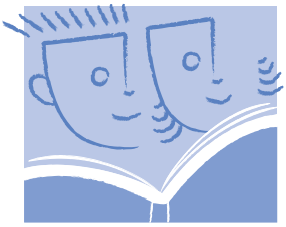
DIBELS: MYTHS and FACTS

Ruth Kaminski & Kelli D. Cummings
Dynamic Measurement Group



Dynamic Indicators of Basic Early Literacy Skills (DIBELS) comprise a set of procedures and measures for assessing the acquisition of early literacy and reading skills from kindergarten through sixth grade. Although DIBELS were designed as a formative assessment and evaluation tool to be used for low-stakes educational decisions, DIBELS have most recently been associated with and embroiled in the controversy related to assessment in general and high-stakes testing in particular. Critics of DIBELS argue that DIBELS data are being used to label and track students and as criteria for decisions about promotion and retention as well as for accountability and teacher evaluation. Under pressure to have children succeed, some teachers are "teaching the test" rather than using DIBELS as an *indicator* and teaching the basic early literacy skills that lead to reading success. As the authors of DIBELS, such uses of DIBELS data are alarming. It is clear to us that some educators have had limited information related to DIBELS, which has led to misconceptions about what DIBELS are and were designed to do. These misconceptions have resulted in some cases of misuse of the data.

We have organized this paper around what we believe are the most common myths about and misuses in DIBELS implementation. Some of the misuses are based on incorrect conclusions about the type of measurement that DIBELS represent (General Outcomes Measurement) or the decisions that DIBELS data are intended to facilitate. Others reflect opinions based on false attributes of the measure or confusion surrounding the underlying assessment concepts related to formative assessment and evaluation. To address confusion around the type of measurement that the DIBELS tools represent, we begin with a description of a framework for educational decision-making, the Outcomes-Driven model. We then highlight the ways in which DIBELS fit within this process.

DIBELS by Design

DIBELS were designed to be used within a formative assessment process to evaluate the effectiveness of interventions for those children receiving support in order to make changes when indicated to maximize student learning and growth. Initial research on DIBELS focused on examining the technical adequacy of the measures for these primary purposes (Good & Kaminski, 2002; Kaminski & Good, 1996), which remain the intended uses of DIBELS to this date (Kaminski & Cummings, 2007).

