

Diana Sharp Consulting

Diana@dianasharp.com

**Supporting Teachers' Data-Driven Instructional
Conversations:
An Environmental Scan of Reading First and STEP Literacy
Assessments, Data Visualizations, and Assumptions about
Conversations that Matter**

Report to the
Information Infrastructure System Project

The John D. and Catherine T. MacArthur Foundation
&
The William and Flora Hewlett Foundation

November 2004

Table of Contents

Introduction	... 4
Section 1. Sources	... 7
Section 2. The Reading First Context	...9
2.01 Key Program Components	... 9
2.02 Assessment Categories	...10
2.03 Assessment Scan and Selection	...13
2.04 Concerns Highlighted by the Scan	...14
Section 3. Key Assessments	...17
3.01 Texas Primary Reading Inventory (TPRI)	...17
3.02 Phonological Awareness Literacy Screening (PALS)	...18
3.03 Dynamic Indicators of Basic Early Literacy Skills (DIBELS)	...19
3.04 Strategic Teaching and Evaluation of Progress (STEP)	...20
3.05 Terminology Issues	...22
3.06 Quick List of Component and Administration Differences	...25
3.07 Vocabulary	...28
3.08 Comprehension	...29
Section 4. Conversations and Visualizations	...31
4.01 The State of the Art in Reading First Schools	...31
4.02 How are Different Visualizations “Actionable?”	...34
4.03 Could Data Visualizations Increase Teachers’ Knowledge of Literacy Development?	...46
4.04 Which Visualizations Are Problematic?	...47

Section 5. Explicit Deep Assumptions	...49
5.01 STEP Assumptions	...50
5.02 DIBELS Assumptions	...57
5.03 TPRI Assumption	...59
5.04 PALS Assumption	...60
 Section 6. Environmental Supports for Conversations	...62
6.01 Links to Lessons that Precede Reading First	...62
6.02 Other Important Supports	...63
 Section 7. Summary and Future Directions	...67
7.01 Conversation Types	...67
7.02 Conversation Evolution	...70
7.03 Reading First Issues	...73
7.04 Final Thoughts	...74
 References	...75
Appendix A. Interviews	...77
Appendix B. Useful Websites	...78
Appendix C. Components of the Key Assessments	...80
Appendix D. Pathway Projects	...85

Introduction

To the casual observer, there is little to link the fields of early reading education and navigation. After all, reading teachers and young students usually interact in relatively small, indoor spaces, confined by schoolhouse walls. Navigators do their work in the unbounded out-of-doors.

Despite these differences, there are those who argue that the two fields now share a common point in their developmental histories: both are in revolutionary stages brought about by recent advances in technology and data-driven decision-making tools. In navigation, the revolution has come about through Global Positioning Software (GPS). Trimble, a leading innovator of GPS, describes it as technology that has “changed navigation forever” (see www.trimble.com). Using man-made “stars” (24 satellites and ground stations) as references, GPS allows users to tell where they are and where they are going anywhere on the planet.

Have similar advances given reading educators the equivalent of GPS tools? Perhaps. Developers of the DIBELS assessment at the University of Oregon, for example, use the analogy of GPS to describe how data from their assessment allow teachers to easily tell where students are in their literacy development, where they need to go, and the best path to get there (e.g., Good, 2002). Moreover, recent technological advances have led to software on handheld computers that teachers can use as they administer literacy assessments. The software (<http://www.wirelessgeneration.com/web/>) provides instant scoring, with links to websites that graph classroom data and provide tips for instructional interventions.

DIBELS is not the only assessment involved in this potential revolution. Reading First, a five-or-six- (estimates vary) billion-dollar federal initiative enacted into law as part of the No Child Left Behind Act of 2002, has focused the spotlight on data-driven decision-making at an unprecedented scale, advocating a variety of scientifically-based reading assessments as possible sources for instructional decisions. Unlike annual standardized tests commonly used to measure summative literacy achievement, the assessments selected under Reading First guidelines must include formative assessments to guide instruction and monitor student progress. The Reading First program has also

led to widespread adoption of literacy assessments across states and districts at an astonishing rate. More teachers are using the same assessments than ever before.

Beyond new assessment tools and new technologies, there may be a third crucial element necessary for a successful revolution in reading education: an increase in the professional conversations among adults in schools. Schools are complex, dynamic, social systems. Researchers in the Information Infrastructure System (IIS) Project (<http://csitech.uchicago.edu:8200/iis>) analyze these systems as part of their efforts to enhance literacy reform, and they emphasize the importance of socio-cultural practices that surround the use of literacy assessments and data-visualization tools. Their work highlights the role of teachers' conversations – with other teachers, with literacy specialists, and principals – as critical in achieving meaningful reform grounded in ambitious instruction.

This work has led to new questions about data-driven decision-making in literacy instruction. For example, how is data-driven decision-making different at the social level – in group conversations – than at the level of an individual teacher looking at data and making decisions? How can we best support “good” data-driven instructional conversations that will impact student literacy learning?

To begin addressing these questions, we must first explore related issues about data-driven decision-making – particularly in the context of Reading First – along with questions about the nature of instructional conversations that matter. As one of the initial steps in this effort, this report will:

- Examine the rapidly evolving state-of-the-art in instructional conversations supported by Reading First assessments and data visualizations,
- Describe how the work of the IIS project with the STEP literacy assessment intersects with those Reading First contexts, and
- Propose directions for increasing our understanding of how to support instructional conversations among the adults in schools, in ways that will make a difference for beginning readers.

The remainder of this report is divided into seven major sections:

- Section 1. Sources
- Section 2. The Reading First Context
- Section 3. Key Assessments
- Section 4. Conversations and Visualizations
- Section 5. Explicit Deep Assumptions
- Section 6. Environmental Supports for Conversations
- Section 7. Summary and Future Directions

In addition, appendices at the end of this report list helpful website resources and describe “pathway projects” to watch for developments in this area.

Section 1. Sources

The sources for this report include:

- Websites for the different Reading First assessments, including technical reports about the assessment contents.
- Reading First websites. Many states have sites specifically geared to their Reading First program, accessible through their state department of education site. Many of these sites include the state's Reading First approved proposal.
- Published books, articles in educational journals, and papers from recent educational conferences.
- Exploratory interviews with other researchers and key figures in the Reading First environment (See Appendix A).
- A meeting with Susan Hall, a national consultant on DIBELS, at the Center for Urban School Improvement, University of Chicago, on July 16, 2004.
- The Yellow Series of the STEP assessment (Center for School Improvement, University of Chicago).
- Visualizations of data from the STEPtool (for the STEP Assessment).
- Wireless Generation user guides for DIBELS, PALS, and TPRI assessments.

In examining the visualizations for Reading First assessments, this report focuses on the visualizations (handheld and Internet) provided by the Wireless Generation (WG) tools, rather than on visualizations available through computer data entry and other websites. This decision was made in part because the WG tools provide instant access to assessment results, potentially increasing the user's interaction with the data, including data-driven instructional conversations. Light, Wexler, & Heinze (2004) outline six factors that influence the functionality of data support systems for practitioners, and the first two factors are:

1. Accessibility. How accessible are the tools, and how do the tools support access to the data or information?
2. Length of feedback loop. How much time passes between the time the data are generated and when results are reported to the end user?

The Wireless Generation tools to date provide the easiest accessibility and fastest feedback loop for assessment data visualizations. For readers interested in other data visualization tools for Reading First assessments, Appendix B provides website addresses for such tools as AIMSweb and the DIBELS data system.

This is an exploratory report based primarily on an analysis of existing information and emerging insights from other researchers; this project did not gather data from observations of teachers' conversations. As the report will show, the field's understanding of how to support teachers' data-driven conversations is in very early stages. We clearly need research that includes systematic observations of teachers' conversations about their Reading First data, and this background report may provide hypotheses to guide those studies.

Section 2. The Reading First Context

Reading First funds have now been awarded to all fifty states, plus the District of Columbia, American Samoa, and the Bureau of Indian Affairs. (For communication purposes, the remainder of the report will refer to all 53 Reading First awardees as “states.”) The final awards were announced in fall of 2003, and by October 1, 2003, 1.78 billion dollars had been distributed.

The most recent major development in the program was the announcement in January 2004 that RMC Research Corporation, based in Portsmouth, NH, won the competitive bid process for a five-year, 36.8 million dollar contract to be the headquarters for a National Center for Reading First Technical Assistance. This national center will coordinate the work of three regional centers operated by Florida State University, the University of Texas at Austin, and the University of Oregon. A major goal of the national center is to provide states and districts with expertise for improving their reading programs; obviously, much of their work will focus on professional development and will likely push the field forward in understanding how to support data-driven instructional conversations.

2.01 Key Program Components

Almost every summary of Reading First employs two major lists. The first list is the set of beginning reading skills that Reading First targets, a set drawn from the National Reading Panel’s (2000) report on instruction linked to scientifically-based reading research (SBRR):

1. Phonemic awareness
2. Phonics
3. Vocabulary development
4. Reading fluency, including oral reading skills
5. Reading comprehension strategies

The second list is the set of components necessary to implement a Reading First plan. Speaking at the Harvard Institute for State-wide Literacy Initiatives in 2002, Chris

Doherty, director of Reading First, outlined the four pillars of a successful Reading First program:

It's about (1) massively increased professional development. It's about (2) a huge reliance on scientifically based programs, materials and instruction. It's about (3) an equally huge reliance on valid and reliable assessments.

And the last piece (4) is "insert state leadership here." The state has to be the one to pull this off. The state, the state, the state. Yes, they're going to be engaging the districts, but the state leadership – it couldn't be done any other way.

2.02 Assessment Categories

The North Central Regional Education Laboratory (NCREL) echoes Doherty's four pillars on its web site (<http://www.ncrel.org/rf/>), which is funded as a Reading First Subgrant Technical Assistance resource. Listing the key elements of successful programs as Assessment, Instructional Strategies and Programs, Professional Development, and Leadership, the site provides a wealth of resources for state education agencies.

In describing the assessment element, the NCREL site uses the official Reading First definitions of three assessment categories – Screening, diagnostic, and classroom-based instructional assessments – found in the *Final Guidance to the Reading First Program* document (US Department of Education, Office of Elementary and Secondary Education, April 2002, see <http://www.ed.gov/programs/readingfirst/legislation.html>). These definitions have short and long versions. The short version appears in the body of the *Guidance* document (p. 25) and the long version appears in the document's Appendix D. Table 1 shows both versions.

Table 1

Assessment Categories from Reading First**Final Guidance Document (U.S. Dept of Education, 2002). Short version (p. 25)**

Screening assessments determine which children are at risk for reading difficulty and need additional support.

Diagnostic assessments provide more in-depth information on a student's skills and instructional needs that forms the basis of a student's instructional plan.

Classroom-based instructional assessments determine whether students are making adequate progress or need more support to achieve grade-level reading outcomes.

Final Guidance Document (U.S. Dept of Education, 2002). Appendix D: Definitions

Screening Reading Assessment – An assessment that is valid, reliable and based on scientifically-based reading research. It is a brief procedure designed as a first step in identifying children who may be at high risk for delayed development or academic failure and in need of further diagnosis of their need for special services or additional reading instruction.

Diagnostic Reading Assessment – An assessment that is valid, reliable and based on scientifically-based reading research. It is used for the following purposes:

- i. identifying a child's specific areas of strengths and weaknesses so that the child has learned to read by the end of grade 3.
- ii. determining any difficulties that a child may have in learning to read and the potential cause of such difficulties.
- iii. helping to determine possible reading intervention strategies and related special needs.

Classroom-Based Instructional Reading Assessment – An assessment that evaluates children's learning based on systematic observations by teachers of children performing academic tasks that are part of their daily classroom experience and is used to improve instruction in reading, including classroom instruction.

The *Guidance* document, however, was not the only document that influenced states' Reading First assessment plans. For many states, the more influential document was a report published by the Institute for the Development of Educational Achievement (University of Oregon) in 2002. Often referred to as the IDEA report, or the ARA report, the full title is *An Analysis of Reading Assessment Instruments for K – 3*. Written by Ed Kame'enui, leader of an assessment committee charged by the National Institute for Literacy with providing assistance to states in selecting assessments, this report employed a set of four assessment categories with definitions, shown below, that are not identical to those in the *Guidance* document:

1. Screening measure - Brief assessment that focuses on critical reading skills strongly predictive of future reading growth and development, and conducted at the beginning of the school year with all children in grades K, 1, 2, and 3 to identify children likely to need extra or alternative forms of instruction.
2. Diagnostic measure - Assessment conducted at any time during the school year when more in-depth analysis of a student's strengths and weaknesses is needed to guide instruction.
3. Progress Monitoring measure - Assessment conducted a minimum of three times a year or on a routine basis (i.e., weekly, monthly, or quarterly) using comparable and multiple assessment forms to (a) estimate rates of reading improvement, (b) identify children who are not demonstrating adequate progress and therefore require additional or different forms of instruction, and/or (c) compare the efficacy of different forms of instruction for struggling readers and thereby design more effective, individualized instructional programs for those at-risk learners.
4. Outcome measure - Assessment for the purpose of classifying students in terms of whether or not they achieved grade level performance or improved.

According to the report, the IDEA assessment committee examined the federal language on assessment categories for Reading First, but they decided to create the new set of four categories and definitions after extensive discussions about (p. 25) “the clear intent of the legislation to improve the reading outcomes of students in Grades K - 3” and “the importance of reliably assessing children’s reading performance more frequently

(i.e., more than once a year) during critical periods of reading growth and development in kindergarten through grade 3...”

Many states adopted the language and categories of the IDEA report; however, the widespread belief that states were required to use these categories and definitions instead of the federal categories and definitions was a misconception, notes Kristi Santi, University of Texas-Houston. Some states mixed terminology; for example, Wyoming, has a Reading First plan with a categorization scheme of screening, diagnostic, and *classroom progress-monitoring devices*. Later sections of this report will return to issues related to the differences in assessment terminology and how they may affect teachers’ instructional conversations.

2.03 Assessment Scan and Selection

In selecting the assessments to feature in this report, along with the STEP assessment used by the Information Infrastructure System Project, I performed an environmental scan of state Reading First plans available on the Internet. Out of the fifty-three Reading First awards, the scan found forty-five states that either had their plans on the web or had sufficient information about their states’ assessment strategy embedded in their Reading First website.

No single assessment could fit all the needs of any state’s complete assessment plan. However, of the 45 state plans reviewed, 39 included the DIBELS (Dynamic Indicators of Basic Early Literacy Skills) assessment either as a requirement or as one of several options for participating schools as part of the overall assessment strategy. The dominance of the DIBELS assessment across Reading First plans made it a natural selection as a key assessment for this report. Eleven state plans in the sample mentioned the TPRI (Texas Primary Reading Inventory) and five state plans mentioned the PALS (Phonemic Awareness Literacy Screening) assessment as required assessments or approved options. Eleven states included at least some components of two of the three – DIBELS, PALS, and TPRI – assessments in their overall plan, and the New Jersey plan gave schools the option of choosing among any of the three as their preferred assessment. The similarity of TPRI and PALS assessments to the DIBELS and STEP assessments, along with their spread to multiple states, justified their inclusion in this report.

Notably, some state plans leave specification of some assessments up to the participating schools and districts, so the numbers quoted above likely underestimate the spread of the DIBELS, TPRI, and PALS assessments. It is also important to note that states vary in the particular sub-tests from these assessments that they include in their assessment plans.

Another potentially important component of the Reading First context is the fact that DIBELS was created by researchers at the University of Oregon and is also a required assessment for Florida teachers, making it of special interest to two of the three regional Reading First Technical Assistance Centers. [Recall that the three regional centers are: University of Oregon, Florida State University, and University of Texas-Austin.] TPRI was created by researchers at the University of Texas-Houston and is used by more than 95% of Texas schools. PALS, on the other hand, does not have a close tie (geographic or otherwise) to the Reading First Technical Assistance Centers. PALS was created by researchers at the University of Virginia and is used by most districts (over 130) in Virginia.

2.04 Concerns Highlighted By the Scan

The scan of Reading First plans and websites also highlighted two concerns about supporting data-driven instructional conversations, concerns specifically related to the Reading First context. The first concern is that the Reading First plan requirements may have in some cases spawned assessment strategies that are more complicated than they need to be. Barbara Foorman, from the University of Texas-Houston, clearly states the problem (see <http://www.ncrel.org/rf/foorman1.htm>):

Part of the challenge is that Reading First has five domains – in other words, there’s phonemic awareness, there’s phonics, there’s fluency, there’s vocabulary, there’s comprehension. And then with respect to assessment, people are looking at screening, diagnosis, progress monitoring, and outcome. And there’s a tendency to cross those [five] content areas with the [four] assessment requirements, creating a matrix with a lot of cells in it. And I think some states and districts have been confused, thinking they need to fill every cell with a different assessment... We’re ending up with a proliferation of assessments that will be very hard for the classroom teacher to manage.

My recommendation is to reduce the burden on the teacher and have the assessments *seamless* – so that the screening, diagnosis, and progress monitoring assessment can indeed be all one piece, all one

assessment, with different time points.... If you overload teachers with too much assessment, there's no way they'll be able to link it to instruction, because they'll have too many scattered pieces of information...

In fact, most state plans have complicated assessment matrixes that not only cross the four assessment categories with five reading components – they also add a third dimension of grade level (K, 1, 2, 3). Figure 1, below, shows an example of such a matrix.

