on the PPVT-R suggests that his comprehension difficulties are more complex than individual word meanings. Tomas's "proficient" score on the state writing assessment also suggests that his difficulty is in understanding rather than in writing answers to comprehension questions. This student's rate of reading, which was quite high compared with rates of fourth-grade students on the Oral Reading Study of NAEP (Pinnell et al., 1995) and other research (Harris & Sipay, 1990), suggests that his decoding is automatic and unlikely to be contributing to his comprehension difficulty. His score in expression is also consistent with students who were rated as "fluent" according to the NAEP rubric, although this seems unusual for a student who is demonstrating difficulty with comprehension.

The evidence suggests that Tomas needs additional instruction in comprehension and

most likely would benefit from explicit instruction, teacher modeling, and think-alouds of key reading strategies (e.g., summarizing, self-monitoring, creating visual representations, evaluating), using a variety of types of material at the fourth- or fifth-grade level (Block & Pressley, 2002; Duke & Pearson, 2002). His comprehension performance on the QRI-II suggests that his literal comprehension is quite strong but that he has difficulty with more inferential and critical aspects of understanding. Although Tomas has strong scores in the fluency category, both in expression and rate, he may be reading too fast to attend to meaning, especially deeper meaning of the ideas in the text. Tomas's teacher should help him understand that the purpose for reading is to understand and that rate varies depending on the type of text and the purpose for reading. Then, the teacher should suggest that he slow down to focus on meaning. Self-monitoring strategies would also help Tomas check for understanding and encourage him to think about the ideas while he is reading. These and other such strategies may help him learn to adjust his rate to meet the demands of the text.

Tomas would also likely benefit from additional support in acquiring academic language, which takes many years for English-language learners to develop (Cummins, 1991). Reading activities such as building background; developing understanding of new words, concepts, and figurative language in his "to-be-read" texts; and acquiring familiarity with genre structures found in longer, more complex texts like those found at fourth grade and above would provide important opportunities for his language and conceptual development (Antunez, 2002; Hiebert, Pearson, Taylor, Richardson, & Paris, 1998). Classroom read-alouds and discussions as well as lots of additional independent reading would also help Tomas in building language and attention to understanding.

Cluster 2—Struggling Word Callers

The students in this cluster not only struggle with meaning, like the Automatic Word Callers in Cluster 1, but they also struggle

with word identification. Makara, a student from Cambodia, is one of these students. Like Tomas, Makara struggled with comprehension. But unlike Tomas, he had substantial difficulty applying word identification skills when reading connected text (QRI-II and state passages), even though his reading of isolated words on the WJ-R was at a fourth-grade level. Such word identification difficulties would likely contribute to comprehension problems. However, Makara's performance on the PPVT-R, which placed him below the 1st percentile compared with other students his age, and his poor performance on the state writing assessment suggest that language may contribute to his comprehension difficulties as well—not surprising for a student acquiring a second language. These language-related results need to be viewed with caution, however, because the version of the PPVT-R available for use in this study may underestimate the language abilities of students from culturally and linguistically diverse backgrounds, and written language takes longer than oral language to develop. Despite difficulty with meaning, Makara read quickly—117 words per minute. At first glance, this may seem unusual given his difficulty with both decoding and comprehension. Closer investigation of his performance, however, revealed that Makara read words quickly whether he was reading them correctly or incorrectly and didn't stop to monitor or self-correct. In addition, although Makara was fast, his expression and phrasing were uneven and consistent with comprehension difficulties.