Supporting Individuals

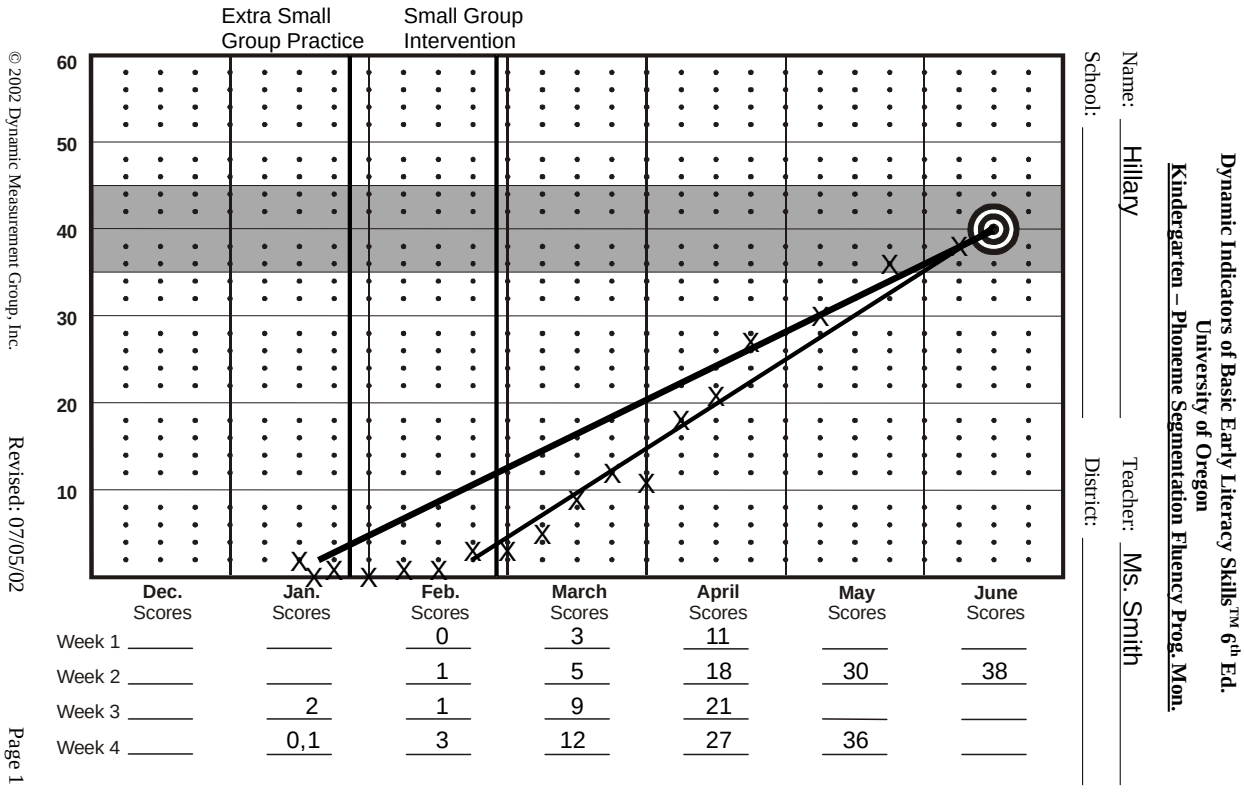
DIBELS were designed to provide information that would help educators match the amount and type of instructional support with the needs of individual students to enable all students to become successful readers. DIBELS were never intended to be used alone as the sole measure of a child's or school's success but rather within a *system* of literacy support that is linked to a model of data-based decision making. The Outcomes-Driven Model (Good, Gruba & Kaminski, 2001; Kaminski & Good, 1998) consists of a set of procedures for making educational decisions that prevent later reading problems by using DIBELS data to: (1) identify need for support early, (2) validate need for support, (3) plan support, (4) evaluate and modify support as needed, and (5) periodically review outcomes for all children.

Within an Outcomes-Driven Model, all children in grades kindergarten through third grade are tested on DIBELS three times each year – fall, winter, and

spring. This "benchmark assessment" is akin to universal screening and is done to identify those children who may need support to achieve reading goals. When a need for support is indicated, additional assessment information is gathered as necessary to validate that need. Progress monitoring is a key element of intervention programs for children who receive support. The purpose of this step of the Outcomes-Driven Model is to ensure that the support is working to get students back on track.

Figure 1 illustrates the use of DIBELS data within an Outcomes-Driven Model as implemented for Hillary, a kindergarten student, from January to June of her kindergarten year. In January, at the winter benchmark, Hillary's performance (below the cutoff on Phoneme Segmentation Fluency) indicated a possible need for support in the area of phonemic awareness. Because DIBELS are only one piece of information that teachers have access to when making educational decisions, it is recommended that teachers use additional assessment information and knowledge about a student to validate a score and the decision to provide additional support. One option for validating scores is to retest the student using alternate forms of the DIBELS measures. In this case, Hillary's teacher, Ms. Smith, wanted to make sure that Hillary's low performance on the winter benchmark assessment was not simply because an unfamiliar person tested her right after the winter break. To validate the benchmark results, Ms. Smith personally retested Hillary with alternate forms of PSF on two different days the following week. Ms. Smith then examined Hillary's pattern of performance across the multiple assessments, and noted that her scores across the repeated assessments were consistent with her benchmark score. Ms. Smith concluded that Hillary's low score was not just a bad day and that she was in need of additional instructional support.