Measures by Essential Reading Components	Screening				Diagnosis				Progress Monitoring				Outcome Assessments			
Grade	K	1	2	3	K	1	2	3	K	1	2	3	K	1	2	3
Phonemic Awareness																
Dynamic Indicators of Basic Early Literacy Skills–5 th Ed. Eng. & Spanish																
Initial Sound Fluency	X	X							X	X						
Phoneme Segmentation Fluency	X	X							X	X			X	X		
Comprehensive Test of Phonological Processing (CTOPP)					X	X										
Phonics																
Dynamic Indicators of Basic Early Literacy Skills–5 th Ed. Eng. & Spanish																
Letter Naming Fluency	X	X														
Nonsense Word Fluency	X	X	X						X	X	X			X	X	
Woodcock Reading Mastery Test–Revised																
Letter Identification					X											
Word Attack					X	X	X									
Stanford Achievement Test–9 th Edition																
Word Reading														X		
Word Study Skills														X	X	
Fluency																
Dynamic Indicators of Basic Early Literacy Skills–5 th Ed. Eng. & Spanish																
Reading–Oral Reading Fluency		X	X	X					X	X	X		X	X	X	
Gray Oral Reading Test IV (GORT-IV): Rate					X	X	X									
Vocabulary																
Woodcock-Johnson III Test of Achievement																
Picture Vocabulary		X	X	X		X	X	X		X	X	X				
Stanford Achievement Test–9 th Edition																
Reading Vocabulary														X	X	
Listening Comprehension														X	X	
Reading Comprehension																
Woodcock Reading Mastery Test – Revised																
Passage Comprehension		X	X			X	X							X	X	
Texas Primary Reading Inventory & Tejas Lee: Reading Comprehension									X	X						
Stanford Achievement Test - 9th edition																
Reading Comprehension														X	X	X
Sentence Reading														X		

Figure 1. Sample assessment matrix from a Reading First plan

In addressing this concern, researchers will need to answer important questions such as:

1. What data might demonstrate the value of seamless assessment systems?
2. Can we measure and evaluate the level of difficulty that teachers encounter in data-based conversations around multiple assessments, compared to conversations around more seamless assessment systems?
3. Can we support teachers who need to integrate information across assessments as they engage in data-driven instructional conversations? How?

The second concern highlighted in the scan of Reading First assessments is that there has been and continues to be confusion in the Reading First community about the terminology associated with the assessment categories, particularly as they relate to specific assessments. To better illustrate the problem, this report will first move into a description of the key assessments and then return to examples of how differences in the Reading First assessment terminology may have hindered the development of a transparent, common language about assessment that could benefit teachers' data-driven instructional conversations. This section will also begin to explore distinctions among the assessments that may affect instructional conversations, with a special focus on comprehension and vocabulary.

Section 3. Key Assessments

The descriptions that follow are intended to give a general overview of each assessment, with highlights of features related to later discussions in this report about potential differences in the instructional conversations that surround the assessments. For more complete descriptions, see each assessment's website, listed in Appendix B.

Another valuable resource for extensive information about TPRI, PALS, and DIBELS – including details about each assessment's reliability and validity evidence – is Rathvon's (2004) *Early Reading Assessment*.

3.01 Texas Primary Reading Inventory (TPRI)

The TPRI continues to be heavily studied by researchers at the University of Texas, and several new features have been added in the past few years. Formerly a K – 2 assessment, it now has a Grade 3 version available for the 2004-2005 year.

The TPRI has two main components: the screening portion and the inventory portion. The screening is an extremely brief measure, designed to determine which children are *not* likely to develop reading difficulties. The inventory section is longer, with branching rules that determine how many sections a student receives. Students who score low on initial skills tested in the inventory do not go on to tests of more advanced skills. All students, regardless of whether they pass the screening, take the oral reading or listening section of the inventory, which includes comprehension questions and – for those who can read orally – yields a score for fluency (number of words correct per minute) and an accuracy rating for the passage (frustrational, instructional, or independent, based on number of words miscalled.) Students are initially placed into the reading passages based on the score from a prior word reading task.

Before Reading First, most teachers administered the screening first, and then administered the inventory to students who did not pass the screening. With the advent of Reading First, researchers now encourage teachers to administer the inventory to all students, because only the inventory provides sufficient information for grouping students and making other data-driven instructional decisions.

The IDEA report (Kame'enui, 2002) judged TPRI to have sufficient evidence for use as a screening, diagnostic, and progress monitoring tool for all five reading

components (phonemic awareness, phonics, fluency, vocabulary, and reading comprehension), although the designation for vocabulary was based only on the kindergarten test of listening comprehension.

A major addition to the 2003-2004 edition was the Fluency Kit, which contains additional sets of oral reading probes that the test developers now require teachers to administer every six weeks for more frequent monitoring of children's progress. Teachers can administer these more often if they choose: there are eight stories each for first, second, and third grade, and teachers have access to all 24 stories to meet the varied needs of students in a classroom. Teachers begin to use the Fluency Kit once students can read at the level of "Story 3" in the original first grade inventory.

Appendix C lists the TPRI sub-tests for Grades K, 1, and 2. (Details for Grade 3 were not available at the time of this writing.)

3.02 Phonological Awareness Literacy Screening (PALS)

The PALS assessment shares many similarities with TPRI. Like the TPRI, it was first developed as a state assessment tool, aligned with state standards. Like the TPRI, it has an initial, brief screening portion, followed by more in-depth tests.

PALS comes in two versions: PALS-K (kindergarten) and PALS 1 – 3. All PALS-K tasks are given to all students. In PALS 1 – 3, the screening portions are termed "entry level tasks." The more in-depth levels, termed "diagnostic" are divided into "Level B" and "Level C" tasks. Students who score low on Level B tasks move on to Level C tasks. As with TPRI, all students – even those who take only the entry-level tasks – also take a "Level A" task of oral reading. As with TPRI, students are placed into a passage depending on their performance on a word-reading task.

As with TPRI, PALS use has evolved with Reading First. Before Reading First, most teachers administered the Level B and C portions of the test only to children who did not pass Level A; now teachers often administer those portions more widely. Initially a fall and spring administered test only, it now has mid-year assessment test. Prior to Reading First, children in grades 1 and 2 who scored high enough at certain test windows were exempt from future PALS administration; now, Reading First schools continue to assess these children.

Unlike with TPRI, PALS developers have not been moved by the recent widespread focus on frequent progress-monitoring to include even more frequent assessment materials, such as the TPRI Fluency Kit or the DIBELS progress-monitoring materials (described below).

PALS was not reviewed for the IDEA (2002) report. However, *PALS and Reading First*, a document on the PALS website, (downloadable from <http://pals.virginia.edu/Reading-First/>) states that PALS is appropriate for screening, diagnostic, and progress-monitoring purposes. The document states that “the use of PALS as an assessment of outcomes is probably not justified,” and notes that in Virginia, student success is measured with a different statewide assessment.

Within the Reading First community, however, PALS’ value as an outcome measure will be further tested. The Nevada Reading First plan was approved for using PALS as a screening, diagnostic, and outcome measure for all five components of reading, with schools able to choose their own progress-monitoring measure. Delaware is also using PALS as an end-of-the-year accountability measure for vocabulary and comprehension in grades K and 1, in addition to its use as a screening and a mid-year assessment in K – 3. In New Hampshire, schools can choose DIBELS or PALS as an outcome measure.

Appendix C lists the sub-tests for PALS-K and PALS 1 – 3.

3.03 Dynamic Indicators of Basic Early Literacy Skills (DIBELS)

DIBELS is another continually evolving assessment. When reviewed by the IDEA report (Kame’enui, 2002), it had sub-tests covering phonics, phonemic awareness, and fluency. It now has a total of 7 sub-tests, covering all five reading components. Most state reading plans include only the phonics, phonemic awareness, and fluency components, but a few – such as Virginia and New Hampshire – have been approved to use DIBELS for vocabulary and/or comprehension.

DIBELS measures are all timed and extremely brief. Students’ scores for each measure are based on their performance during a one-minute task. With instructions and sample questions, each measure takes about three minutes to administer.

Unlike TPRI and PALS there is no separate screening tool. All students receive the same measures during benchmark periods (three times a year in most cases; four in some).

From their scores at these benchmark periods, students are flagged in terms of their risk level for achieving literacy goals. Students designated as “High-risk” or “Some risk” may fail to reach literacy goals unless they receive extra support.

DIBELS stresses the importance of not just reaching literacy goals, but reaching them on time. Students who score low on a measure on or after the critical target date for that skill level are scored as “Deficit” or “Emerging,” to indicate that they have missed a benchmark goal.

In addition to the benchmark measures, DIBELS has extensive materials for repeated testing. No other assessment can yet match this feature. There are approximately 20 different forms for each measure at each grade level. These additional materials are called “progress-monitoring” materials, and their purpose is to help teachers ensure that low-scoring children are not just making improvement, but are making progress at a fast enough rate to meet subsequent goals. Teachers are encouraged to administer these progress monitoring tests as often as weekly – or at the least frequent end, every six weeks – to children at the highest risk levels, and to use these measures to make decisions about whether the additional support a child receives is working.

The IDEA report (Kame’enui, 2002) judged DIBELS to be an appropriate measure for screening, progress-monitoring, and outcome purposes. It was not approved for diagnosis, presumably due to the brevity of its measures. As with PALS, the Reading First community appears to be exploring extended uses. The North Carolina plan lists TPRI as its preferred measure for diagnosis, but it lists DIBELS as another diagnostic option for phonemic awareness, phonics, and fluency. The Washington state plan was not available, but the May 2004 application for Reading First schools in Washington lists DIBELS as the required diagnostic assessment for phonemic awareness, phonics, and fluency (see

<http://www.k12.wa.us/curriculuminstruct/reading/readingfirst/GrantInformation.aspx>).

Appendix C lists the DIBELS measures for each grade level.

3.04 Strategic Teaching and Evaluation of Progress (STEP)

The Strategic Teaching and Evaluation of Progress (STEP) assessment measures the literacy growth of children in grades K – 3 and organizes their performance around a developmental map. This map shows a typical developmental path segmented into twelve discrete steps. Within each step, students are expected to meet targets across such areas as phonemic awareness, concepts about print, letter-sound correspondence, developmental spelling, letter/word knowledge, reading accuracy, reading rate, fluency, comprehension, and use of reading strategies. Kindergarten students who are on track in their literacy development move through a Pre-Reading assessment and Steps 1 – 2, first graders move through Steps 2– 6, second graders move through Steps 6 – 9, and third graders move through Steps 10 – 12.

Each Step has an individual assessment battery. When children meet the targets on a particular Step battery, they are considered to have achieved that step. Each teacher, has the same complete set of assessment materials for all twelve steps and can administer any Step battery to any child in the class as judged appropriate. When a teacher suspects, through observations of classroom work, that a child has developed new skills, the teacher can administer the next Step battery to verify the progress and gain insight into the child’s continuing needs. Teachers administer the first assessment at the beginning of the year to begin instructional planning, and they ideally assess each child at least three times during the year to monitor progress and adjust instruction.

A key feature of STEP is that each Step battery includes – in addition to tests of isolated skills – an attractive book, complete with illustrations, that resembles “real” books tied to a particular reading level. As part of the battery, the child reads the book; or, at early steps, reads the book with the examiner. There is a book introduction that happens prior to the reading, just as in most classroom book activities. The teacher takes a running record to record the child’s reading and then engages the child in a “comprehension conversation.” This conversation is one of the most challenging pieces to administer – and, according the developers – is one of the most powerful components of the assessment. Only partially scripted, the conversation can include prompts as judged appropriate by the teacher to elicit the depth of a child’s understanding about the

text, and to help the teacher answer such questions as: How did the child make sense of the story? How did the child draw conclusions and extend meaning? How did the child use information from the text to support his or her thinking?

Unlike the TPRI Fluency Kit and the DIBELS progress monitoring materials, the STEP assessment is not designed for highly repeated testing of any child. There are two versions of each Step battery, allowing for cases in which a child does not achieve the targets of a step and must be re-tested later, after additional instruction. The versions are separated into the *Yellow* series and the *Purple* series, and teachers typically have both series to use throughout the year.

Like the PALS assessment, the STEP assessment was not reviewed for the IDEA (Kame'enui, 2002) report. David Kerkow, Lead Researcher and STEP developer, states that STEP fits all of the IDEA categories (screening, diagnostic, progress-monitoring, and outcome). Reports on the development, validity, and reliability of STEP are forthcoming.

3.05 Terminology Issues

As noted in the previous section, the scan of Reading First plans revealed a lack of consensus – and in some cases, apparent confusion – about the assessment category terms (screening, diagnosis, progress-monitoring, and outcome) used in the IDEA (Kame'enui, 2002) report. This confusion may give rise to confusion in teacher's instructional conversations, making it relevant for this report.

Part of the problem no doubt lies in the subtle terminology differences for “progress-monitoring” in the DIBELS materials and in the IDEA report. The IDEA report specifically tied “progress-monitoring” to two different types of cycles: “a minimum of three times a year” and “a routine basis (weekly, quarterly, monthly).” DIBELS reserves the term “progress-monitoring” only for the latter, more frequent administration times for children who score low on the “three-times-a-year” assessment. DIBELS calls the “three-times-a-year” assessment a “benchmark assessment” to distinguish it from the “progress-monitoring assessment.”

TPRI, on the other hand, prior to the introduction of the Fluency Kit, had only the three-times-a-year assessment material, which were termed “progress-monitoring” assessments. The confusion was especially high in states such as Colorado, where teachers had a choice between DIBELS and TPRI. In May 2004, Colorado issued an

assessment update in an attempt to straighten out the terminology; the solution was to add additional hyphenated terminology (see

http://www.cde.state.co.us/cdecomp/rf_assessment.htm):

There is some confusion in the way the terms “progress monitoring” and “benchmark” are understood and used, especially at schools using DIBELS...To help clarify the terminology, CRF [Colorado Reading First] will begin using the categories:

- *Progress-monitoring-benchmark* describes the beginning, mid, and end-of-year required assessments.
- *Progress-monitoring-intensive* is used for those more frequent interim assessments for strategic or intervention students.

Whether this removes the confusion remains to be seen.

Other states using DIBELS bypass the problem by focusing on the Reading First terminology for screening [“determine which students are at risk for reading difficulty and need additional support”], rather than the IDEA definition of “screening” which specifically tied it to assessments “conducted at the beginning of the school year” only. Susan Hall, a national consultant on DIBELS, reports that in states such as Louisiana, the beginning-mid-end year assessments of DIBELS (the “benchmark assessments”) are referred to as “screenings.”

The IDEA definition of “screening” as a “beginning of the year” assessment can also cause confusion about the usability of the TPRI for kindergarten, because TPRI intentionally does not assess kindergarten children until the middle of year, to give them time to acclimate to school. However, TPRI still fits the Reading First definition of “screening” because it assesses the child’s risk level. Kristi Santi, University of Texas-Houston, reports that this confusion led the TPRI developers to come out with a statement justifying their decision to delay kindergarten testing until mid-year.

The terminology problems continue with the category of “diagnostic.” In the IDEA definitions, language about using assessments to understand “students’ strengths and weaknesses to guide instruction” appears only in the “diagnostic” category. For TPRI and PALS, this use of the term aligns nicely with the “inventory” section of the TPRI and the more in-depth levels of PALS. However, DIBELS – which is usually judged to not be a diagnostic assessment and does not have multiple levels – is also being

used to do more than “screen” students or “progress-monitor” high risk students: it is being used to guide instruction right from the beginning of the year. Florida, for example, does not consider DIBELS to a diagnostic assessment, but the state encourages teachers to use DIBELS for “creating small groups” and “guiding instruction” (Roehrig, 2004).

It is worth noting that at a recent national DIBELS Summit conference (March 11-13, 2004, Santa Ana Pueblo, New Mexico), conference leaders urged teachers and Reading First leaders to go beyond the “risk” categories of DIBELS scores and look at children’s error patterns on test items, to better understand student’s strengths and weaknesses and to use this information in planning. For example, they clarified that the Oral Reading Fluency passages could be used as a “double-duty” measure to help teachers learn about children’s development of advanced alphabetic principle knowledge. (The DIBELS Nonsense-Word-Fluency task assesses basic alphabetic principle knowledge using VC and CVC patterns.) In looking at the words that children missed on the Oral Reading Fluency passages, teachers could look for answers to such questions as: “Can the child read multiple syllable words? Silent-e words? Words with vowel digraphs?” Such use of DIBELS appears to place it more closely in the territory of diagnostic assessment.

Susan Hall, author of the forthcoming book, *Designing Interventions with DIBELS Data*, responds to questions about whether using DIBELS to guide instruction in this way is inappropriately “diagnostic.” She states, “This (analysis of error patterns) is just allowing teachers to begin intervention. If the student fails to make adequate progress, then it’s time to do diagnosis.”

This conception of the “diagnosis” label is echoed in the May 2004 assessment update (see http://www.cde.state.co.us/cdecomp/rf_assessment.htm) put out by Colorado Reading First. Citing Florida’s use of the terms, it emphasizes that “Although they [diagnostic tests] can be given as soon as a screening test indicates a child is behind in reading growth, they will usually be given only if a child fails to make adequate progress after begin given extra help in learning to read.” Notably, the Florida Reading First plan also specifies that two of its three assessments classified as diagnostic options (*Diagnostic Assessment of Reading*, and *Early Reading Diagnostic Assessment*) are more

appropriately administered by a reading specialist or school psychologist than a classroom teacher.

Finally, Barbara Foorman (see <http://www.ncrel.org/rf/foorman1.htm>) offers a conception of diagnosis as an assessment that you do “several times across the year, each time when you’re thinking about re-grouping for instruction.”

The “outcome” conception also lacks clear-cut consensus in the Reading First community. TPRI and PALS developers both consider their assessments to be inappropriate as an outcome tool. Foorman, in advocating for a seamless assessment system (<http://www.ncrel.org/rf/foorman1.htm>), notes, “Outcome is a different animal, so I can certainly see different outcome assessments. So a teacher may be doing one assessment that’s a combined screening, diagnostic, progress-monitoring [assessment], and another assessment later, at the end of the year, that is for outcome.” There is not, however, agreement about whether outcome measures must and should be distinct from the assessments used for the other purposes. For example, DIBELS, approved as an outcome measure by the IDEA (Kame’enui, 2002) report, is used in some Reading First plans (e.g., Maine, New York) for screening, progress monitoring, and outcome purposes.