Makara likely needs instruction and practice in oral and written language, as well as in constructing meaning in reading and writing, self-monitoring, and decoding while reading connected text. All this needs to be done in rich, meaningful contexts, taking into account his background knowledge and interests. Like Tomas, Makara would benefit from teacher or peer read-alouds, lots of experience with independent reading at his level, small-group instruction, and the kinds of activities aimed at building academic language that we described earlier, as well as a more foundational emphasis

on word meanings. Makara also needs instruction in self-monitoring and fix-up strategies to improve his comprehension and awareness of reading for understanding. Decoding instruction is also important for him, although his teacher would need to gather more information using tools such as miscue analysis or tests of decoding to determine his specific decoding needs and how they interact with his knowledge of word meanings. Makara clearly cannot be instructed in fourth-grade material; most likely, his teacher would need to begin with second-grade material that is familiar and interesting to him and a good deal of interactive background building. At the same time, however, Makara needs exposure to the content and vocabulary of grade-level texts through activities such as teacher read-alouds, tapes, and partner reading so that his conceptual understanding continues to grow.

Cluster 3—Word Stumblers

Students in this cluster have substantial difficulty with word identification, but they still have surprisingly strong comprehension. How does that happen? Sandy, a native English speaker from a middle class home, is a good example of this type of student. Sandy stumbled on so many words initially that it seemed unlikely that she would comprehend what she had read, yet she did. Her word identification scores were at second-grade level, and she read the state fourth-grade passages at frustration level. However, a clue to her strong comprehension is evident from the difference between her immediate word recognition accuracy score and her acceptability score, which takes into account self-corrections or errors that do not change the meaning. In other words, Sandy was so focused on reading for meaning that she spontaneously self-corrected many of her decoding miscues or substituted words that preserved the meaning. She attempted to read every word in the reading selections, working until she could figure out some part of each word and then using context clues to help her get the entire word. She seemed to over-rely on context because her decoding skills were so

weak (Stanovich, 1994). Remarkably, she was eventually able to read the words on the state fourth-grade reading passages at an independent level. But, as we might predict, Sandy's rate was very slow, and her initial attempts to read were choppy and lacked flow—she spent an enormous amount of time self-correcting and rereading. After she finally self-corrected or figured out unknown words, however, Sandy reread phrases with good expression and flow to fit with the meaning. Although Sandy's overall fluency score was low, her primary difficulty does not appear in the area of either rate or expression; rather, her low performance in fluency seems to be a result of her difficulty with decoding.

With such a strong quest for meaning, Sandy was able to comprehend fourth-grade material even when her decoding was at frustration level. No doubt her strong language and vocabulary abilities (i.e., 99th percentile) were assets. As we might predict, Sandy was more than proficient at expressing her ideas when writing about her experiences. She understands that reading and writing should make sense, and she has the self-monitoring strategies, perseverance, and language background to make that happen.

Sandy needs systematic instruction in word identification and opportunities to practice when reading connected text at her reading level. She is clearly beyond the early stages of reading and decoding, but her teacher will need to determine through a more in-depth analysis precisely which decoding skills should be the focus of her instruction. At the same time, Sandy needs supported experiences with texts that will continue to feed and challenge her drive for meaning. For students like Sandy, it is critical not to sacrifice intellectual engagement with text while they are receiving decoding instruction and practice in below-grade-level material. Furthermore, Sandy needs to develop automaticity with word identification, and to do that she would benefit from assisted reading (i.e., reading along with others, monitored reading with a tape, or partner reading) as well as unassisted reading practice (i.e., repeated reading,

reading to younger students) with materials at her instructional level (Kuhn & Stahl, 2000).

Cluster 4—Slow Comprehenders

Almost one fourth of the students in this sample were Slow Comprehenders. Like other students in this cluster, Martin is a native English speaker and a relatively strong decoder, scoring above fourth-grade level on all measures of decoding. His comprehension was at the instructional level on the fourth-grade QRI-II selections, and his vocabulary and writing ability were average for his age. On the surface, this information is puzzling because Martin failed the fourth-grade state test.