Figure 1



Through consultation with the reading support team, Ms. Smith decided to place Hillary in a small group for some extra practice in phonemic awareness skills, using the phonemic awareness activities from the core reading curriculum. Because Ms. Smith wanted to be sure that the support was effective, she assessed the skills of students in this group weekly on PSF. Using the progress-monitoring booklet for PSF, Ms. Smith administered a PSF probe to Hillary using alternate forms every week and plotted her performance on the graph on the front of her progress monitoring booklet.

The small-group extra practice started at the beginning of February. In the school that Hillary attended, end of month data meetings were conducted as part of regular grade-level meetings. By the first end-of-month data meeting in February, Ms. Smith had four additional PSF scores. Although the extra small-group practice was working for the other children in the group, Hillary's performance continued consistently below the aimline. The team decided the small group extra practice was not sufficient for Hillary to achieve the benchmark goal for PSF by the end of Kindergarten. Hillary was moved to another instructional group that included two other students from different kindergarten classrooms. Her program was modified to provide explicit instruction on earlier phonemic awareness skills with additional modeling, examples, and practice. At the data meeting at the end of March, Hillary's performance climbed to the aimline. By the end of April, Hillary's performance continued to hug the aimline and the team decision was that the additional support was effective to keep Hillary on track to achieve the end-of-year benchmark goal for phonemic segmentation. Hillary continued to receive small group support and the frequency of progress monitoring was reduced to every other week. Hillary met the benchmark goal of 35 by the end of May and maintained her level of performance until the end of the school year, when outcomes were reviewed.

As illustrated by the example, the use of DIBELS within an Outcomes-Driven Model is prevention-oriented and is designed to preempt reading difficulties and support all children to achieve adequate reading outcomes. The model incorporates conceptual foundations regarding crucial early literacy skills for assessment and instruction and is focused on empirically validated outcomes for each early literacy skill. The model builds on reliable and valid measures of essential early literacy skills to be used to document growth toward outcomes, as well as a set of steps for using the data generated by the measures. The use of DIBELS to identify students in need of support and evaluate the effectiveness of the support provided is consistent with the most recent reauthorization of the Individuals with Disabilities in Education Improvement Act (IDEIA, 2004) allowing use of a Response to Intervention (RTI) approach to identify children with learning disabilities. In an RTI approach to identification, early intervention is provided to students who are at risk for the development of learning difficulties. Data are gathered to determine which students are responsive to the intervention provided and which students are in need of more intensive support (Fuchs & Fuchs, 2006). Rather than labeling or tracking, DIBELS data are used to prevent school failure and provide support that is linked to student need.

Supporting Systems

DIBELS data may also be aggregated and used at a systems level (e.g., school, district, agency, state) formatively, to identify need for support within a school or district. Through use of DIBELS data collection across a school year, administrators have access to data on all students in the system. The data can be used to identify the percentage of students who are on track and who are making adequate progress. Aggregation of DIBELS data at the systems level provides information that may be used to examine the effectiveness of the instructional supports within a classroom, school, or district to help determine when changes should be made. As with decision making for individual students, when used at the systems level DIBELS data should be used *formatively* to *identify needs for support*. Possible targets of opportunity for support for your school-wide reading model may include aspects of the system such as:

a. The curricula and programs used in the school–including both the core reading program and any supplemental materials or interventions
b. Fidelity of implementation of curricula/instructional programs
c. Time allocated for instruction or intervention
d. Instructional grouping
e. Content and delivery model for professional development

Similar to formative assessment of individual students, systems-level DIBELS data are designed to be used by in-house staff of the programs with the intent to *improve* the programs. In this way, systems-level DIBELS data are helpful in evaluating overall effectiveness of support across a school year and mobilizing resources to improve programs at the systems level.

At the systems level the outcomes of the instructional support for all children are reviewed at the end of each benchmark period. The system-wide data informs needed modifications to the system of support, including the core and supplemental programs being implemented. The bottom line in the Outcomes-Driven Model is the achievement of crucial literacy outcomes for both individual students and systems at the classroom, school, and district levels. The outcomes drive the decisions. If outcomes for individual children and/or groups of children are adequate, the instruction and curricula are deemed to be adequate. However, if outcomes are not adequate, a change is warranted. Changes that improve outcomes are maintained; changes that maintain poor outcomes are abandoned. Because data are collected on all students on an ongoing basis, instructional and or curricular modifications can be made in a timely fashion to ensure all children achieve the goal of becoming established readers.