Clearly, researchers interested in teachers’ conversations about their Reading First assessment data need to be alert for confusions resulting from the terminology of assessment categories. The final section of this report will return to this issue and propose possible directions. At this point, the next section will turn to distinctions among the assessments at the level of their task components and administration procedures, distinctions that may affect teachers’ conversations.

3.06 Quick List of Component and Administration Differences

The distinctions in this section are not exhaustive; they are a starting point for hypotheses about differences in the test components and administration procedures that may affect teachers’ conversations. If these differences do affect teachers’ conversations, more likely than not they are also affecting their individual decision-making and instruction. For example, Mary Fowler, University of Virginia, notes, “If something is on the assessment, then teachers tend to think that something’s important and teach it. If something’s not on the assessment – like Concept of Word – then all of a sudden it’s not an important concept to teach.” One could similarly argue that what’s on the assessment,

or the way the assessment is given, also influences how important teachers think it is to talk about a particular skill. What follows is a starter list of differences, intended to help conceptualize the distinctions among the assessments. It is too early yet to judge which of these distinctions will ultimately matter for teachers' conversations and children's reading outcomes:

- PALS, STEP, and TPRI all have spelling tasks; DIBELS does not.
- TPRI is the only assessment to delay kindergarten administration until the middle of the year.
- PALS has the fewest possibilities for administration, no more often than three times a year. DIBELS has the most possibilities for administration with its extensive progress-monitoring materials for use as often as weekly. TPRI has a Fluency Kit for administering probes every six weeks or more for low-scoring children. STEP is flexibly designed for administration about three or four times during the year with the possibility of administering the same Step battery twice, if a child fails it the first time. Developer David Kerbow notes that the philosophy of STEP is that teachers will be heavily involved in careful classroom observation of children's progress in authentic literacy tasks, rather than in using as many formal assessments as teachers who use DIBELS' progress monitoring tools.

- DIBELS is the only assessment that asks children to read nonsense words.

There is a growing controversy in the reading field about the use of nonsense words in assessment, with some authorities clearly advocating their use – as a way of avoiding the under-identification of children who have memorized sight words but have underdeveloped alphabetic principle skills (Rathvon, 2004) – and others arguing heavily for the use of real-word reading (Fuchs, Fuchs, & Compton, in press). Naomi Hupert, from the Center for Children and Technology, notes that in some – albeit infrequent – cases, teachers in New Mexico misunderstood the purpose of the nonsense-word task and mistakenly began to infuse nonsense-word-reading tasks into their instruction.

- TPRI and DIBELS do not include any real-word reading tasks for kindergarten administration; PALS includes this as an optional task for kindergarteners who are reading. STEP teachers in kindergarten are free to administer any Step level – including those with real reading tasks – to kindergarten children who are reading.

- PALS is the only assessment that asks a child (kindergarten only) to do “pretend” or “non-decoding” reading [not their terms] with a memorized rhyme (Concept of Word task). This type of early reading in a predictable book is notably different from the heavy emphasis on reading in decodable texts that is widespread in Texas.

- PALS and STEP are the only assessments that record the types of errors (e.g., substitutions, insertions, omissions) children make in their readings. PALS also codes whether children misread function or content words. Mary Fowler, University of Virginia, elaborates:

“Do they say ‘a’ for ‘the’ all the time? Is their score coming down because of tiny function word errors, or is it coming down because they couldn’t read “vegetable” or some other big word they couldn’t decode. This gives some insight into what’s causing difficulty when they’re reading. Is it a decoding issue or is it more skipping over some of the smaller words?”

- STEP and PALS are the only assessments with a fluency rubric for scoring children’s prosody with text.

- PALS and TPRI place students into a level of passage reading for the fluency and comprehension measures based on the child’s performance on a word recognition measure. DIBELS passages are all end-of-grade level, based on Curriculum-Based-Measurement research (Shinn, 1989). Foorman, Santi, & Berger (2002) note their preference for passage placement, stating, “This approach has advantages...because the students are placed into passages at their instructional level and can read, therefore, with sufficient accuracy that fluency and comprehension can be measured without decoding confounds. Teachers can use differential growth in fluency as a basis for forming small group instruction.” In STEP, oral reading also occurs in a book designed to be close to the child’s instructional level.

- TPRI and STEP allow children to look back at the text when answering comprehension questions; PALS does not.

- TPRI is the only key assessment with a pure listening comprehension task (for kindergarten only).

By far, the biggest distinctions among the assessments appear to lie in the areas of comprehension and vocabulary. The next sections examine these component areas in more detail.

3.07 Vocabulary

The *Guidance* document for Reading First defines Vocabulary skill as follows (see <http://www.ed.gov/programs/readingfirst/legislation.html>):

Vocabulary Development – Development of stored information about the meanings and pronunciation of words necessary for communication. There are four types of vocabulary:

- Listening vocabulary – the words needed to understand what is heard
- Speaking vocabulary – the words used when speaking
- Reading vocabulary – the words needed to understand what is read
- Writing vocabulary – the words used in writing

None of the key assessments specifically measure all of these areas apart from other skills. Moreover, DIBELS is the only assessment with a targeted vocabulary sub-test that is completely separate from other skill measures. In its Word-Use-Fluency (WUF) measure, the teacher presents the student with a word and asks the child to use the word in a sentence. If the child uses the word in accordance with its meaning, the teacher records the number of words in the child's sentence. Benchmarks for the WUF sub-test are under review and the developers expect them to be available shortly. At present, teachers use local norms to identify children who may need extra support in vocabulary.

PALS is used in some Reading First plans to measure vocabulary, but as the *PALS and Reading First* document (<http://pals.virginia.edu/Reading-First/>) notes, vocabulary is measured in PALS with the Word Recognition test. This test looks only at the child's ability to read words and does not assess knowledge of word meaning.

As noted above, TPRI was judged in the IDEA report (Kame'enui, 2002) to assess vocabulary, but only for kindergarten, and only as embedded in the listening comprehension sub-test. Kristi Santi, University of Texas-Houston, states further:

Vocabulary is measured in Grade 3 through the use of two questions in the Reading Accuracy, Fluency, and Comprehension portion of the Inventory.

In Kindergarten through Grade 2 there is no direct measure of vocabulary, although we are working a vocabulary measure to incorporate into the Inventory section of the TPRI and Tejas LEE. Vocabulary is, in a sense, embedded in the reading and comprehension portions of the inventory.

Santi also notes that Texas teachers are encouraged to use classroom-based assessments to progress-monitor vocabulary development.

STEP also does not have a separate measure of vocabulary, although portions of the comprehension conversations may give teachers insight into children's vocabulary knowledge.

3.08 Comprehension

PALS and TPRI both use one-answer multiple-choice questions to assess comprehension. In TPRI, these questions are clearly divided into questions about explicit and implicit information.

DIBELS and STEP both use story-retelling as a measure of comprehension, but the results are scored very differently. DIBELS uses a word count approach: the number of words that a child uses to retell a story should be about 50% of the child's word-per-minute score when reading the stories for the Oral Reading Fluency (ORF) measure. If a child is reading 60 words a minute, then the child should use about 30 words to retell the story. If the child is reading at the rate of 60 words per minute but the retell score is 15 words or less (25% of the word-per-minute score), then comprehension may be low, and the authors caution that the ORF measure may not be a good indicator of the child's reading ability.

STEP, in contrast, scores retellings with a four point rubric, ranging from *1 – Response does not indicate understanding* to *4 – Response shows exceptional understanding*. Each point on the rubric is illustrated with examples, such as “Facts are not stated in sequence” (1 on the scale), and “Shows insight into characters” (4 on the scale). Teachers also record any significant confusions or any noted instances of interpreting the text.

DIBELS uses retelling as early as first grade; STEP employs it only at Step 8, typically in the middle of second grade. In addition to the retelling measure, the STEP assessment includes an emphasis on comprehension at all STEP levels that represents one

of the largest distinctions among the key assessments. [In fact, the Colorado Reading First site – http://www.cde.state.co.us/cdecomp/rf_assessment.htm; see the May 2004 update – cautions teachers that TPRI and DIBELS only provide “dipstick” measures of comprehension.] In STEP’s Comprehension Conversation, described briefly above, the questions move beyond one-answer questions and beyond explicit/implicit text understanding. The conversations include these four types of questions: Factual/Literal questions, Inferential questions, Critical thinking questions (with a variety of possible answers), and Personal-Opinion questions (e.g., students decide whether they agree with the author, or consider whether they agree with characters’ thoughts, statements, and actions).

Teachers are encouraged to prompt children for complete, elaborate answers, using such prompts as “What in the book makes you think that? Why do you think that? What do you mean when you say....” Assessment developer David Kerbow, University of Chicago, notes:

If you’re asking a critical thinking question, and their [the student’s] first answer is just a recounting of an incident, we want to be able to follow up with a prompt that encourages a child to elaborate. It’s what you want to do in a classroom situation, and it gives you a better window into how the child is making meaning with that book. If you don’t know that, where do you go in your instruction?

Looking at these component task differences represents just one level in exploring the distinctions among the assessments that may have implications for supporting teachers’ data-driven instructional conversations. At another level, the data-visualizations may affect how teachers talk about their data; and at an even deeper level, there may be fundamental distinctions among the assessments and/or their use in Reading First programs that illuminate differences in the field about what is considered to be a “good” data-driven instructional conversation. The next major sections explore these other levels.

Section 4. Conversations and Visualizations

4.01 The State of the Art in Reading First Schools

What do we know about data-driven decisions in the real world of public schools designated as eligible for Reading First support due to low reading achievement?

We know that they are happening (in some places more than others), and not just among individuals. Group conversations about data from all of the three Reading First key assessments are part of the state-of-the-art landscape.

Mary Fowler, University of Virginia, reports:

Since it's (PALS) a state-wide assessment, not just Reading First, and we've been doing it for several years, you hear stories about people who are really using the information and it's valuable – and others [where teachers] aren't even giving the assessments themselves and are barely looking at them. In our courses and Reading First academies, we talk about how important the data is and you need to be having those conversations... at the grade level in the school.... Definitely, there are schools where teachers get together, talk about the data, and use it to plan instruction.

Kristi Santi, University of Texas-Houston, adds:

We haven't looked per se at the social level (of data-driven decisions), but we do encounter it every day when our mentors are in schools working with the teachers. In the IERI grant, we're out in 40 schools with the on-site mentoring condition, where master teachers go out and have group meetings about how to interpret results [from TPRI] and plan lessons. They have three or four meetings to get to that lesson plan, and then they observe those lessons. We just finished Year 1 ...[and we've seen] how teachers' involvement and enthusiasm helps get more teachers to the meeting, and everybody's different perceptions influence how well the meetings go.

Naomi Hupert, Center for Children and Technology, speaking about Reading First teachers who used DIBELS in New Mexico last year for the first time, says:

We're seeing what teachers are starting to ask for, as a result of thinking about what to do when they get data. Teachers are having a novel experience of getting data right away. Usually they don't get it for weeks, months, a year, and it's meaningless. Now they get it instantly. So what we're seeing is that they have this information, and they're having

meetings with literacy coordinators and coaches and are saying to each other, “So now what? Now what do we do with this information?”

There was general agreement among the sources consulted for this report that our understanding about data-driven conversations is still in very early stages. At this point, most of the supports for conversations about the key Reading First assessments appear to focus on a “first-level” of data: a summary for teachers that shows in one display the performance of children across all the sub-tests. Efforts to help teachers use data are still primarily exploring how to help teachers look at these one-page summaries, use them to form instructional small groups, and then link appropriate instructional activities to the appropriate small groups.

Marguerite Held, a teacher trainer at the University of Texas-Houston, gives a typical example:

[I]nitially, after the teachers give the TPRI, we ask them as a grade level to bring in their classroom summary sheet that has all their data on it. As a group, we talk about how they can analyze their scores. We have them do some color-coding as a group for skills that are still developing, and some percentages for skills that are still developing. This helps them determine where their weaknesses are. We do this as a whole group, and teachers discuss each other’s groups. I really encourage that – I think you learn from other teachers. We show strategies to group their students; we do this all together with lots of discussion. If the school is farther behind, then I’ll do a model (not any of their data) and take them through a generic sheet. [With some schools], we go straight into their own data.

Mary Fowler, University of Virginia, describes similar experiences in PALS schools:

“[Teachers] are probably going to be looking at overall scores. . . what kind of grouping decision they might be able to make based on the data. They might be able to look at the reading portion and decide how the kids fit in their reading groups, or look at the spelling and see what needs there are for grouping the kids for that . . . mostly taking the data and saying, “How does that inform my instruction?”

Of course, forming instructional groups based on data is one thing; knowing what to do with those groups is another. TPRI and PALS both have extensive on-line text and video materials for teachers, with suggestions on intervention activities and suggestions for choosing appropriate activities for different small groups. TPRI is moving toward

extremely explicit links: Their latest development uses a webtool and relational database that ties activities to the individual words that children miss on the assessment. Teachers administer the TPRI using a palm computer and Wireless Generation software, upload the data onto the Internet; the webtool analyzes the specific errors that children made on the test items and then pinpoints specific lessons (from basals available to Texas teachers) that include practice with words similar to those that the children missed.

DIBELS has been somewhat behind the curve in developing materials for teachers, although their related “Big Ideas in Beginning Reading” website has many instructional suggestions. Also, sources for this report indicated that Sopris-West is developing a CD to help teachers make the data-to-instruction link. In addition, Wireless Generation is creating an ACT feature as part of the DIBELS tool that will provide some advice for teachers – on their palm computers - about what their data indicates they should do to support children.

Currently, however, teachers using DIBELS have to go farther on their own to get the resources they need to link data to instruction, rather than directly accessing these resources from the same place that gives them their reports. In Florida, for example, teachers get their DIBELS reports from Florida’s Progress Monitoring and Reporting Network (PMRN) (see <http://www.fcrr.org/pmrn/>). Alysia Roehrig, Florida State University, notes:

If you don’t know what to do with [students] once you’ve grouped them, then it’s not worthwhile. That’s the really hard part. In our wonderful system [PMRN], there aren’t instructional tips. Just [reports such as], “they’re high-risk on fluency.”

Naomi Hupert’s observations of New Mexico DIBELS users exemplify the importance of giving teachers easy access to instructional advice, now that data on student needs is so much more timely and capable of evoking conversation:

It’s still novel that [teachers] use data to make decisions. Before teachers had the Palm tool, they used paper and pencil [for DIBELS], and the coordinators made graphs to show them where the kids fell out. Those initial experiences for the coordinators and the teachers, seeing it quickly from administering it, those were the instigators of conversations, like: “Oh, I had no idea that my kids were so scattered out, so much of a division between the good readers and the struggling, or that they were so

concentrated in one area....” Teachers were expressing surprise and interest.

The next thing that happens after surprise is the question, “OK, so what do I do now, and where do I go with this information?” We heard that teachers were having discussions with each other about strategies and were feeling at sea about what strategies were needed to address different needs.

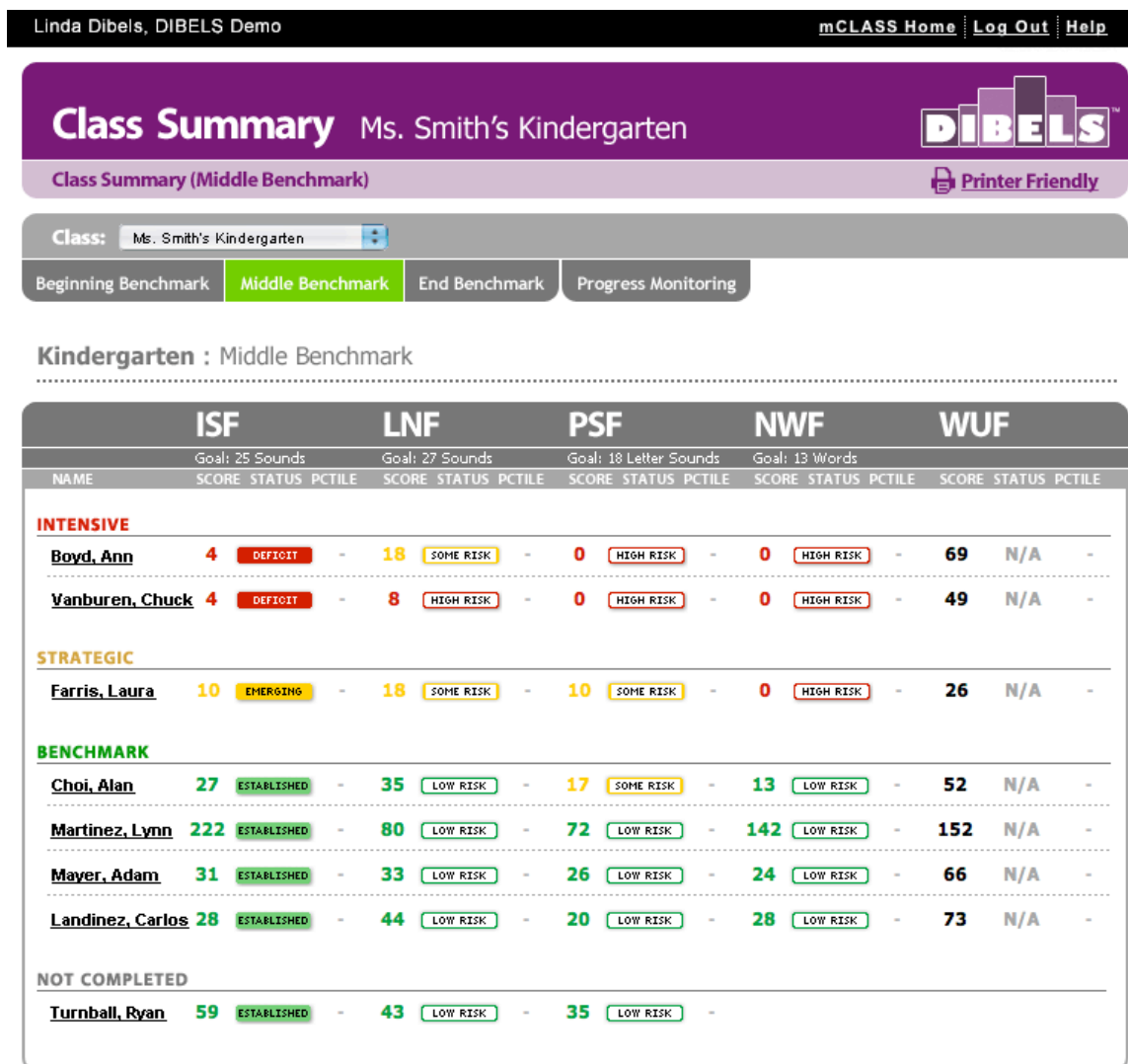
4.02 How are Different Visualizations “Actionable?”