Insight about Martin's reading performance comes from several sources. First, Martin was within two points of passing the state assessment, so he doesn't seem to have a serious reading problem. Second, although his reading rate is quite slow and this often interferes with comprehension (Adams, 1990), results of the QRI-II suggest that Martin's comprehension is quite strong, in spite of his slow rate. This is most likely because Martin has good word knowledge and understands that reading should make sense, and neither the QRI-II nor the state test has time limits. His strong score in expression confirms that Martin did, indeed, attend to meaning while reading. Third, a close examination of his reading behaviors while reading words from the WJ-R tests, QRI-II, and state reading selections revealed that he had some difficulty reading multisyllabic words; although, with time, he was able to read enough words to score at grade level or above. It appears that Martin has the decoding skills to attack multisyllabic words, but they are not yet automatic.

The outstanding characteristic of Martin's profile is his extremely slow rate combined with his relatively strong word identification abilities and comprehension. Our work with him suggests that, even if Martin were to get the additional two points needed to pass the state test, he would still have a significant problem with rate and some difficulty with automatic decoding of multisyllabic words, both of

which could hamper his future reading success. Furthermore, with such a lack of automaticity and a slow rate, it is unlikely that Martin enjoys or spends much time reading. As a result, he is likely to fall further and further behind his peers (Stanovich, 1986), especially as he enters middle school where the amount of reading increases dramatically. Martin needs fluency-building activities such as guided repeated oral reading, partner reading, and Readers Theatre (Allington, 2001; Kuhn & Stahl, 2000; Lipson & Wixson, 2003). Given his word identification and comprehension abilities, he most likely could get that practice using fourth-grade material where he will also encounter multisyllabic words. It is important to find reading material that is interesting to Martin and that, initially, can be completed in a relatively short time. Martin needs to develop stamina as well as fluency, and to do that he will need to spend time reading short and extended texts. In addition, Martin might benefit from instruction and practice in strategies for identifying multisyllabic words so that he is more prepared to deal with them automatically while reading.

Cluster 5—Slow Word Callers

The students in this cluster are similar to Tomas, the Automatic Word Caller in Cluster 1. The difference is that Tomas is an automatic, fluent word caller, whereas the students in this cluster are slow. This group is a fairly even mix of English-language learners and native English speakers who have difficulty in comprehension and fluency. Andrew is an example of such a student. He has well-developed decoding skills, scoring at the seventh-grade level when reading words in isolation and at the independent level when reading connected text. Even with such strong decoding abilities, Andrew had difficulty with comprehension. We had to drop down to the second-grade QRI-II passage for Andrew to score at the instructional level for comprehension, and, even at that level, his retelling was minimal. Andrew's score on the PPVT-R, corresponding to first grade (the 4th percentile for his age), adds to the comprehension picture as well. It suggests that Andrew

may be experiencing difficulty with both individual word meanings and text-based understanding when reading paragraphs and longer selections. Like Martin, Andrew's reading rate was substantially below rates expected for fourth-grade students (Harris & Sipay, 1990; Pinnell et al., 1995), averaging 62 words per minute when reading narrative and expository selections. In practical terms, this means he read just one word per second. As we might anticipate from his slow rate and his comprehension difficulty, Andrew did not read with expression or meaningful phrasing.

The relationship between meaning and fluency is unclear in Andrew's case. On the one hand, students who realize they don't understand would be wise to slow down and monitor meaning. On the other hand, Andrew's lack of automaticity and slow rate may interfere with comprehension. To disentangle these factors, his teacher would need to experiment with reading materials about which Andrew has a good deal of background knowledge to eliminate difficulty with individual word meanings and overall comprehension. If his reading rate and expression improve under such conditions, a primary focus for instruction would be meaning. That is, his slow rate of reading and lack of prosody would seem to be a response to lack of understanding rather than contributing to it. In contrast, if Andrew's rate and expression are still low when the material and vocabulary are familiar, instruction should focus on both fluency and meaning. In either case, Andrew would certainly benefit from attention to vocabulary building, both indirect building through extensive independent reading and teacher read-alouds as well as more explicit instruction in word learning strategies and new words he will encounter when reading specific texts (Nagy, 1988; Stahl & Kapinus, 2001).