We would argue that making improvements in the system of support are some of the most important decisions that educators can make. However, such decisions are distinctly different from the use of DIBELS as a high stakes test. DIBELS data were *not* designed to make decisions at the school level regarding if a school/program/teacher should continue to receive funding. Rather, when implemented according to the prevention-oriented framework of the Outcomes-Driven Model, DIBELS data can be used to improve the effectiveness of a system and provide frequent feedback regarding the impact of those changes.

Our vision for the use of DIBELS includes the measures as one component of an effective, school-wide literacy model. The data from the DIBELS measures can be used to identify targets of opportunity for school improvement and suggest appropriate interventions for struggling readers. When educators use DIBELS data to evaluate effectiveness, prevention efforts are strengthened and student outcomes improve. However, myths surrounding the appropriate use of DIBELS may impede its use as an effective tool. We highlight some of the most common myths associated with DIBELS below, and then provide evidence regarding the validity of the measures for their intended uses.

Myths about DIBELS

Myth 1: *DIBELS measure splinter skills and do not prepare children to read for meaning.*

Fact: *DIBELS are deliberately designed to be indicators of critical early reading skills rather than comprehensive assessments of all reading and reading-related skills.*

DIBELS measures, by design, are *indicators* of each of the Basic Early Literacy Skills. For example, DIBELS do not measure all possible phonemic awareness skills such as rhyming, alliteration, blending, and segmenting. Instead, the DIBELS measure of phonemic awareness, Phoneme Segmentation Fluency (PSF), is designed to be an *indicator* of a student's progress toward the long-term phonemic awareness outcome of segmenting words. The notion of DIBELS as *indicators* is a critical one. It is this feature of DIBELS that distinguishes it from other assessments and puts it in a class of assessments known as General Outcome Measures.

DIBELS were developed based on measurement procedures used for Curriculum-based Measurement (CBM) by Deno and colleagues through the Institute for Research on Learning Disabilities at the University of Minnesota in the 1970s-80s (e.g., Deno & Mirkin, 1977; Deno, 1985; Deno & Fuchs, 1987; Shinn, 1989). Like CBM, DIBELS were developed to be economical and efficient *indicators* of a student's progress toward achieving an important outcome. Although DIBELS materials were initially developed to be linked to the local curriculum like CBM (Kaminski & Good, 1996), current DIBELS measures are generic and draw content from sources other than any specific school's curriculum. The use of generic CBM methodology is typically referred to as General Outcome Measurement (GOM) (Fuchs & Deno, 1994).

General Outcome Measures (GOMs) like DIBELS differ in meaningful and important ways from other commonly used formative assessment approaches. The most common formative assessment approach that teachers use is assessment of a child's progress in the curriculum, often called mastery measurement. End of unit tests in a curriculum are one example of mastery measurement. Teachers teach skills and then test for mastery of the skills just taught. They then teach the next set of skills in the sequence and assess mastery of those skills. Both the type and difficulty of the skills assessed change from test to test; therefore scores from different times in the school year cannot be compared. Mastery-based formative assessment such as end of unit tests addresses the question, "has the student learned the content taught?" In contrast, GOMs are designed to answer the question, "is the student learning and making progress toward the long-term goal?"

In much the same way as an individual's temperature or blood pressure can be used to indicate the effectiveness of a medical intervention, GOMs in the area of education can be used to indicate the effectiveness of our teaching. However, the powerful predictive validity of the measures does not mean that their content should become the sole components of our instruction. In other words, unlike mastery based assessment in which it is appropriate to teach the exact skills tested, each DIBELS indicator represents a broader sequence of skills to be taught. (For an example of sequence of skills related to and leading to the goals, please see Curriculum Maps at http://reading.uoregon.edu.) DIBELS measures are designed to be brief so that our teaching doesn't have to be.