What is clear from the current stage of Reading First implementations is that the state-of-the-art in providing data visualizations currently outpaces the state-of-the-art in using them. All of the key assessments, including STEP, provide multiple views of the data beyond the overall class summaries. For example, the mClass DIBELS (Wireless Generation) web tool offers all of the following views, from the overall class summary down to the student’s responses to any individual test item (p. 24, DIBELS User Guide):

- The Benchmark Class Summary Page provides a convenient summary of your students’ reading skills, based upon their recent Benchmark Assessment. This high-level snapshot is designed to help you flag a student who may need your attention.
- The Progress Monitoring Class Summary Page provides a similar view of your class. It provides information not about Benchmark Assessments but about the on-going progress-monitoring assessments that you have administered to your class.
- The Student Summary Page provides summary information on all the Benchmark assessments you have conducted for a selected student, plus information about the student’s current Progress Monitoring. This page gives you a graphical and numerical representation of the student’s performance across Benchmark periods.
- The Student Monitoring Page provides a detailed review of a particular student’s performance on a measure that is being progress-monitored. The information is represented numerically as well as on a longitudinal graph.
- The Probe Details Page allows a teacher or assessor to analyze the complete reproduction of any probe that a particular student has taken.

As the field moves towards a better understanding of how to help teachers use various data views, what questions might help us learn about the value of particular data visualizations? One way to frame these questions may be to organize them around the concept of what is “actionable” for the teacher. (Thanks to Larry Berger, Wireless Generation, for suggesting this concept.)

For example, assigning students to receive additional instruction, such as one-on-one tutoring, that goes beyond regular classroom instruction, is one action that teachers can take. The overall class summaries of each of the key Reading First assessments provide actionable information for this decision in different ways. DIBELS, for instance, provides a highly clear and comprehensive view for this decision, because it specifically divides children into “Intensive” groups (needing the most additional instruction), “Strategic” groups (needing less or less-intensive additional instruction) and “Benchmark” groups (students who should remain on target with regular classroom instruction). The same view also shows teachers the overall risk category for students on each DIBELS sub-test. Figure 2, below shows this view:



© 2004 Wireless Generation. All rights reserved.

Figure 2. DIBELS class summary view

The PALS class summary also separates students into three groups; unlike the DIBELS view, it does not simultaneously provide information about the child's needs in specific areas. Instead, it shows teachers the child's overall summed score across tasks, as in Figure 3, below:

Class Summary Ann Parks Class



School: Greenbrier Elementary Class: Ann Parks Class

First Grade 2004–2005

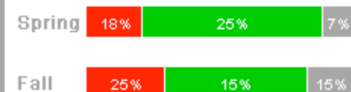
Class Owners: Ann Parks, Ginny Wean

	FALL	MID-YEAR	SPRING
IDENTIFIED [b = 28 / 102]			[b = 81 / 102]
Jennifer Rodriguez	20	30	66
Joe Smith	18	25	67
Mary Lee	100	45	68
Jake Himmel	25	45	68
MET BENCHMARK [b = 28 / 102]			[b = 81 / 102]
Jason Gianakouros	90	64	81
Stephanie Babcock	20	Incomplete	84
Mary Lee	Ø	Ø	89
+ Chris Wren	85	71	102
+ Simon Goldsmith	96	Ø	102
INCOMPLETE [b = 28 / 102]			[b = 81 / 102]
Faith Reynolds	27	50	-
Owen Higgins	22	40	-
ACCESS DENIED [b = 28 / 102]			[b = 81 / 102]
Marcie Fromm	26	Ø	Ø
Gregory Boyers	24	Ø	Ø
NOT ASSESSED [b = 28 / 102]			[b = 81 / 102]
Samuel Driscoll	22	70	-
Jenny Lin	103	90	-

© 2004 Wireless Generation. All rights reserved.

Guide to Class Summary

Class Performance History



Class Summary Key

Scores display in red if the student scores below summed benchmark score.

Fence Sitters

Student has met the summed score Benchmark but did not meet the Benchmark for specific tasks.

+ High benchmark

Some first and second grade students who read above grade level are designated with High Benchmark status and are exempt from future PALS administrations.

Student has multiple PALS Assessment results



Performance Groups

Identified

Students did not meet the benchmark for the current assessment window. These Students require intensive instruction.

Met Benchmark

Students in this section met the benchmark for the current assessment window.

Incomplete

Students in this section have not completed all required tests for this assessment window.

Access Denied

Due to your school's security settings, you do not have permission to see results for students who were assessed by another teacher.

Figure 3. PALS class summary view

The TPRI class summary (see Figure 4) is less explicit and specific about the level of additional instruction that children might need, in that it separates children into just two groups – those who passed the screening and those who didn't – and then lists children within those groups alphabetically, as a default view. (Recall that the screening is designed to identify children unlikely to be at risk, which means that the children who fail the screening may have needs ranging from intensive ones to moderate needs.)

Unlike the other assessment views, however, the TPRI class summary can also be sorted by different criteria. Teachers can sort the data into views that rank order children by fluency score, by the Story Number that indicates their instructional level passage, or their Reading comprehension score. These additional views may help teachers see which of the children who did not pass the screening are farthest behind in what they can read and understand.

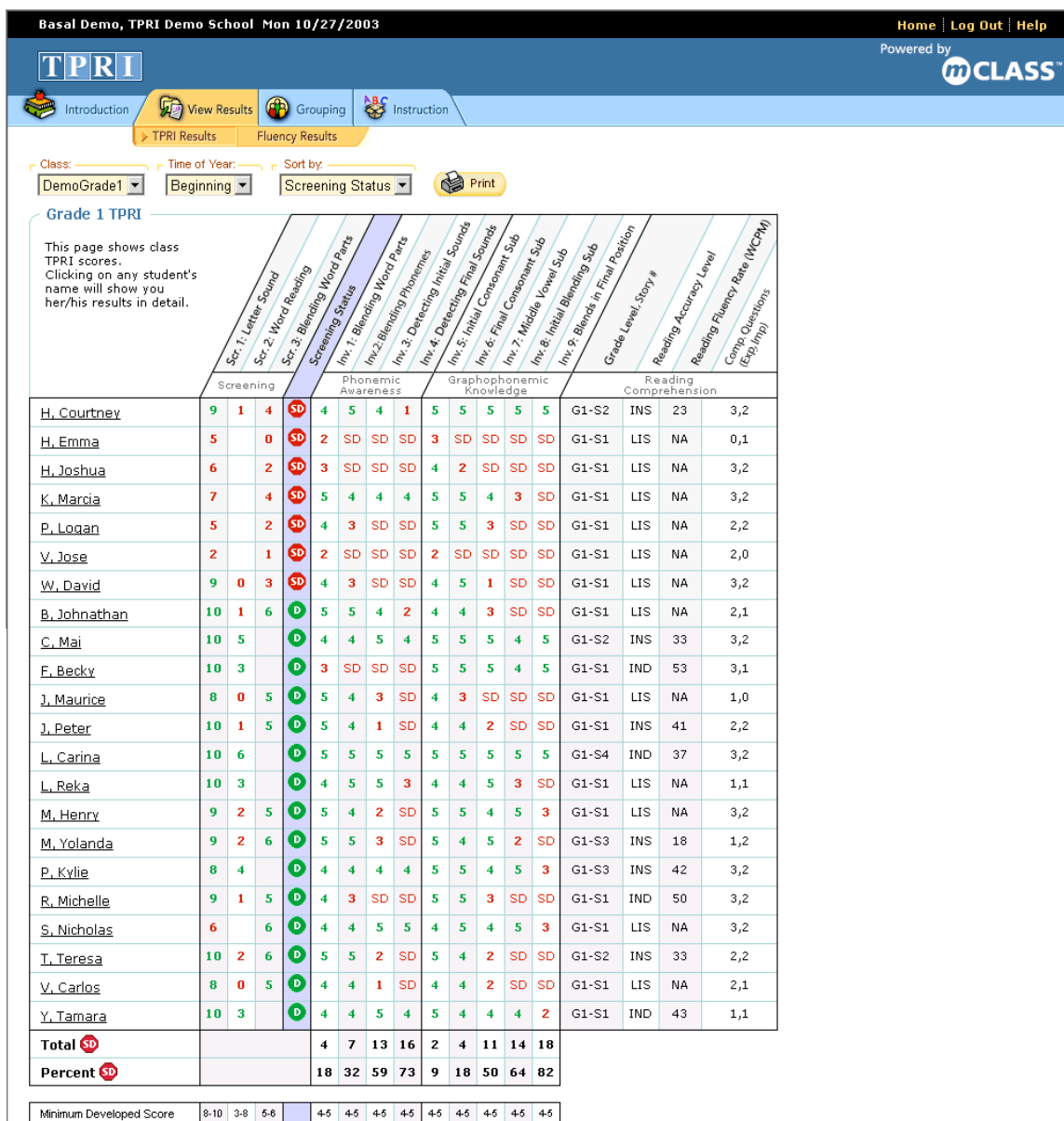
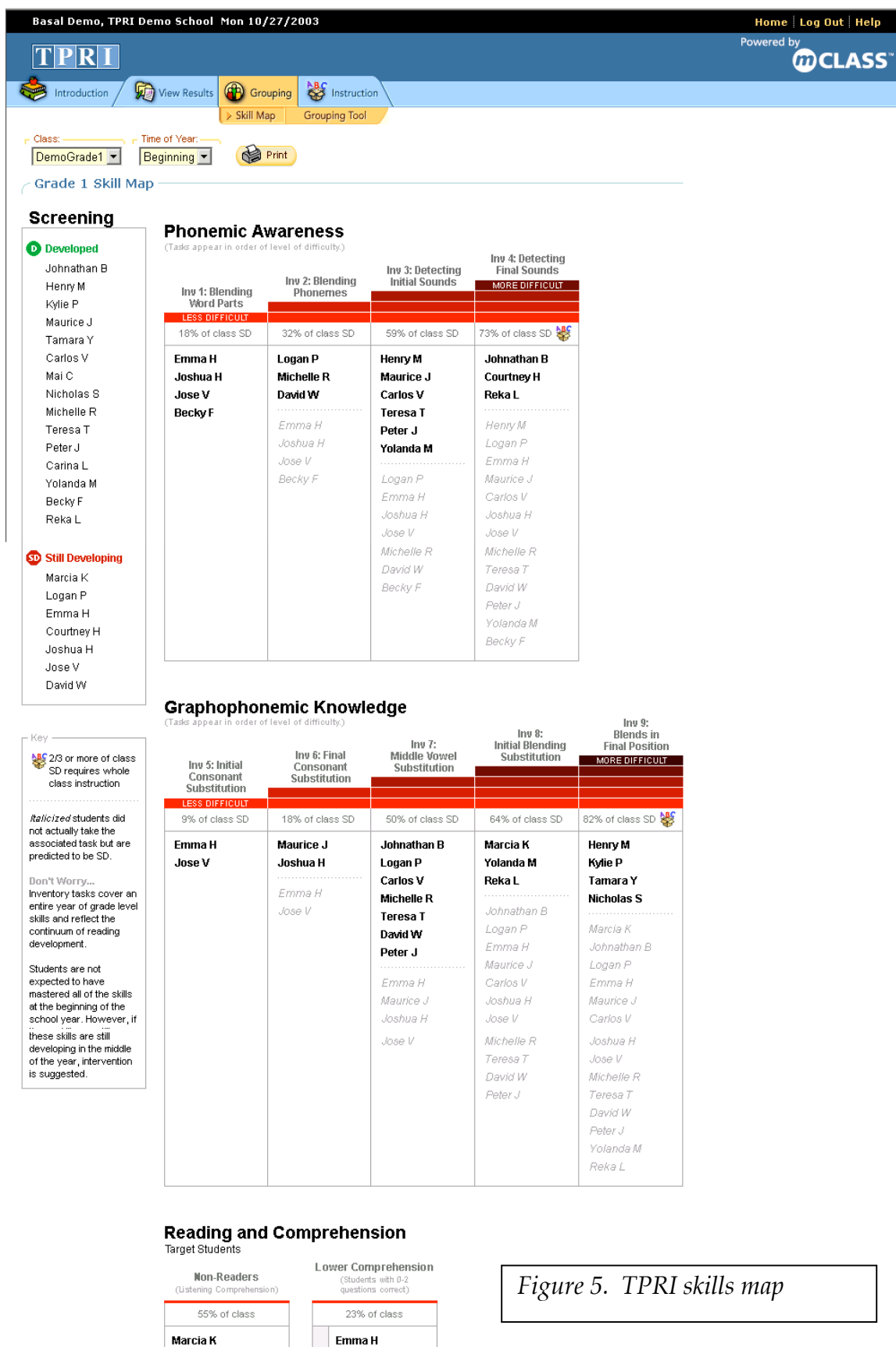


Figure 4. TPRI class summary page

After deciding which students need instruction that takes place in addition to classroom work, the most critical actions for a teacher to take are to (a) assign students to groups to work on a particular skill during class, and (b) decide on the instructional activities to use with that group. In DIBELS, for example, not all “strategic” instruction level children may need the same skill. Even all the children who score as “high risk” in a particular skill area may not need the same subskills. For example, some children who score as “high risk” in alphabetic principle skills (measured by the Nonsense Word Fluency task) may need help knowing which sounds go with which letters; others may know the letter-sound correspondences but need help blending sounds together more fluently. Thus, another question we can ask about visualizations is: How actionable is a particular visualization for making decisions about classroom instructional groupings and the activities to use with those small groups?

TPRI, for example, provides some of its most actionable information for instructional groupings and decisions not in the class summary view, but in an additional view called the skill map (See Figure 5, below). Students are grouped not just according to the overall needs they have (such as needing help with phonemic awareness), but according to the specific subskills within that skill area that teachers should focus on in working with small groups of students.



Naomi Hupert, Center for Children and Technology, points out that a big challenge for teachers, once they have grouped students, is to design small group instruction that integrates skill practice into real reading and writing activities; otherwise, instruction gets reduced to a series of isolated skill exercises. The Reading First data visualizations may, in some cases, initially steer teachers more toward isolated practice activities for small groups needing help in a particular skill. Overcoming this tendency may be a challenge that requires more than a single year's experience with Reading First, notes Hupert. In describing her group's observations of New Mexico teachers in their first year of Reading First (using DIBELS), she says:

“[W]e noticed an absence of contextualization of skill-based work. If teachers were working on segmenting words, like “playful,” there often wasn't a time when the words that were examined individually were used in the context of reading or when the skill of segmenting was integrated into the reading context. There was not the cross over from skill-based work into authentic reading. This raised a red flag for us and is something we tried to keep track of. It's not the fault of the teachers, but a misinterpretation of Reading First

Providing instruction in individual skills can be easier to adopt as an approach than to try and overhaul how you present the reading of a piece of text, trying to integrate every aspect of reading skill into that. This is the first year of this effort, and it could be that it's changing how teachers are doing things, and that takes time. The first step may be to have a greater emphasis on phonics, phonemic awareness, and vocabulary. The more complex issues – comprehension, fluency – may come later, as the first set of skills become more natural to teachers in their teaching process. It's hard to say and may be too early to tell.”

Moreover, Marguerite Held, University of Texas-Houston, stresses that teachers may need to know more than just the fact that a group of students needs help on a skill such as blending phonemes; they need to know about the pre-requisite skills involved. She gives this example:

After we talk about the grouping, this is a real test for them [the teachers]: to look at that data and translate it into some type of instruction. You'll say, “This child is having difficulty blending the sounds together. What type of instruction does this child need?” ... You can really tell if a teacher is on your wavelength, because sometimes a teacher will want to go all the way back to rhyming, instead of picking up just a little below that point [of blending]. I have found that so many times – they'll go back

to something that's real comfortable. When you talk about prerequisite skills, you don't have to go all the way back to the beginning.

Finally, PALS visualizations provide one of the most actionable features for the question of what to teach an individual student in the area of spelling. The data report screen for a student's individual responses to spelling test items contains a sidebar that states – in highly actionable and explicit language – “Teach these features....” The sidebar then points to particular spelling features (e.g., digraphs, blends, nasals, CVCe) that the student used but confused, suggesting that these features lie in the student's ideal learning zone. The only thing missing is a button that would list the other students who need to be taught those same features, helping the teacher to take direct action in forming a group and developing spelling lessons.