It is interesting that 50% of the students in this cluster scored at Level 1 on the state test, the lowest level possible. State guidelines characterize these students as lacking prerequisite knowledge and skills that are fundamental for meeting the standard. Given such a definition, a logical assumption would be that these students

lack basic, early reading skills such as decoding. However, as the evidence here suggests, we cannot assume that students who score at the lowest level on the test need decoding instruction. Andrew, like others in this cluster, needs instruction in meaning and fluency.

Cluster 6—Disabled Readers

We call this group Disabled Readers because they are experiencing severe difficulty in all three areas—word identification, meaning, and fluency. This is the smallest group (9%), yet, ironically, this is the profile that most likely comes to mind when we think of children who fail state reading tests. This group also includes one of the lowest numbers of second-language learners. The most telling characteristic of students in this cluster, like Jesse, is their very limited word identification abilities. Jesse had few decoding skills beyond initial consonants, basic consonant-vowel-consonant patterns (e.g., *hat*, *box*), and high-frequency sight words. However, his knowledge of word meanings was average, like most of the students in this cluster, which suggests that receptive language was not a major problem and that he does not likely have limited learning ability. With decoding ability at the first-grade level and below, it is not surprising that Jesse's comprehension and fluency were also low. He simply could not read enough words at the first-grade level to get any meaning.

As we might anticipate, the majority of students in this cluster were not proficient in writing and scored at the lowest level, Level 1, on the state fourth-grade reading test. It is important to remember, however, that children who were receiving special education intervention were not included in our sample. So, the children in this cluster, like Jesse, are receiving all of their instruction, or the majority of it (some may be getting supplemental help), from their regular classroom teachers.

Jesse clearly needs intensive, systematic word identification instruction targeted at beginning reading along with access to lots of reading material at first-grade level and below. This will be a challenge for Jesse's fifth-grade

teacher. Pedagogically, Jesse needs explicit instruction in basic word identification. Yet few intermediate-grade teachers include this as a part of their instruction, and most do not have an adequate supply of easy materials for instruction or fluency building. In addition, the majority of texts in other subject areas such as social studies and science are written at levels that will be inaccessible to students like Jesse, so alternative materials and strategies will be needed. On the social-emotional front, it will be a challenge to keep Jesse engaged in learning and to provide opportunities for him to succeed in the classroom, even if he is referred for additional reading support. Without that engagement and desire to learn, it is unlikely he will be motivated to put forth the effort it will take for him to make progress. Jesse needs a great deal of support from his regular classroom teacher and from a reading specialist, working together to build a comprehensive instructional program in school and support at home that will help him develop the skill and will to progress.

Conclusions and Implications

Our brief descriptions of the six prototypical children and the instructional focus each one needs is a testimony to individual differences. As we have heard a thousand times before, and as our data support, one-size instruction will not fit all children. The evidence here clearly demonstrates that students fail state reading tests for a variety of reasons and that, if we are to help these students, we will need to provide appropriate instruction to meet their varying needs. For example, placing all struggling students in a phonics or word identification program would be inappropriate for nearly 58% of the students in this sample who had adequate or strong word identification skills. In a similar manner, an instructional approach that did not address fluency and building reading stamina for longer, more complex text or that did not provide sufficient reading material at a range of levels would miss almost 70% of the students who demonstrated difficulty with fluency. In addition to these important cautions about

overgeneralizing students' needs, we believe there are several strategies aimed at assessment, classroom organization and materials, and school structures that could help teachers meet their students' needs.

First and most obvious, teachers need to go beneath the scores on state tests by conducting additional diagnostic assessments that will help them identify students' needs. The data here demonstrate quite clearly that, without more in-depth and individual student assessment, distinctive and instructionally important patterns of students' abilities are masked. We believe that informal reading inventories, oral reading records, and other individually tailored assessments provide useful information about all students. At the same time, we realize that many teachers do not have the time to do complete diagnostic evaluations, such as those we did, with every student. At a minimum, we suggest a kind of layered approach to assessment in which teachers first work diagnostically with students who have demonstrated difficulty on broad measures of reading. Then, they can work with other students as the need arises.