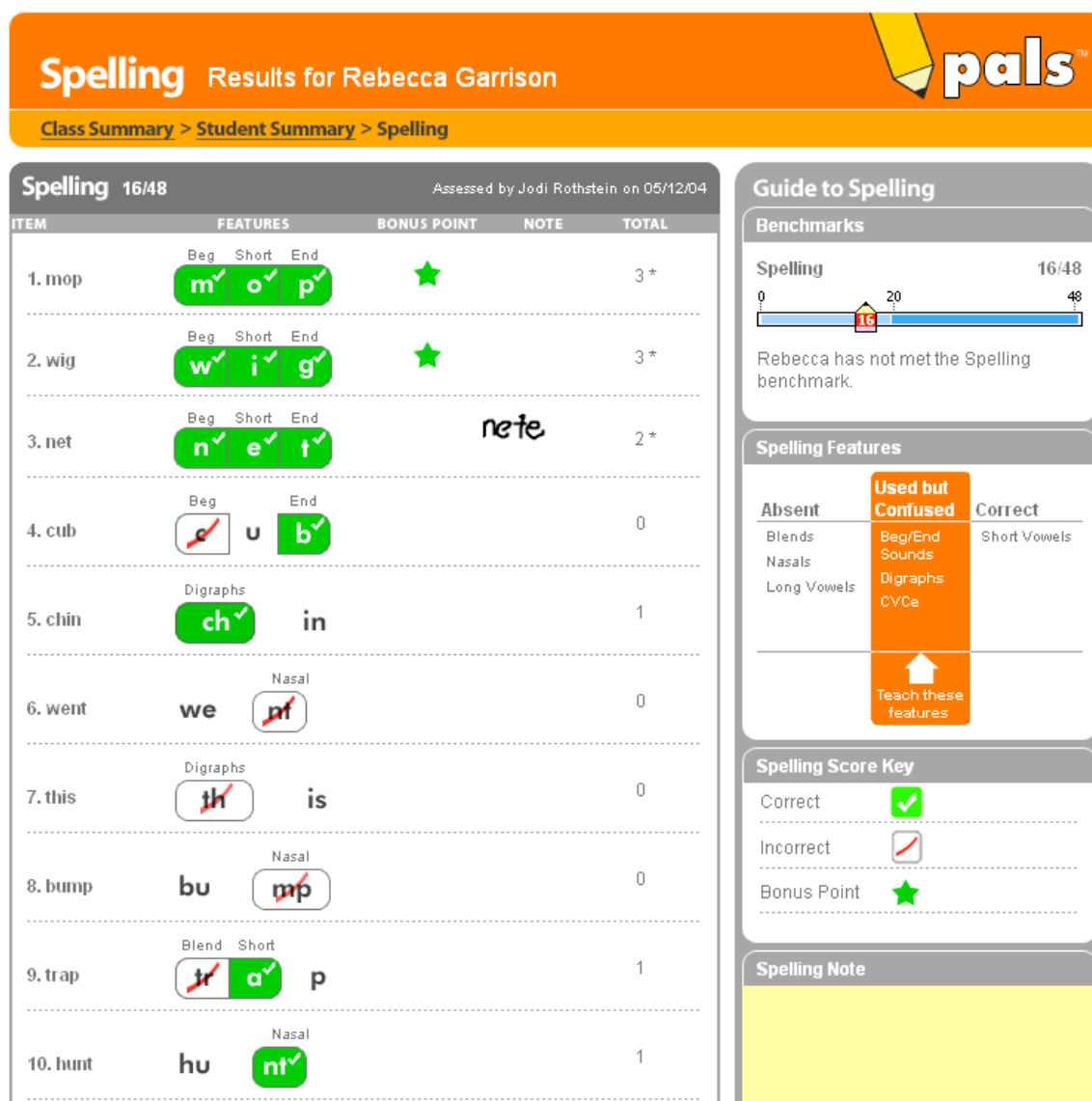


Figure 6. PALS item level details, spelling

Mary Fowler, University of Virginia, also notes that PALS training includes discussions about multiple types of groupings, not just homogenous groups of children who share the same needs. The link from data to instruction is certainly clearer for homogenous groups, but that may not be a reason to exclude other grouping possibilities (e.g., heterogeneous groups, high level/low level pairs) from the data-driven conversations that teachers have in planning their instruction. Future research is needed in this area.

Finally, DIBELS' progress-monitoring graphs provide a clear visual signal for teachers about when instructional changes needed; these graphs are actionable in suggesting that a teacher change a student's instructional supports, but they do not explicitly tell the teacher what the change should be. As shown in Figure 7, these graphs mark student progress toward a goal, with the slope of the student's progress line showing whether the student will reach the goal on time. If the slope of the progress line is inadequate, DIBELS teachers are encouraged to change the students' instructional supports.

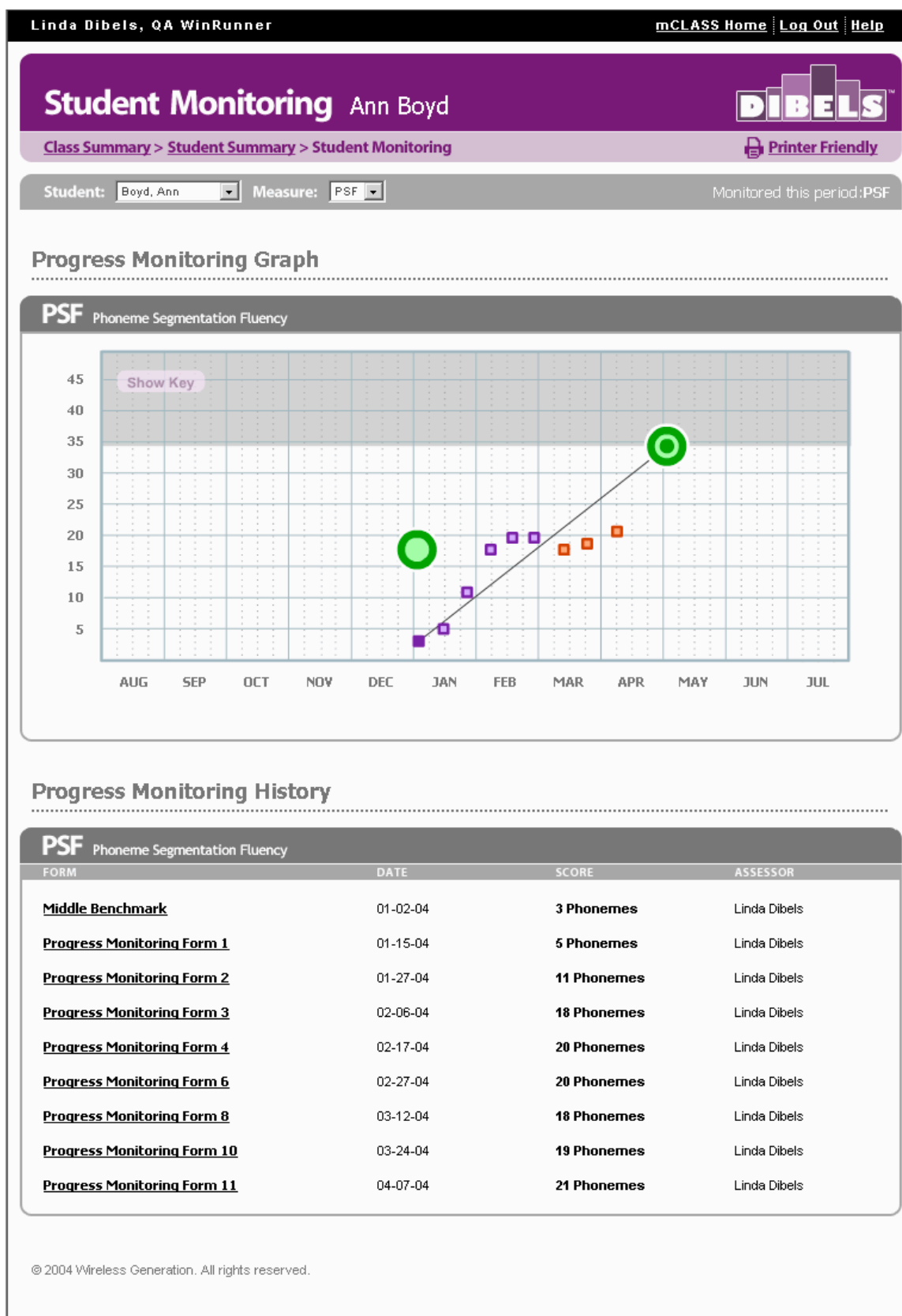


Figure 7. DIBELS progress monitoring graph

4.03 Could Data Visualizations Increase Teachers' Knowledge of Literacy Development?

There is little known about how data visualizations may impact teachers' overall understanding of literacy development, but certain visualizations have features that incorporate representations of literacy models in ways that may support this knowledge. The TPRI skills-map makes explicit to some degree a model of skill development, since the skills on the inventory for phonemic awareness and grapho-phonemic knowledge are presented in the order that children tend to develop them. The visualizations for the STEP assessment, not yet used in Reading First schools, may go the farthest in attempting to explicitly link student data displays to a model of literacy development. STEPS visualizations superimpose student progress on the progression of literacy steps in its developmental model. Figure 8, below, shows an example display of a teacher's classroom, with students grouped according to the STEP model.

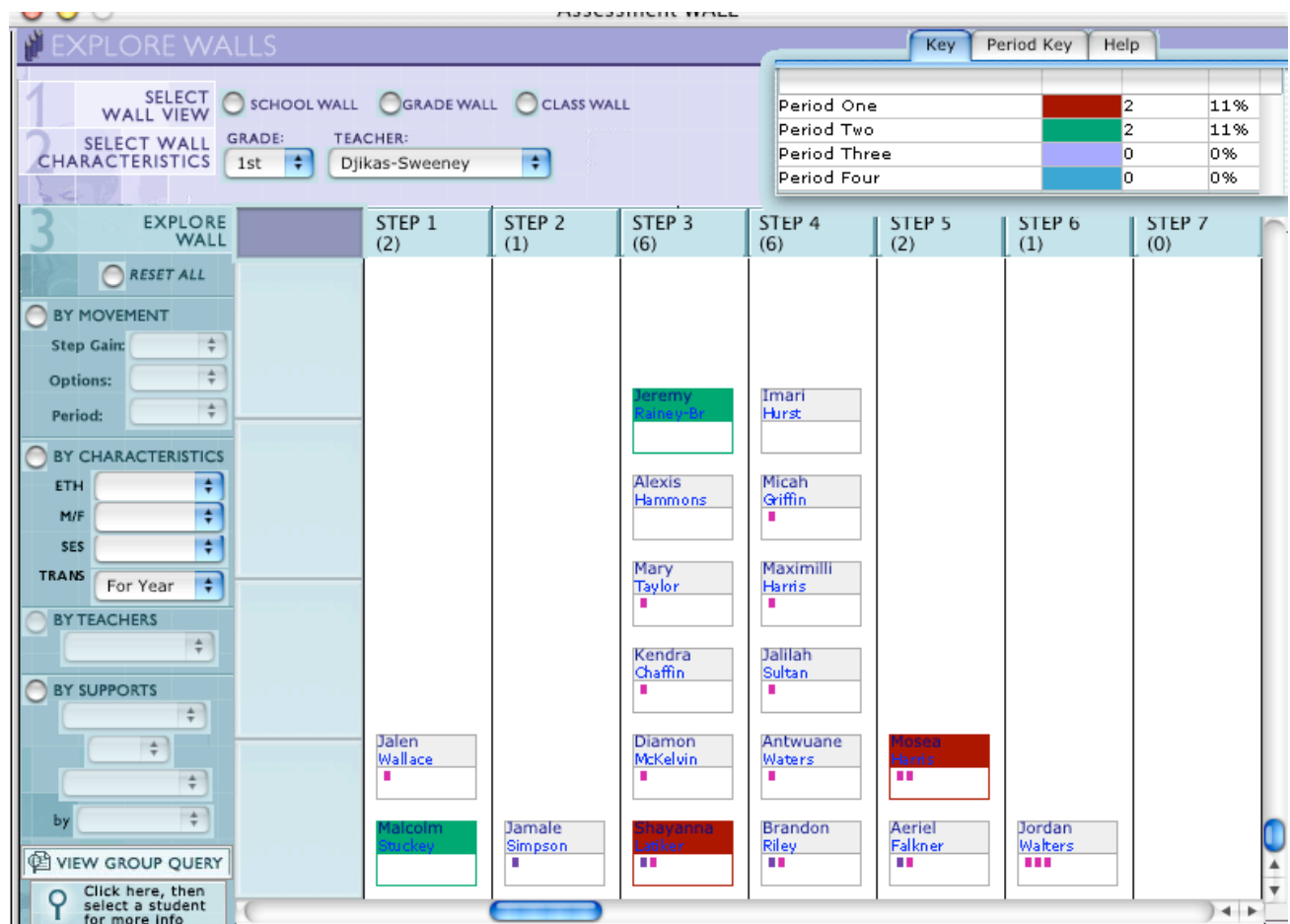


Figure 8. STEptool classroom view

As with DIBELS, these STEP groupings in and of themselves don't necessarily mean that all students at the same step have the same needs; teachers need to go other views for more specific information about the needs of students at a particular step. The manual is explicit in this regard, stating (p. 61), "Simply knowing the STEP level is not enough to inform and provide appropriate instruction."

4.04 Which Visualizations Are Problematic?

Due to the early stages of the field in supporting teachers' conversations around data visualizations, it's too early to know which visualizations or features of visualizations will ultimately be judged as problematic. Researchers have encountered some difficulties but can only speculate on this question.

Alysia Roehrig, Florida State University, notes, "We're teaching about DIBELS in an online course [about measurement] for middle school teachers. Even board-certified teachers really struggled with the box and whiskers graph." [Note: This type of

graph appears on the Florida Progress Monitoring and Reporting Network, not on the Wireless Generation visualizations.]

Marguerite Held, University of Texas-Houston, reports that sometimes the labels for skills in the TPRI displays don't resonate with teachers. She explains, "When they see 'blending word parts,' [teachers ask], 'What does that look like instructionally?' We have to make sure that teachers understand the terms on the assessment [visualization] and what [they] look like."

Mary Fowler, University of Virginia, states that benchmark changes across difficulty levels in word lists can cause problems:

Once when teachers were looking at the growth of the children, it had their benchmark score in the fall, which had to be, say 30. And in the spring, their benchmark score had to be 25. Logically that doesn't seem to make sense. But at the beginning of the year, [children] had to read the pre-primer word list, and at the end of the year they had to read the first grade list... the task at the end of the year was harder. So the benchmark ended up being lower, but for a lot of people that was confusing. It was hard to see that there was growth there, if the number was going down.

As is evident from this section, there is general consensus among the sources used for this report about what makes a data-driven conversation "good." Conversations are good if teachers are (a) using the data to assign children to extra support structures, such as one-on-one tutoring, (b) using the data to come up with reasons for forming classroom instructional groups, and (c) making a link between the data and instruction to plan appropriate activities for each of those groups. But are there other, deeper assumptions about the nature of "good" instructional conversations? The next major section explores this question.

Section 5. Explicit Deep Assumptions

A deep understanding about how to support teachers' data-driven instructional conversations may require that we first make explicit our deep assumptions about the nature of "good" instructional conversations. Some of these assumptions may cut across conversations about different assessments, embedded there by the developers who created the assessments, or modified by the Reading First sites who are using them. Other assumptions may be more closely tied to the design of a particular assessment; in some cases, different assessments may have underlying deep assumptions that conflict with each other.

A key principle of the Information Infrastructure Project is that deep assumptions matter; that educational change often succeeds or fails by what lies beneath the surface. Like an iceberg, there is often much more beneath the surface of educational reform efforts than is visible from a bird's-eye-view. With this principle in mind, this section will dive into a description of the deep assumptions that may underlie one or more of the Reading First assessments plus the STEP assessment.

Naturally, the developers and key supporters of each assessment are the people who should be most involved in making explicit the deep assumptions about their assessment. It would be inappropriate for this report to attempt an outsider's list of assumptions for each key assessment. Instead, this section will be driven by a preliminary list of deep assumptions about the STEP assessment, a list created through conversations with David Kerbow, the lead developer of STEP. This section will also attempt a preliminary analysis of whether these assumptions may or may not be shared across other Reading First assessments, based on initial conversations with members of the Reading First community. The section will end with preliminary suggestions about a few deep assumptions that may be more closely tied to specific Reading First assessments than to others or to the STEP assessment. Ultimately, however, this section can only serve as a starting point, with suggestions to the field for looking deeper at assessments and uncovering our assumptions about teachers' conversations that matter for students.

5.01 STEP Assumptions

The following list of STEP assumptions is loosely organized by how widely each assumption may be shared by other key assessments. Widely-shared assumptions begin the list; assumptions that are more specific to STEP will end it. Sources involved with TPRI and PALS research discussed some of these assumptions explicitly in interviews for this report, but such discussions did not take place with DIBELS sources due to schedule conflicts; hence, examples related to DIBELS are fewer and more speculative.

- *Conversations should include discussions of children's strengths, not just weaknesses.*

This assumption is embodied in the design of each STEP level: each level is defined by what a child can do, not by what the child cannot do. For example, children at STEP 3 (end of kindergarten/beginning of first grade) can:

- Hear and record many initial and ending sounds in words.
- Use a core of high-frequency words and some word families.
- Add up clues from text and pictures to make sense of and talk about stories.

By referencing a child's STEP level in a conversation, teachers should be referencing the child's strengths. Their knowledge of the next STEP level should invite conversations about where the child should be moving; failure on a STEP test should also lead to conversations about skills to focus on.

This assumption appears to be shared widely across the other key assessments. For example, Kristi Santi, University of Texas-Houston, and Mary Fowler, University of Virginia, reported that TPRI and PALS support this assumption. Marguerite Held, University of Texas-Houston, notes that teachers tend to focus on students' weaknesses, but TPRI mentors work to point out students' strengths shown in the assessment data. She adds:

It's a catch-22, because you do want them [the teachers] to see [students'] strengths. At the same time, when you form your groups, you're thinking about needs that [students] have...we do try to get [teachers] to realize [students] have strengths. For example, if a child isn't reading yet but has good listening comprehension – well, that's a strength. Once you get those tools together to read, at least you know he should be able to understand. Whereas if a child doesn't have any reading skills and

also has poor listening comprehension, you think – ‘Oh, you’ve got a lot of work to do.’

- *Conversations should include discussions of children’s continual movement forward in their literacy development, not just their attainment of grade-appropriate skill levels.*

STEP visualizations explicitly highlight children’s continued movement in their literacy development, even after they have attained grade level goals. In the earlier Figure 8, for example, teachers can immediately see, via the “dots” under a child’s name, whether children have stagnated in their literacy development – even when they are performing at a grade-appropriate level. Each dot represents the number of Steps that a child has moved from the beginning of the year.

STEP also reinforces the importance of movement by specifying multiple goal steps for each grade, not a single end-of-grade-level goal. Moreover, teachers using STEP can administer any STEP level to any child, making it possible to look closely and continually at the literacy skills of high-performing children who already meet grade-level goals early in the year.

PALS developers are working to add material that will allow teachers to measure growth in high-performing children and check for movement. The assessment stops at third grade, but developers are adding fourth and fifth-grade level material appropriate for high-performing third graders. Mary Fowler also notes that PALS supporters recognize the dangers of focusing only on the growth of the lowest-scoring children. She says:

One of the things that can happen with PALS is that, say, six [children] get identified and get intervention all year long. Five students, say, didn’t qualify for intervention, and then they end up looking lower than the kids who got the intervention. That’s what’s good about the mid-year [assessment], is that those kids can then feed into the intervention program before it’s all the way at the end of the year, and you realize those kids weren’t moving forward. It definitely comes up, and people talk about it.

Kristi Santi, University of Texas-Houston, also agrees with the importance of this assumption and describes two ways that TPRI addresses it. First, TPRI now includes the

Fluency Kit for assessments every six weeks, which can help teachers see if students are stagnating. She notes:

[You] have to look at kids who are lagging behind and make sure they're making progress. But we also stress that for kids who are in what we call the "most independent" group, that they're continually making progress – which is why we talk about continuing to measure fluency and other assessments, because you don't want those kids to go into a holding pattern.

Second, Santi notes that TPRI now provides an option for teachers to begin administering fluency assessments to high-performing first graders at the very beginning of the year, to more closely monitor their growth. She reports:

[In the beginning of first grade], we expect a lot of kids will be at the listening comprehension [stage], so it doesn't make sense to go to the Fluency Kit right then...But we had a lot of teachers ask, "What if my students are reading?" So we put that provision in there, so you can see if they're making progress.

DIBELS does not include standard provisions for looking at fluency of high-performing first graders, nor does it include visualizations that focus attention on movement of children who are not currently at risk. However, Naomi Hupert of CCT cautions against assuming that DIBELS will encourage teachers to ignore the progress of high-performing students, to the detriment of their literacy growth. She states:

In taking a few steps back and looking at NCLB and Reading First, it's easy to say that [teachers will ignore high-performing students' growth or lack of growth], because all you need to do is focus on the kids who are performing most poorly, and you're done. It doesn't matter what the kids at the middle or top are doing – they could be bored out of their mind as long as the others are meeting benchmark.

But of course the other way to look at it is that any effort to help teachers use data will help all students – if [teachers] are provided with adequate and useful professional development to help them understand what they're seeing. You need to differentiate your instruction, have groups, then you start to have kids with similar skills, and you may have kids who don't need instruction in [basic] reading. You can assign them something more challenging, because you've already taken the first step of identifying what they're capable of doing. Prior to using any assessment, teachers might have a sense that these kids are doing well, but it may not have been articulated or spelled out. It could go both ways – I don't know if there's enough information out there.

Alysia Roehrig, Florida State University, makes a similar point:

[Teachers think] these [low scorers] are the kids [they] need to help most, so all [their] energy goes there. They're triaging. It's such a tricky balancing act that a lot of teachers struggle with that, and I wouldn't want to blame DIBELS.

- *Conversations should include a strong tie between what the child does on the assessment and what the child and teacher do in normal classroom activities (reading and writing).*

Several of the STEP activities were specifically created to mirror what happens in the classroom. Most notably, the book reading assessment and comprehension conversation should be highly similar to teacher-student interactions around books. David Kerbow notes that teachers at the second and third-grade levels, particularly, should use data from this portion of the assessment to support conversations about how the child is a problem-solver, how a child elaborates answers, and about the information that a child brings to the text.

Similarly, the Developmental spelling tasks link to classroom writing activities. STEP also has a clear mapping from the book assessment portion to classroom leveled books. Kerbow notes:

We settled on the Guided Reading levels of Fountas and Pinnell, and we have maps to take them to Reading Recovery levels or to broad grade-level applications . . . Teachers most definitely bring in these reading levels and use the overall assessment to make decisions about the books to use in their instruction. There are direct links – STEP 6 book is Guided Reading level I. You may work with a step lower or higher with those students, but there are direct links to what book you'll use in small groups.

DIBELS, in contrast, has sub-tests that appear to be designed more as measures of underlying skills, rather than as mirrors of classroom activities. The Nonsense Word Fluency measure, for example, was discussed earlier in this report as a measure of alphabetic principle skill. It is specifically not meant to infuse nonsense-word reading into classroom activities, but it is sometimes misunderstood by teachers as implying that nonsense word-reading is a valuable instructional practice. Also, the DIBELS oral reading passages, as noted earlier, measure end-of-year text level reading, while TPRI

and PALS attempt to place children in a level of text that is similar to their current instructional level.

PALS is the most similar to STEP in providing links from the scores on the oral reading measure to leveled classroom books. TPRI does not make such explicit links to leveled books (outside of basals), but Marguerite Held, University of Texas-Houston, notes that charges of a conflict between TPRI and leveled books (particularly in Guided Reading lessons) are undeserved:

Some people in the Guided Reading world see us coming and say, “Oh, they won’t like this” – and that’s not true. I worked with a group that did Guided Reading, and we tried to rework their lesson design a little bit, to get them to implement a little more of the word work into the lessons – where they do some word sorts and things like that to focus on [for example] the ‘long a’ sound spelled ‘ai.’ So they’ll have a little more opportunity to practice that....

If a student is reading in Story 1 [on TPRI], he’s reading at text appropriate for the beginning of grade one. If he’s reading at story 5, that’s text that’s more appropriate for the beginning of second grade. And so story 2 is a little harder, story 3 middle of the year, story 4 moving toward end of first grade. I tell [teachers] that, and then they have to make that correlation to their leveled text.”

Though TPRI may not mirror some classroom activities as closely as STEP, it has other ways of making a strong link between the data and instruction, and these will be further discussed later in this section.

- *Conversations should center on the child as a meaning-maker.*

STEP developer David Kerbow emphasizes that good conversations about STEP data should emerge from and return to the central question, “How does this child make meaning from text?” even when teachers are discussing data about isolated skills. He explains:

When teachers talk together, in the best of conversations, they talk about the entire Step, not just the isolated skill. For example, at Step 6, they would be talking about fluency and phrasing, how students are understanding character and perspective at that level, and where their development spelling is going. They should see students using problem-solving for words in their writing and in their reading. Since we’re not just talking about a student’s reading rate and phonemic awareness, [STEP] tends to encourage teachers to talk about how [skills] fit together.

Kerbow notes that this focus on meaning-making and STEP's heavy emphasis on comprehension measures happened as a result of dissatisfaction with measures of reading rate and accuracy as comprehension predictors:

With STEP, we started with reading levels and running records and tried to analyze the students' strategies. We found that at 2nd grade, we had students who were reading at the top text levels – but then when we compared their standardized scores on ITBS, there was enormous variance. These students, who could read text on a 5th grade level, were scoring between the 25th and 90th percentile [on ITBS]. The argument that reading level and rate is sufficient is a weak one, and this is especially true at 2nd and 3rd grade. The variance of comprehension scores really widens for a given level of text reading.

Mary Fowler, University of Virginia, agrees that this assumption is important for and consistent with PALS. Kristi Santi, University of Texas-Houston, suspects that teachers using TPRI may vary in how much emphasis they place on meaning-making within their discussions. She says, “I think that for TPRI, it would really depend on the knowledge level of the teachers as to whether you would get that type of conversation.” She gives an example of how coaches for TPRI bring up the topic of integrating skills into larger activities:

One of the questions we get is: What happens if a child finishes an activity before group time is over? This is where say that if you're doing a word-sort activity, the kids can move on to make a sentence out of the words, and then a story. We try to get it so that you're not just thinking of activities as isolated skills, reading in isolation, but turning them into extension activities, where you're incorporating [words] into writing or skill maps or conversations.

As noted in the previous section, Naomi Hupert, of CCT, reports that her group is watching closely to see if New Mexico teachers using DIBELS will be successful in integrating skills into meaningful reading and writing activities in their instruction, rather than tending toward isolated skill practice.

- *Cross-grade conversations are valuable.*

By working in small schools – some with one classroom per grade – where cross-grade conversations were the only ones possible, STEP developers discovered that cross-

grade conversations can enable teachers to share ideas about supporting students across a wide range of literacy development. Through sharing these ideas, teachers may also become more knowledgeable about the skills that teachers of higher grades view as key for success and the skills that teachers in lower grades have focused on.

The twelve-step literacy framework in the STEP assessment provides a common anchor for discussions about children's literacy development across grades K – 3. All teachers share the same set of assessment materials, further supporting cross-grade discussions.

Kristi Santi, University of Texas-Houston, acknowledges the value that could come from cross-grade discussions, but reports that schedule constraints in larger schools can be a problem:

In the IERI study, we started meeting with all teachers at the same time to talk about TPRI in general, and timewise it didn't work out well. I think it's a great idea for a first grade teacher to know what's on the kindergarten assessment and how kids did, or for a first grade teacher to know how to administer the kindergarten inventory with a child. To see the continuum of skills and where children fall. We haven't been very successful at it, mainly because it's hard to get a time when the K, 1, and 2 teachers can get together.

Mary Fowler, University of Virginia, concurs and says that to her knowledge, cross-grade conversations in Virginia PALS schools are not common, but could help teachers learn more about their students. She notes that most information about a child's previous year's experience gets "stuck in a folder and unless there's specific concern about a child, most teachers don't go back to the other grade teacher to discuss [their students]."

DIBELS developers have written about the importance of teacher conversations around data, but they appear to stress within-grade, not cross-grade, conversations. Good, Kaminski, Smith, Simmons, Kame'enui, and Wallin (2003), for example, write, "(p. 224) ...change is enhanced through *grade-level* discussion of teacher reports in which collegial relationships and necessary support systems are created to help make instructional and programmatic decisions based on data." [italics added]

- *Conversations should include discussions that look at socio-cultural sub-groups in the school.*

STEP researchers have created visualization tools that allow teachers to easily question how the school's reading program is meeting the needs of socio-cultural sub-groups defined by ethnicity, gender, SES level, time in the school, and participation in specific interventions, such as Reading Recovery. Teachers can easily view data that highlights the performance of these sub-groups within their own classrooms.

Administrators or groups of teachers and administrators can view the performance of these sub-groups across grades or across the school.

These tools appear to be more prominent for visualizations of STEP data than for data from the other key assessments reviewed in this report. None of the sources interviewed knew of schools where teachers looked specifically at the performance of these sub-groups on Reading First assessments.

5.02 DIBELS Assumptions

A review of the DIBELS assessments and the developers' writings suggests three possible deep assumptions that were not uncovered in the review of STEP.

- *Conversations should include an element of urgency.*

DIBELS materials stress the high odds for failure that low-performing students face, unless they receive urgent, additional instruction. With its progress-monitoring materials, DIBELS – more than any other assessment – allows teachers to change course quickly if a low-performing student's progress is not rapidly observed in an area of concern. The “risk” language used to label performance also appears to suggest that teachers should focus their discussions on identifying the students most urgently in need of attention and asking, “How can we address these urgent needs and diligently monitor progress?”

None of the other assessments appear to convey urgency to the same degree as DIBELS. The TPRI reports, for example, label low-performing children as “Still developing,” but that label is not as explicit about whether or how much a child is in trouble with his or her reading development. PALS reports show teachers which

students are “identified” as needing intensive support, but this is also less urgent language, and Mary Fowler, University of Virginia notes, “I just don’t know if [PALS] people would use that term [“urgent”].” Further, as noted earlier, neither TPRI nor PALS suggests weekly testing for very low-scoring children, to suggest that instructional decisions and changes need to happen very quickly.

David Kerbow, STEP developer, notes, “We’ve begun to identify places at the beginning, middle, and end of the year when a student’s performance warrants immediate attention. It’s not the clean rule that DIBELS brings to the table, but it serves the same purpose.” He does not subscribe, however, to the belief in weekly testing for any student:

We don’t encourage that. With STEP, we identify weaknesses, where the student is struggling, and then you target your instruction and you observe during classroom work. Even if we create more forms, I wouldn’t encourage assessments every week or every two weeks.”

• *Conversations should include discussions about a student's rate of progress, not just progress alone.*

Related to the notion of urgency, DIBELS materials stress the importance of reaching goals on time. Its progress-monitoring graphs (see Figure 7) help teachers visualize whether a student's rate of progress is adequate. Teachers are often pleased – as they should be – when lower-skilled students make progress of any kind, but the DIBELS focus on rate of improvement suggests that it is important for their conversations to explore whether this progress is fast enough to keep students from falling further behind peers.

Marguerite Held, University of Texas-Houston, agrees that conversations about rate are important, and she points out that the Fluency Kit of TPRI can be helpful for looking at students’ progress in oral reading. She also notes that it can be difficult to stress the importance of rate to teachers:

[Teachers] will say, “ This student is a really good reader, and I don’t understand why you’re concerned that at the end of first grade he’s reading 30 words correct per minute. He can answer every question.” We talk about increasing that fluency rate, and that yes, he’s making progress, but he needs to be making more progress in this area and here are some things you can do. I think it’s a real problem . . . particularly in 2nd grade, where we see the issue more. We talk about how these children need to read longer texts, to read faster.... I was working with a first grade student

in a Guided Reading book, and he finally read the sentence, and he was so excited that he called his mom. But, my goodness, he was at the end of first grade reading at the level of beginning of first grade, so he had a long way to go. It's a hard situation. You don't want to beat teachers down, but at the same time, you have to be realistic.

When asked if rate comes up in many PALS conversations, Mary Fowler, University of Virginia, suspects, "Probably not, because doing the constant assessment is not a key part of PALS; it's more [about] doing the assessment over a long period of time. Teachers might say, Oh, Suzy's made tremendous progress, but Joey hasn't moved. So it may come up but not as much."

- *Conversations should include discussions by a team about how the school is meeting needs compared to previous years and compared to other schools.*

In a recent chapter (Good, Kaminski, Smith, Simmons, Kame'enui, & Wallin, 2003), DIBELS developers provide data from over 400 schools on kindergarten students' performance on the DIBELS assessments. Schools can use these tables to compare how well they are doing in addressing the needs of students who score at different levels during the year. Schools at the 95th percentile, for example, help 81% of their lowest students – those whose scores at the middle of the year suggest they need intensive instruction on phonemic awareness – to reach the end-of-kindergarten DIBELS goal for phonemic awareness. Schools at the 25th percentile and below help only 7% or fewer of such students to reach that end-of-kindergarten goal. In addition to these reference tables, the chapter provides templates for guided reports that schools can complete at the beginning, middle, and end of the year. These guided reports list questions that the school teams should ask in looking at their DIBELS data and comparing it to previous year's data and data from other schools.

No other assessment reviewed for this report appeared to have the extensive available data or the guided report structures already in place for school teams to engage in this kind of conversation.

5.03 TPRI Assumption

More than any other assessment, TPRI appears to have this underlying deep assumption:

- *Good instructional conversations should be supported by providing extremely explicit links between the data and instructional materials the teacher already has.*

As noted, TPRI developers are exploring the use of handheld tools and website information for linking students' specific word-reading errors (on word lists and passages) to activities and lessons in the basal materials that all Texas teachers have readily available.

These links are in addition to the already-available Intervention Activities Guide. Kristi Santi, University of Texas-Houston, notes that TPRI trainers stress ways that teachers can use existing materials for activities in the guide. She explains:

When you plan the lesson, some of the activities in the Intervention Guide require that you have or make materials. So we really work with the teacher to show them that they can pull materials that come with their curricula, so they're not having to invent activities or make materials. [For example], say the activity is Building Words, and you're using the Open Court basal. Use the mini-cards from Open Court, and use them in a center activity, so kids can build words. . . . The bottom line, we feel, is that if you want teachers to use the results to inform instruction, you have to make it the least time-consuming event that you can, and that includes teachers not having to make materials, and using things they already have.

5.04 PALS Assumption

Mary Fowler, University of Virginia, suggests that some PALS conversations may have this underlying assumption:

- *Good instructional conversations should help teachers use their data to navigate between basals and other reading materials.*

Fowler notes:

One of the big things about Reading First is that you need to be using assessment to drive instruction. At the same time, [the state] says, "You need to use your basal." Well, if your kids are not reading at a 2nd grade level, the 2nd grade basal isn't going to be appropriate. It's a big contradiction. What I tell people is, if you've adopted this basal, then put it in a bookroom, so that in any classroom you might have five first-grade basals, five second-grade basals, and so on. That way, you can meet the needs of your kids. And that works for some schools. For others they say, "No way, you're in the second grade, you're using the second-grade basal. . . I used to hear crazy things like, they have all these wonderful little books and trade books in their classroom, and now they have the basal, so

they've put all [the other books] in storage, and they're not using them anymore. And [we tell them], "No, no, no, that's one other resource; it's not one or the other.