However, we caution that simply administering more and more assessments and recording the scores will miss the point. The value of in-depth classroom assessment comes from teachers having a deep understanding of reading processes and instruction, thinking diagnostically, and using the information on an ongoing basis to inform instruction (Black & Wiliam, 1998; Place, 2002; Shepard, 2000). Requiring teachers to administer grade-level classroom assessments to all their students regardless of individual student needs would not yield useful information or help teachers make effective instructional decisions. For example, administering a fourth-grade reading selection to Jesse, who is reading at first-grade level, would not provide useful information. However, using a fourth- or even fifth-grade selection for Tomas would. Similarly, assessing Jesse's word identification abilities should probably include assessments of basic sound/symbol correspondences or even phonemic awareness, but assessing decoding of multisyllabic words would

be more appropriate for Martin. This kind of matching of assessment to students' needs is precisely what we hope would happen when teachers have the knowledge, the assessment tools, and the flexibility to assess and teach children according to their ongoing analysis. Both long-term professional development and time are critical if teachers are to implement the kind of sophisticated classroom assessment that struggling readers need.

Second, the evidence points to the need for multilevel, flexible, small-group instruction (Allington & Johnston, 2001; Cunningham & Allington, 1999; Opitz, 1998). Imagine, if you will, teaching just the six students we have described, who could easily be in the same class. These students not only need support in different aspects of reading, but they also need materials that differ in difficulty, topic, and familiarity. For example, Tomas, Makara, and Andrew all need instruction in comprehension. However, Tomas and Andrew likely can receive that instruction using grade-level material, but Makara would need to use easier material. Both Makara and Andrew need work in vocabulary, whereas Tomas is fairly strong in word meanings. As second-language learners, Tomas and Makara likely need more background building and exposure to topics, concepts, and academic vocabulary as well as the structure of English texts than Andrew, who is a native English speaker. Furthermore, the teacher likely needs to experiment with having Tomas and Makara slow down when they read to get them to attend to meaning, whereas Andrew needs to increase his fluency through practice in below-grade-level text.

So, although these three students might be able to participate in whole-class instruction in which the teacher models and explicitly teaches comprehension strategies, they clearly need guided practice to apply the strategies to different types and levels of material, and they each need attention to other aspects of reading as well. This means the teacher must have strong classroom management and organizational skills to provide small-group instruction. Furthermore, he or she must have access to a

wide range of books and reading materials that are intellectually challenging yet accessible to students reading substantially below grade level. At the same time, these struggling readers need access to grade-level material through a variety of scaffolded experiences (i.e., partner reading, guided reading, read-alouds) so that they are exposed to grade-level ideas, text structures, and vocabulary (Cunningham & Allington, 1999). Some of these students and their teachers would benefit from collaboration with other professionals in their schools, such as speech and language and second-language specialists, who could suggest classroom-based strategies targeted to the students' specific needs.

The six clusters and the three strands within each one (word identification, meaning, fluency) clearly provide more in-depth analysis of students' reading abilities than general test scores. Nevertheless, we caution that there is still more to be learned about individual students in each cluster, beyond what we describe here, that would help teachers plan for instruction. Two examples make this point. The first example comes from Cluster 1, Automatic Word Callers. Tomas had substantial difficulty with comprehension, but his scores on the vocabulary measure suggested that word meanings were likely not a problem for him. However, other students in this cluster, such as Maria, *did* have difficulty with word meanings and would need not only comprehension instruction like Tomas but also many more language-building activities and exposure to oral and written English. The second example that highlights the importance of looking beyond the cluster profile is Andrew, our Slow Word Caller from Cluster 5. Although we know that in-depth assessment revealed that Andrew had difficulty with comprehension and fluency, we argue above that the teacher must do more work with Andrew to determine how much fluency is contributing to comprehension and how much it is a result of Andrew's effort to self-monitor. Our point here is that even the clusters do not tell the entire story.