Section 6. Environmental Supports for Conversations

The previous three sections have illuminated some differences about the key assessments, their accompanying data visualizations, and assumptions about the nature of good data-driven conversations. The good news is that despite these differences, there is growing consensus about how to create environments that support teachers' social practices around assessment.

6.01 Links to Lessons that Precede Reading First

Some of this knowledge draws upon work outside of the Reading First community. Daniel Light, from the Center for Children and Technology, notes that many districts have been working through issues of data-driven decision-making for years, using a variety of assessment and analyses tools ranging from the GROW Network reporting system for high-stakes standardized tests (see <http://www.grownetwork.com/>) to homegrown data warehouse systems. Light reports that CCT's work with one such district (District 10) in Florida has underscored many critical components for successful data-driven conversations, including:

- *A focus on student learning combined with a high level of confidence between teachers and administration that the data will not be used in a punitive way.*

Mary Fowler, University of Virginia, agrees:

One of the things that's often misinterpreted is the whole purpose of our PALS assessment. It's designed to inform your instruction; it's designed to identify kids who need extra help; it's *not* designed to evaluate whether you're a good or bad teacher. It's important for teachers to understand that the goal is helping the children.

Fowler also notes that teachers' fears about sharing data often extend from the possibilities of formal consequences from administrators to informal judgments by other teachers:

Sometimes with their assessments they get a little [worried]: 'Oh my kids aren't as good as so-and-so's.' [You need to create] an environment where everyone's not going to be judged.

- *Confidence that there will be resources in place to support what's uncovered through the conversations.*

When questions, findings, or perspectives emerge as teachers discuss the data, there needs to be a response from the administration. If, for example, the teachers find that they need more leveled reading materials, or need professional development for working with ESL children, the administration needs to find ways to fill those needs. Otherwise, teachers can grow frustrated, believing that they have found a problem and discovered what they need to do, but have no one to support the next steps.

- *Professional development to strengthen teachers' assessment literacy.*

Light clarifies that teachers do not need to be psychometricians, but many need more information about how to use data for decisions. Alysia Roehrig, Florida State University, notes that her group is developing such a course for the state that teachers will take for their certifications. "If teachers can understand measurement, we can empower them to understand what a test means and why they do a test a certain way."

- *Professional development that goes beyond general instructional strategies and includes specific classroom practices.*

Light explains, "Teachers might hear, for example, 'You need Differentiated Instruction.' Well yeah, thanks. What does that mean? What resources do I have to make that happen? Do I differentiate by grouping kids or by differentiating within whole class activities?"

6.02 Other Important Supports

Mary Fowler, University of Virginia, adds these components to the list of important supports, based on experience with PALS implementations:

- *Built-in time for conversations.*

Fowler notes, "[When it doesn't happen], the reason it doesn't happen is that no one has ever set aside time for that conversation. They have so much going on that unless they make it a priority, it won't happen."

- *On-site data and conversation coaches.*

“[You need to have] someone who’s knowledgeable about the assessment and interpreting it, being there and helping to guide any questions,” says Fowler.

University of Texas-Houston researchers are exploring the feasibility of TPRI coaches that travel to different sites, and they are already learning about what does and doesn’t work. Marguerite Held reports:

Last year was our first year for mentoring, and we unfortunately had a really heavy case load...eight or nine schools per mentor, and you couldn’t get around quickly enough. You’d work with a teacher and couldn’t get back until two weeks later... I call it, ‘drive-by mentoring.’...It’s not as effective as if you could get by every week....This year (2004-2005) we’re having fewer schools and teachers [per mentor], and we’ll be able to get more in depth about what’s in that Intervention Activities Guide, making connections to each piece and to the big picture, which of course is reading.

Held suspects that the ideal caseload for traveling mentors may be about four schools.

Another component that may be critical for good instructional conversations sometimes lies in tension with goals for reliable and valid testing, as well as with time constraints on teachers:

- *Direct involvement by the teachers with all phases of the assessment process.*

In some Reading First schools, for example, SWAT teams (School-Wide Assessment Teams) administer the tests, score the tests, enter the data, and then deliver reports to teachers. This is not the preferred operating procedure, according to Roland Good, co-developer of DIBELS, who states (see <http://www.ncrel.org/rf/good2.htm>):

I’d like to see teachers being directly involved in that assessment process. If they are directly involved, the assessment takes on additional meaning, additional clarity, additional vividness to them. If you get a score of 37, and you don’t know what that means as you sit with a child, then it’s just a 37. But if you know what that looks like when a child does it, it takes on additional meaning that links to what you can do to teach core skills.

Good clarifies, however, that this does not mean that each teacher works in complete isolation. He notes, “But I would also like to see teachers not take over – completely – ownership of being involved, but to give that away and to share it with the team. So I’d like teacher direct involvement, but I’d like a shared involvement as well.”

Alysia Roehrig, Florida State University, also suggests that it may be important to re-think the roles of coaches and teachers in some schools. She says:

In our schools the coaches score and enter data. They said you don’t really appreciate the data unless you’ve “dibbled” them. When you’ve actually “dibbled” them, you learn a lot about the kids’ ability and about the test itself. We have the coaches doing it for reliability issues and validity issues, but if you did it yourself as a teacher, it might help [you] understand it better.

Notably, the STEP assessment is different from the other assessments in having an explicit and heavy emphasis on teacher observation notes (recorded during the assessment), which also underscores the importance of teachers as the assessment givers.

Page 61 of the STEP manual, for example, stresses:

Recording your comments and analysis of the child’s reading behaviors and the conversations you have about comprehension are crucial. Without these notes, it becomes difficult to design instruction that builds upon a child’s strengths and responds to his or her learning needs. Simply knowing the STEP level is not enough to inform and provide appropriate instruction.

This emphasis is consistent with Sharp and Risko’s (2003) finding in a small study of Ohio teachers using the TPRI and Wireless Generation palm/website tools. Teachers reported that they got many of their insights about a student while giving the assessment, and that the post-assessment data displays often served more to remind them of those insights than to generate insights.

Finally, in supporting teachers’ conversations, there may be an important but often overlooked feature in what gets communicated to teachers up front:

- *Clear communication to teachers that a primary purpose of the assessment is to facilitate their conversations.*

In his advice to Reading First sites (see <http://www.ncrel.org/rf/good2.htm>) DIBELS co-developer Roland Good explicitly states, “An important function of assessment is to build a common language and vocabulary for the team.” The STEP assessment also underscores the value of this function. On the first page of the manual text, in the list of five main purposes of the assessment, the authors write that STEP enables teachers and administrators to: “Develop a common language for the discussion of student progress and make the tracking of student growth a topic of open conversation across classrooms.”

Section 7. Summary and Future Directions

This final section summarizes the main points of this report within three discussions: Conversation Types, Conversation Evolution, and Reading First Issues. Each of these discussions includes suggestions for future work.

7.01 Conversation Types

Exploring the various components, visualizations, and assumptions of the key assessments in this report leads inevitably to the recognition that there is no single type of data-driven instructional conversation. Instead, one can organize many of the various aspects of conversations discussed here around five major types of conversations: Identifying conversations, Planning conversations, Evaluating conversations, Revising conversations, and Conceptualizing conversations. There is no intended strict hierarchy here, either in the importance or in the temporal occurrence of this proposed typology; in fact, any single conversation among teachers may include elements from each of these types.

Identifying conversations. In these conversations, teachers ask, “Which children need extra supports?” Visualizations used here need to be “actionable” in allowing teachers to assign low-scoring children – especially those who continue to score low in the middle of the year – to supports that go beyond the regular classroom instruction groupings. For example, some children may be assigned to “double-dipping” or two small groups of reading instruction per day instead of one. Some children may be assigned to go back and work on particular skills, such as early phonemic awareness, after classroom instruction has already moved on. Some children may be assigned to one-on-one instructional time with reading specialists or tutors. These conversations represent one part of the “grouping” emphasis in Reading First. Most current implementations use training and/or mentors that focus on the use of overall class data summaries as the sources for these conversations. As noted in Section 5 (Explicit Deep Assumptions) one of the challenges in supporting teachers’ conversations is helping them to focus on student strengths as well as weaknesses.

Planning conversations. These conversations take off from Identifying conversations, and they are at the epicenter of current attention in Reading First.

Questions in these conversations go beyond, for example, assigning low-scoring children to extra support structures (e.g., one on one tutoring, or two small groups of reading instruction a day of reading versus one). In planning conversations, teachers discuss questions such as, “How do I divide up my entire classroom into groups as part of my regular instruction?” “What do I do with each small group in my classroom?” and “What happens in the extra support structures for low-skilled children?”

As with Identifying conversations, most of the strategies employed in Reading First implementations currently emphasize the use of overall class data summaries as the sources for driving these conversations, at least as a first step. Further analyses of children’s responses to specific test items may be more important sources for some assessments than others. As an example, teachers working with DIBELS need to move beyond the overall summaries to decide how to structure instruction for children who score as “some risk” on the alphabetic principle. Some children in this group may need to work on letter-sound relationships; other children may know these relationships but need practice in blending. In this case, only additional views of the data that show individual performance provide the level of information that teachers need to group children who need the same kind of instruction within a particular skill area and plan appropriate activities. TPRI, in contrast, breaks down areas such as alphabetic principle skills into finer sub-skill categories directly on the overall class summary page; it also provides a skill map with additional “actionable” information that can support Planning conversations.

All of the state-of-the-art Reading First implementations also recognize that data views are not enough in themselves. Sites that strive for highly supported Planning conversations need to provide teachers with knowledge about the instructional activities that support specific skill needs. Knowing that children need to work on blending phonemes, for example, is different from knowing what instructional activities will help children learn to blend phonemes. Knowing that children need help with implicit comprehension is different from knowing how to strengthen your comprehension instruction.

As noted earlier in this paper, a major concern related to planning conversations is that they might focus too heavily on discussions about isolated skill activities, rather than

on activities that integrate skills into real reading and writing practices. The STEP assessment, in particular, recognizes this possibility and incorporates a naturalistic book reading activity as part of the assessment, to support teachers in focusing on the integration of skills in reading and moving beyond a tendency to interpret the data as a call for drills of isolated skills. Some researchers are also concerned that the absence of certain skill areas on assessments (e.g., the absence of Concept of Word skills on DIBELS) may negatively influence the content of teachers' planning conversations.

Finally, as noted earlier, another challenge is helping teachers to look at all levels of student performance, including the high performers. It may be helpful here to adopt language from the GROW Network. Daniel Light, from CCT, notes that in GROW reports, when children perform at high levels in a particular area, the reports flag this area as a place where children could benefit from advanced work. This represents a potentially important mindset that contrasts from the mindset of "these children have reached the goals, so I don't have to worry about them." Instead, by using the language of "these children could benefit from more advanced work," GROW specifies that these children also have a need. Providing these children with more advanced work is an "actionable" suggestion. Bringing this need into Reading First conversations may also encourage early grade teachers to bring comprehension and vocabulary – two skills that are (a) important for readers of all levels to continue developing but (b) are less addressed than other skills in some of the key assessments – into more of the Reading First conversations.

Evaluating conversations. These conversations happen throughout the year, and they include such questions as "Is what we're doing working?" "Is it working for all children?" and "Are we supporting children as strongly as we could?" DIBELS provides some of the strongest support for these conversations through its progress-monitoring data views and through its structured conversation templates, complete with comparison data from other schools.

One of the concerns about these conversations is that they might often exclude discussions about whether highly-skilled children are continually moving forward in their development or are merely stagnating. The STEP assessment provides strong support for

this discussion through its visualizations that highlight student movement, not just overall skill level.

Revising conversations. These conversations happen when Evaluating conversations signal the need for a change. In Revising conversations, teachers ask, “What is our Plan B, now that Plan A hasn’t worked?” and “What are we failing to address in our instruction?” The PALS and STEP assessments provide detailed information about children’s oral reading performance – beyond rate and accuracy – that may be helpful here. All of the assessments also provide access to items analyses that could reveal insights for these conversations; however, the sources for this report suggest that teachers’ use of these item-level data views is currently even less-explored territory than their use of overall class data summaries.

Conceptualizing conversations. These conversations happen when any of the other conversations reveal differences in the way that individuals within the discussion group think about literacy development, or when conversations reveal gaps in the group’s knowledge about the underlying skills of reading. In these conversations, teachers discuss such questions as “What do we mean by fluency?” and “How does all this phonemic awareness fit in, anyway?” and “What is literacy development supposed to look like?” These ongoing conversations in the context of teachers’ own assessment data may be critical for helping teachers to make the link between data and instruction, and to truly understand why certain instructional activities support particular skill needs.

The STEP assessment provides support for conceptualizing conversations by placing student performance directly and explicitly (in its visualizations) along a continuum of literacy development. The TPRI skill map is another example of a visualization that directly links student assessment data to a pathway of skill development. Other assessments are linked to models of literacy development (see e.g., Good, Simmons, & Kame’enui, 2001), but the links and models are typically conveyed to teachers in separate visualizations used during training, and not directly on the data report screens.

7.02 Conversation Evolution

None of the conversation types described above are meant to be “better” or “worse” than other types. But to support data-driven conversations that matter for

students, we need to have some way to identify conversations – of any type – that are better or worse than other conversations of that same type. What does the best kind of Planning conversation look like? The best kind of Conceptualizing conversation? What’s an example of a Planning or Conceptualizing conversation that isn’t as good?

Our understanding of what makes conversations “good” or “not good” is still developing, but there are some points of consensus, particularly for Planning conversations. Good Planning conversations include clear rationales for grouping decisions. They include discussions of specific instructional activities that will address the needs highlighted by the data. And they include links to fully developed understandings of literacy development. A poor conversation, on the other hand, may not show sophisticated understanding of literacy development – as in the example of teachers who decide to go all the way back to the early skill of rhyming for a group of students who score low on phonemic awareness, when what those particular students really need is help in a more advanced skill such as identifying middle sounds.

Further understandings of what makes a data-driven conversation “good” may be advanced by a closer look at the deep assumptions behind the assessments, something that this report began to explore but that requires more conversations among the developers of the different assessments. Additionally, work by researchers at CCT (Center for Children and Technology) promises to shed more light in this area. Drawing from the field of organization and management theory, these researchers have created a framework for understanding individual data-driven decision-making. Within this framework, there is a continuum between data, information, and knowledge:

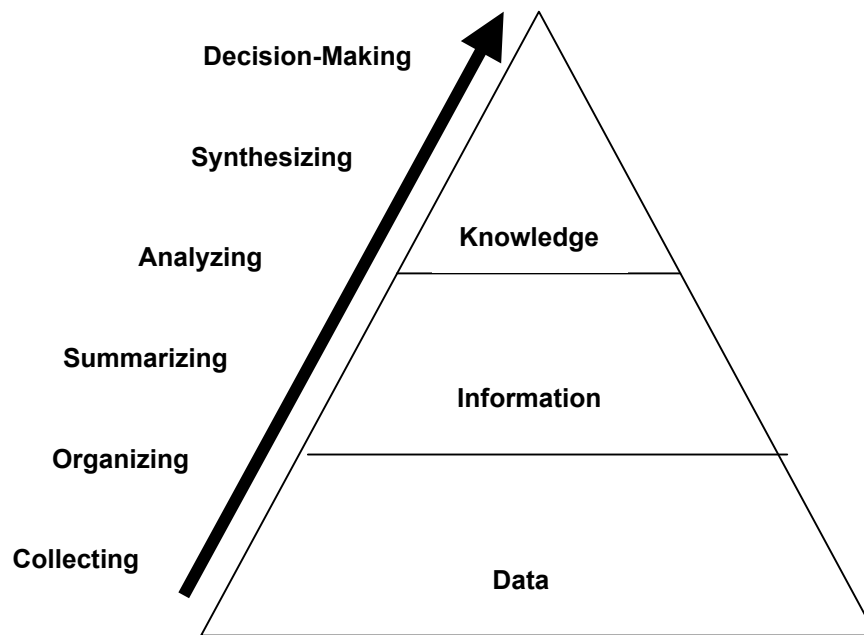
(from Light, Wexler, & Heinze, 2004)

Data exist in a raw state. They do not have meaning in and of itself, and therefore, can exist in any form, usable or not. Whether or not data become information depends on the understanding of the person looking at the data.

Information is data that is given meaning when connected to a context. It is data used to comprehend and organize our environment, unveiling an understanding of relations between data and context. Alone, however, it does not carry any implications for future action.

Knowledge is the collection of information deemed useful, and eventually used to guide action...teachers’ ability to see connections between students’ scores on different item-skills analysis and classroom instruction, and then act on them, represents knowledge.

Light et al. (2004) then describe six broad steps, drawn from the literature (Ackoff, 1989; Drucker, 1989), involved in transforming data to knowledge, as shown below in Figure 9:



*Figure 9: The process of transforming data into knowledge
(Light et al., 2004, p. 4, reproduced with permission)*

Teachers' conversations may follow these transformational steps as they progress toward better and better conversations. Early conversations may focus heavily on organizing and summarizing the data. Over time, these conversations may move to the level of analyzing and synthesizing the data as a basis for making decisions. Future research may find it helpful to look for these levels in teachers' conversations about their data and strive to scaffold the conversations from lower to higher levels. Just as children's comprehension evolves from lower to higher levels, so teachers' conversations may evolve as they gain more experience with data-driven conversations. [Thanks to Susan Hall for this analogy.]

CCT researchers suspect that the six skills described above don't necessarily evolve in clear stages. Ellen Mandinach, CCT, explains:

We're working on the framework of the six skills...because it is very much cyclical. A discussion is not a static thing, it's a recurring, dynamic phenomenon. [We] realize that there is a six-point continuum, but you go back and forth across the three levels [Data, Information, Knowledge] and six skills.