Finally, from a school or district perspective, we are concerned about the disproportionate number of second-language students

who failed the test. In our study, 11% of the students in the school district were identified as second-language learners and were receiving additional instructional support. However, in our sample of students who failed the test, 43% were second-language learners who were *not* receiving additional support. Tomas and Makara are typical of many English-language learners in our schools. Their reading abilities are sufficient, according to school guidelines, to allow them to exit supplemental ESL programs, yet they are failing state tests and struggling in the classroom. In this district, as in others across the state, students exit supplemental programs when they score at the 35th percentile or above on a norm-referenced reading test—hardly sufficient to thrive, or even survive, in a mainstream classroom without additional help. States, school districts, and schools need to rethink the support they offer English-language learners both in terms of providing more sustained instructional support over time and of scaffolding their integration into the regular classroom. In addition, there must be a concerted effort to foster academically and intellectually rigorous learning of subject matter for these students (e.g., science, social studies) while they are developing their English-language abilities. Without such a focus, either in their first language or in English, these students will be denied access to important school learning, will fall further behind in other school subjects, and become increasingly disengaged from school and learning (Echevarria, Vogt, & Short, 2000).

Our findings and recommendations may, on one level, seem obvious. Indeed, good teachers have always acknowledged differences among the students in their classes, and they have always tried to meet individual needs. But, in the current environment of high-stakes testing and accountability, it has become more of a challenge to keep an eye on individual children, and more difficult to stay focused on the complex nature of reading performance and reading instruction. This study serves as a reminder of these cornerstones of good teaching. We owe it to our students, their parents, and ourselves to

provide struggling readers with the instruction they really need.