Appendix D gives more information about the CCT research underway in this area.

7.03 Reading First Issues

This report suggests that one of the potential obstacles to successful conversations is the current confusion about assessment terminology used in the federal language, in the IDEA report (Kame'enui, 2002), and in the assessments themselves. Additionally, some researchers are concerned that the conceptual structure of four assessment categories (screening, diagnosis, progress monitoring, and outcome) combined in a matrix with five components of reading (phonemic awareness, phonics, fluency, comprehension, and vocabulary) has led unintentionally to assessment plans that are currently overly complex and disjointed.

Resolving the terminology issue, then, may require more than a consensus about the meanings of assessment category terms. Researchers, such as Barbara Foorman, are calling for a move to more seamless systems, in which the same assessments serve as many purposes as possible. Additionally, from the teachers' point of view, it may make sense to think more along the line of strategies – that is, how the data is “actionable” in the classroom – rather than in terms of the four IDEA categories. Put simply, teachers primarily need to have data for Plan A strategies and Plan B strategies. Plan A data sources give the teacher information about grouping students, assigning them to extra supports, designing instructional activities, and ensuring – throughout the year – that the instruction in place is helping all students move forward at an appropriate rate. When the data indicate that instruction is not working, then teachers need Plan B strategies. These strategies may include different assessments – some more appropriately administered by a psychologist or reading specialist – or different interpretations of the data gathered under the Plan A assessments, including closer looks at item-level responses.

There is not a consensus at this point about whether a more seamless assessment system needs to have outcome assessments that are separate from what a teacher uses for Plan A and Plan B strategies. TPRI and PALS developers see a distinctiveness about outcome measures that lies outside the purposes of their assessments; DIBELS is often used for multiple purposes including outcomes; and STEP developers see the STEP assessment as covering outcomes as well as other purposes. Ideally, a highly seamless system would eliminate the need for separate outcome measures, but that may not be realistic given the entrenchment of state outcome measures that are separate from formative assessments for teachers' use.

There is another way that Reading First assessments can strive for seamlessness, and that is through further development of measures for vocabulary and comprehension, reducing the need for completely separate assessments targeting those two skills. DIBELS is the only key assessment reviewed here with a separate vocabulary [word meaning] measure, although TPRI is developing one. STEP is the only assessment reviewed here with a heavy emphasis on measures of comprehension, sufficient for extensively informing instruction.

7.04 Final Thoughts

This report began with the question of whether the combination of new technologies, new reading assessments, and new support for social practices around data-driven decision-making could together create a revolution in the field of reading education as profound as the revolution in navigation brought about by GPS tools. As this report clearly shows, there is much we still need to learn – especially about teachers' data-driven instructional conversations– before we will know the answer. However, there can be little doubt that with the massive resources and statewide programs recently put in motion (and quickly gaining speed) through Reading First, there are plenty of immediate opportunities for pursuing the question. For the reading research field, this translates into an open challenge:

Keep up. And find out.

References

- Ackoff, R. L. (1989). From data to wisdom. Journal of Applied Systems Analysis, 16, 3 – 9.
- Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. Educational Researcher, 32(6), 3 – 12.
- Drucker, P. F. (1989). The new realities: In government and politics/In economics and business/In society and world view. New York: Harper & Row.
- Foorman, B. R., Santi, K. L., & Berger, L. (2004). Scaling assessment-driven instruction using the Internet and handheld computers. Presented at a conference in Washington DC, Nov 3-4, 2003, entitled, “Conceptualizing Scale-Up: Multidisciplinary Perspectives.
- Fuchs, L., S., Fuchs, D., & Compton, D. L. (in press). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. Exceptional Children.
- Good, R. H., Simmons, D.C., & Kame’enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. Scientific Studies of Reading, 5, 257-288.
- Good, R. (2002). Reading First Initiative: The purpose of assessment is to change life trajectories for children. PowerPoint presented at the New York State K – 3 Teacher Reading Academy. Downloaded from www.emsc.nysed.gov/nyc/REA/REASummerInstituteFollowUp/NYSTeacherReading-RolandGood.ppt.
- Hupert, N., Martin, W., Heinze, C., & Perez, H. E. (2004). Trends in use of handheld technology to support student reading assessment. Presented at the 2004 National Educational Computing Conference in New Orleans, LA. Available at <http://www2.edc.org/cct> under Publications > Speeches and Presentations.
- Kame’enui, K. (2002). An Analysis of Reading Assessment Instruments K – 3: Final Report. Retrieved from: <http://idea.uoregon.edu/assessment/>

Light, D., Wexler, D., & Heinze, J. (2004). How practitioners interpret and link data to instruction: Research findings on New York City Schools' implementation of the Grow Network. Paper presented at AERA 2004, San Diego, CA.

National Reading Panel (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Washington, DC: NIH.

Rathvon, N. (2004). Early reading assessment. New York: The Guilford Press.

Roehrig, A. D. (2004). The relationship between retention/promotion decision-making processes and student reading gains. Paper presented at the annual meeting of the American Psychological Association, Honolulu, HI.

Sharp, D., & Risko, V. (2003). All in the palm of your hand: Lessons from one school's first steps with handheld technology for literacy assessments. Report to the Information Infrastructure System project, Network on Teaching and Learning, John D. and Catherine T. MacArthur Foundation.

Shinn, M. R. (1989). Curriculum-Based Measurement: Assessing special children. New York: Guilford.

Appendix A

Interviews

Mary Fowler, Assistant Professor, Curry School of Education, University of Virginia. Interviewed via phone on August 5, 2004.

Marguerite Held, teacher-trainer, Center for Academic and Reading Skills, University of Texas Health Science Center at Houston. Interviewed via phone on July 23, 2004.

Naomi Hupert, Senior Research Associate, EDC's Center for Children and Technology, Co-Director of the New Mexico Reading First Evaluation project. Interviewed via phone on July 21, 2004.

David Kerbow, Senior Research Analyst, Consortium on Chicago School Research at the University of Chicago, co-developer, STEP literacy assessment. Interviewed via phone on June 29, 2004.

Daniel Light, Senior Research Associate, EDC's Center for Children and Technology. Researcher with the Grow Network project in New York City schools. Interviewed via phone on July 21, 2004.

Ellen Mandinach, Senior Research Associate, EDC's Center for Children and Technology (CCT), Director of NSF funded study, 2003 – 2006: Creating an Evaluation Framework for Data-Driven Instructional Decision-Making. Interviewed via phone on July 22, 2004.

Alysia D. Roehrig, Assistant Professor of Educational Psychology, Florida State University; Center Research Faculty, Florida Center for Reading Research. Interviewed via phone on July 22, 2004.

Kristi Santi, Assistant Professor at the Center for Academic and Reading Skills, University of Texas Health Science Center at Houston. Project coordinator for IERI project, Scaling up Assessment-Driven Intervention Using the Internet and Handheld Computers. Interviewed via phone on July 22, 2004.

Appendix B

Useful Websites

<http://www.ed.gov/programs/readingfirst/index.html>

Official Reading First site from U.S. Department of Education

<http://www.tpri.org/>

Official site for the TPRI (Texas Primary Reading Inventory) assessment

<http://pals.virginia.edu/>

Official site for the PALS (Phonological Awareness and Literacy Screening) assessment

<http://dibels.uoregon.edu/>

Official DIBELS (Dynamic Indicators of Basic Early Literacy Skills) website. Includes a link to download the assessments (free).

<http://reading.uoregon.edu>

Big Ideas in Beginning Reading website (University of Oregon). Includes information for teachers about the five components of reading (Phonemic Awareness, Alphabetic Principle, Fluency, Comprehension, Vocabulary) and instructional practices.

<http://dibels.uoregon.edu/data/index.php>

DIBELS data system site (University of Oregon) for online data entry and generated reports. Includes sample reports.

<http://www.edformation.com/products/DIBELS.htm>

AIMSweb, a commercial system for managing, charting, and reporting DIBELS data

<http://www.fcrr.org/pmrn/>

Florida Center for Reading Research's Progress Monitoring and Reporting Network

<http://usi.uchicago.edu/tools.html#STEP>

Site for the STEP (Strategic Teaching and Evaluation of Progress) literacy assessment, University of Chicago, Center for Urban School Improvement

<http://usi.uchicago.edu/research.html#iis>

Site for the Information Infrastructure System project

<http://www.ncrel.org/rf/>

Reading First Subgrant Technical Assistance website by the North Central Regional Education Laboratory (NCREL)

<http://www.sedl.org/readingfirst/welcome.html>

Reading First database, Southwest Educational Development Laboratory (SEDL)

<http://www.wgen.net/web/>

Wireless Generation. Handheld to web assessment solutions

<http://idea.uoregon.edu/assessment/>

Site for the IDEA report, *Analysis of Reading Assessment Instruments for K – 3* (Kame'enui, 2002)

<http://www.grownetwork.com/>

Grow Network, a data reporting system for states and districts that focuses on end-of-year standardized assessment data

Appendix C: Components of the Key Assessments

Texas Primary Reading Inventory (TPRI)

		Skill Area	Task Name or Description
Kindergarten			
	Screening		
		Graphophonemic Knowledge	Letter Sound
		Phonemic Awareness	Blending Onsets-Rimes and Phonemes
	Inventory		
		Book and Print Awareness	Warm-up activity-not scored
		Phonemic Awareness	Rhyming, Blending Word Parts, Detecting Initial Sounds, Detecting Final Sounds
		Graphophonemic Knowledge	Letter-Name Identification, Letter-to-Sound Linking
		Listening Comprehension	Five comprehension questions for a story
Grade 1			
	Screening		
		Graphophonemic Knowledge	Letter Sound
		Word Reading	List of eight words
		Phonemic Awareness	Blending Word Parts/Phonemes
	Inventory		
		Phonemic Awareness	Blending Word Parts, Blending Phonemes, Detecting Initial Sounds, Detecting Final Sounds
		Graphophonemic Knowledge	Initial Consonant Substitution, Final Consonant Substitution, Middle Vowel Substitution, Initial Blending Substitution, Blends in Final Position
		Word List/Passage Placement	List of 15 words
		Reading Accuracy, Fluency, Comprehension	Passage reading, comprehension questions

Texas Primary Reading Inventory (TPRI), continued

		Skill Area	Task Name or Description
Grade 2			
	Screening		
		Word Reading	List of eight words
	Inventory		
		Graphophonemic Knowledge	Spelling: (1) CVC, CVCe, R-controlled Vowels, and Blends; (2) Long Vowels, Digraphs, and Orthographic Patterns; (3) Blends, Digraphs, Compounds, Past Tense, Homophones, and Orthographic Patterns; (4) Plural, Digraphs, Blends, Consonant Doubling, Past Tense, Inflectional Endings, and Changing y to i.
		Word List/Passage Placement	List of 15 words
		Reading Accuracy, Fluency, Comprehension	Passage reading, comprehension questions

Phonological Awareness Literacy Screening (PALS)

		Skill Area	Task Name or Description
Kindergarten			
		Phonological Awareness	Rhyme Awareness, Beginning Sound Awareness
		Alphabet Recognition	List of 26 lower-case letters for students to name
		Concept of Word	Uses a memorized rhyme and asks students to touch words while reading and identify words from the book outside the text.
		Letter-Sounds	23 upper case letters and three digraphs
		Spelling	Five CVC words
		Word Recognition in Isolation	Optional, for students with some reading ability
Grades 1 - 3			
	Entry Level/ Level A		
		Spelling	Scored by phonics features
		Word Recognition in Isolation	Graded word list
		Oral Reading/Fluency	Timed passage reading scored with running record, calculated rate, and optional fluency rubric
		Comprehension	Optional multiple choice questions
	Level B		
		Alphabetics	Alphabet Recognition, Letter Sounds, Concept of Word
	Level C		
		Phonemic Awareness	Blending, Sound-to-Letter

Dynamic Indicators of Basic Early Literacy Skills (DIBELS)

	Skill Area	Task Name or Description
Kindergarten		
	Phonological/Phonemic Awareness	Initial Sound Fluency, Phonemic Segmentation Fluency
	Phonics (Alphabetic Principle)	Nonsense Word Fluency
	Fluency	Oral Reading Fluency
	Comprehension	Retell Fluency
	Vocabulary	Word Use Fluency
	Letter Naming Fluency	Identification of upper and lower case letter names – considered a separate domain from phonics
Grade 1		
	Phonemic Awareness	Phonemic Segmentation Fluency
	Phonics (Alphabetic Principle)	Nonsense Word Fluency (basic), Oral Reading Fluency (advanced)
	Fluency	Oral Reading Fluency
	Comprehension	Retell Fluency
	Vocabulary	Word Use Fluency
	Letter Naming Fluency	Identification of upper and lower case letter names – considered a separate domain from phonics
Grade 2		
	Phonics (Alphabetic Principle)	Nonsense Word Fluency (basic), Oral Reading Fluency (advanced)
	Fluency	Oral Reading Fluency
	Comprehension	Retell Fluency
	Vocabulary	Word Use Fluency
Grade 3		
	Fluency	Oral Reading Fluency
	Comprehension	Retell Fluency
	Vocabulary	Word Use Fluency
Grades 4, 5, 6		
	Fluency	Oral Reading Fluency
	Comprehension	Retell Fluency

Strategic Teaching and Evaluation of Progress (STEP)

	Step Level	Task Components
Kindergarten		
	Pre-Reading	Name Assessment, Rhyming Words, Concepts about Print
	Step 1	Concepts about Print, Letter and Letter-Sound Identification, Matching First Sounds, Developmental Spelling, Reading Record, Comprehension Conversation
	Step 2	Letter and Letter-Sound Identification, Segmentation, Developmental Spelling, Reading Record, Comprehension Conversation
Grade 1		
	Step 3	Letter and Letter-Sound Identification, Segmentation, Developmental Spelling, Reading Record, Comprehension Conversation
	Steps 4 - 6	Developmental Spelling, Reading Record, Fluency and Reading Rate, Comprehension Conversation
Grade 2		
	Step 7	Developmental Spelling, Reading Record, Fluency and Reading Rate, Comprehension Conversation
	Step 8	Developmental Spelling, Reading Record, Fluency and Reading Rate, Comprehension Conversation, Story Retelling
	Step 9	Developmental Spelling, Reading Record, Fluency and Reading Rate, Comprehension Conversation, Story Retelling, Comprehension Writing
Grade 3		
	Steps 10 - 12	Developmental Spelling, Reading Record, Fluency and Reading Rate, Comprehension Conversation, Story Retelling, Comprehension Writing

Appendix D

Pathway Projects

Reading First has propelled the field forward in implementations that aim to support teachers' data-driven decision-making in early literacy. Many of these implementations include opportunities for supporting teachers' instructional conversations. These implementation projects may provide alternative pathways for understanding the major issues of interest to the IIS group.

As the Information Infrastructure System project continues to explore this area, it may be helpful to establish connections with these "Pathway Projects" – or at least keep track of their progress. Even in cases where projects may use contrasting approaches, their experiences may provide solutions to dilemmas that arise in IIS project work.

Florida Center for Reading Research

The Florida Center for Reading Research has a four-part mission

(From <http://www.fcrr.org/>):

1. To serve as part of Florida's "Leadership Triangle" for the Just Read, Florida! initiative and to provide technical assistance and support to all districts and schools receiving a Reading First Award
2. To conduct applied research that will have an immediate impact on policy and practices related to literacy instruction and assessment in Florida
3. To disseminate information about research-based practices related to literacy instruction and assessment for children in pre-school through 12th grade
4. To conduct basic research on reading, reading growth, reading assessment, and reading instruction that will contribute to the scientific knowledge of reading

All Reading First sites in Florida use the DIBELS assessment and enter the data into the Florida Progress Monitoring and Reporting Network, which provides visualizations and reports. The center is keenly interested in supporting teachers to use DIBELS data for creating small groups, progress monitoring, guiding instruction, and parent conferencing. Current studies include a focus on how teachers use the data as part of their retention/promotion decision-making process (Roehrig, 2004).

Scaling up Assessment-Driven Intervention Using the Internet and Handheld Computers

This IERI project (PI, Barbara Foorman) at the Center for Academic Reading Skills, University of Texas – Houston, began in 2002. It is a major three-year study in 255 Texas schools, aimed at questions of how technology tools can facilitate assessment-driven instruction. Of particular interest to the IIS project are these elements:

- The project is focused on capturing the depth of change and the shift of reform ownership (Coburn, 2003).
- The project's website offers video segments of small group, differentiated instruction K – 2.
- The website includes a Teacher's Lounge, where teachers can electronically converse with each other or post questions to research staff. The IERI project will track the topics of these conversations.

The study design includes three conditions of tools: paper only, paper plus desktop (web entry), and handheld computer plus desktop. There are also three types of interpretation support: teacher only (materials available in testing materials and intervention guide), website mentor (master teacher insight available on web), and on-site mentor.

Reading First Evaluation for the State of New Mexico, Center for Children and Technology, EDC

This project is assessing the impact of 491 K – 3rd grade teachers administering DIBELS three times during the year with over 7000 students, using the PalmPilots during the year to over 7000 students. The project began in the 2003-2004 school year, using a

combination of school visits, observations, and interviews with administrators and teachers (Hupert, Martin, Heinze & Perez, 2004).

Creating an Evaluation Framework for Data-Driven Instructional Decision-Making, Center for Children and Technology, EDC

This NSF-sponsored project, funded from 2003-2006, is synthesizing research across projects in Albuquerque Public Schools (using handhelds with Reading First data), New York City schools (using data from the GROW network), and Broward County Public Schools (using a locally developed data-warehousing and data-support system). The project aims to create a framework and accompanying systems models that can be used to evaluate data-driven instructional decision making across a variety of contexts.