References

- Adams, M.J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Aldenderfer, M., & Blashfield, R. (1984). *Cluster analysis*. Beverly Hills, CA: Sage.
- Allington, R.L. (2001). *What really matters for struggling readers*. New York: Longman.
- Allington, R.L., & Johnston, P.H. (2001). What do we know about effective fourth-grade teachers and their classrooms? In C.M. Roller (Ed.), *Learning to teach reading: Setting the research agenda* (pp. 150–165). Newark, DE: International Reading Association.
- Antunez, B. (2002, Spring). Implementing reading first with English language learners. *Directions in Language and Education*, 15. Retrieved October 15, 2003, from <http://www.ncela.gwu.edu/ncbepubs/directions>
- Black, P., & Wiliam, D. (1998). *Assessment and classroom learning*. *Assessment in Education*, 5(1), 7–74.
- Block, C.C., & Pressley, M. (2002). *Comprehension instruction: Research-based best practices*. New York: Guilford.
- Cummins, J. (1991). The development of bilingual proficiency from home to school: A longitudinal study of Portuguese-speaking children. *Journal of Education*, 173, 85–98.
- Cunningham, P.M., & Allington, R.L. (1999). *Classrooms that work* (2nd ed.). New York: Longman.
- Duke, N.K., & Pearson, P.D. (2002). Effective practices for developing reading comprehension. In A.E. Farstrup & S.J. Samuels (Eds.), *What research has to say about reading instruction* (pp. 9–129). Newark, DE: International Reading Association.
- Echevarria, J., Vogt, M.E., & Short, D. (2000). *Making content comprehensible for English language learners: The SIOP model*. Boston: Allyn & Bacon.
- Edmondson, J., & Shannon, P. (2002). The will of the people. *The Reading Teacher*, 55, 452–454.
- Elmore, R.F. (2002, Spring) Unwarranted intrusion. *Education Next*. Retrieved March 21, 2003, from <http://www.educationnext.org>
- Goodnough, A. (2001, May 23). Teaching by the book, no asides allowed. *The New York Times*. Retrieved March 21, 2003, from <http://www.nytimes.com>
- Harris, A.J., & Sipay, E.R. (1990). *How to increase reading ability* (9th ed.). New York: Longman.
- Helfand, D. (2002, July 21). Teens get a second chance at literacy. *Los Angeles Times*. Retrieved March 21, 2003, from <http://www.latimes.com>
- Hiebert, E.H., Pearson, P.D., Taylor, B.M., Richardson, V., & Paris, S.G. (1998). *Every child a reader: Applying reading research to the classroom*. Ann Arbor, MI: Center for the Improvement of Early Reading Achievement, University of Michigan School of Education. Retrieved March 21, 2003, from <http://www.ciera.org>
- Klein, S.P., Hamilton, L.S., McCaffrey, D.F., & Stecher, B.M. (2000). What do test scores in Texas tell us? *Education Policy Analysis Archives*, 8(49). Retrieved March 21, 2003, from <http://epaa.asu.edu/epaa/v8n49>
- Kuhn, M.R., & Stahl, S.A. (2000). *Fluency: A review of developmental and remedial practices* (CIERA Rep. No. 2-008). Ann Arbor, MI: Center for the Improvement of Early Reading Achievement, University of Michigan School of Education. Retrieved March 21, 2003, from <http://www.ciera.org>
- Linn, R.L. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4–16.
- Linn, R.L. (n.d.). *Standards-based accountability: Ten suggestions*. CRESST Policy Brief. 1. Retrieved March 21, 2003, from <http://www.cse.ucla.edu>
- Lipson, M.Y., & Wixson, K.K. (2003). *Assessment and instruction of reading and writing difficulty: An interactive approach* (3rd ed.). Boston: Allyn & Bacon.
- McNeil, L.M. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.
- Nagy, W.E. (1988). *Teaching vocabulary to improve reading comprehension*. Urbana, IL: ERIC Clearinghouse on Reading and Communication Skills and the National Council of Teachers of English.
- National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 004 769). Washington, DC: U.S. Government Printing Office. Retrieved March 21, 2003, from <http://www.nationalreadingpanel.org>
- Olson, L. (2001). Overboard on testing. *Education Week*, 20(17), 23–30.
- Opitz, M.F. (1998). *Flexible grouping in reading*. New York: Scholastic.
- Paterson, F.R.A. (2000). The politics of phonics. *Journal of Curriculum and Supervision*, 15, 179–211.
- Pinnell, G.S., Pikulski, J.J., Wixson, K.K., Campbell, J.R., Gough, P.B., & Beatty, A.S. (1995). *Listening to children read aloud*. Washington, DC: U.S. Department of Education.
- Place, N.A. (2002). Policy in action: The influence of mandated early reading assessment on teachers' thinking and practice. In D.L. Schallert, C.M. Fairbanks, J. Worthy, B. Malock, & J.V. Hoffman (Eds.), *Fiftieth yearbook of the National Reading Conference* (pp. 45–58). Oak Creek, WI: National Reading Conference.
- Riddle Buly, M., & Valencia, S.W. (2002). Below the bar: Profiles of students who fail state reading tests. *Educational Evaluation and Policy Analysis*, 24, 219–239.
- Samuels, S.J. (2002). Reading fluency: Its development and assessment. In A. Farstrup & S.J. Samuels (Eds.), *What research has to say about reading instruction* (pp. 166–183). Newark, DE: International Reading Association.
- Shepard, L.A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29, 4–14.
- Snow, C.E., Burns, M.S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Stahl, S.A., & Kapinus, B.A. (2001). *Word power: What every educator needs to know about vocabulary*. Washington, DC: National Education Association Professional Library.

- Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 360–407.
- Stanovich, K.E. (1988). Explaining the difference between the dyslexic and garden-variety poor reader: The phonological-core variable-difference model. *Journal of Learning Disabilities*, 21, 590–612.
- Stanovich, K.E. (1994). Romance and reality. *The Reading Teacher*, 47, 280–290.