Myth 2: *DIBELS measures are timed, so they are all about speed, not about real reading.*

Fact: *DIBELS measures are timed because fluency and automaticity in each underlying early literacy skill is a critical component to becoming a skilled reader. Both accuracy and fluency in early literacy skills are critical to successful reading, including comprehension.*

Because all DIBELS measures are timed, it is sometimes perceived that the emphasis of the measures is simply on speed. In actuality, DIBELS measures are timed for two reasons. The first reason that timing is used for DIBELS measures is that fluency in early literacy provides information about student performance above and beyond accuracy alone (Hudson, Lane, & Pullen, 2005; Pikulski & Chard, 2005, Shaywitz, 2003). A child who performs a task fluently, that is, both accurately *and* quickly, has learned the skill to mastery is automatic in performing the underlying skills and is much more able to remember, maintain, and apply the skill than a child who does not. Fluency of the component early literacy skills during the early school years provides a foundation upon which later skills are built. Fluency in oral reading is the gateway to comprehension. Research on oral reading fluency clearly shows that, in general, children who read fluently are more likely to comprehend what they read. Not only does fluency facilitate comprehension, but it also places an upper limit on the number of concepts that are encountered in a given passage of text (Howell & Nolet, 2000).

The second reason for timing is that measuring fluency results in scores that are more sensitive to small changes in growth and development of skills over relatively short periods of time (e.g., one week). For example, a child who is highly accurate but takes 3

seconds per item to respond to a task would achieve a maximum score of 20 within one minute. Although the child may have already been 100% accurate, as he/she masters the skill and becomes more fluent, his/her score will continue to go up. This improvement in fluency that is reflected in the overall score is *meaningful* in terms of its summary of student performance and would have been masked by attending to accuracy alone.

For all timed measures, the timing is as unobtrusive as possible. A silent count-down timer is used and no attention or emphasis is given to the timing aspect of the assessment. Most students are not even aware that timing is occurring. Students are never told to read as fast as possible; the DIBELS directions clearly state that students are to "do your best reading." When the assessment results are translated into effective instructional strategies, the emphasis is always on both accuracy and fluency—where accuracy is a requisite for fluency building. It doesn't do anyone any good to do something wrong, very, very fast.

Myth 3: *DIBELS are too hard. Children are penalized (i.e., scored "wrong") on things they have not yet had a chance to learn.*

Fact: *Because they are designed to measure growth, not simply status on a skill at a single point in time, scores increase as children learn. The DIBELS measures are sometimes perceived as "being too hard" when in fact they are designed so that students score lower early and do better as they are taught and learn essential skills.*

Because DIBELS were designed to be used in a repeated fashion to monitor progress of students in an ongoing manner the measures, by necessity, need to be brief, efficient, and instructionally relevant (Deno, 1985; Marston, 1989). The DIBELS measures are also sensitive, in that scores on the measures *change* as a result of student learning. Because they are designed to measure *growth*, not simply status on a skill at a single point in time, children need to start with low scores so scores can increase as children learn. For example, it is not expected that children in the middle of kindergarten will know all of their letter sounds and be able to correctly read all of the words on the Nonsense Word Fluency task (see Table 1). The measure and the benchmark goals are designed in such a way as to capture incremental progress in student learning. By the middle of first grade, however, the expectation is that children be fluent not only in naming the correct sound for letters, but also with blending the sounds to accurately and fluently read the words as whole words.

In a series of early research studies on DIBELS, measures that had floor effects (i.e. large numbers of students scoring at zero) were discarded (Gruba, 1997; Johnson, 1996; Kaminski, 1992; Koehler, 1996; Laimon, 1994; Otterstedt, 1993). For example, Phoneme Segmentation Fluency (PSF), an indicator of phonemic awareness, was initially administered at the beginning of kindergarten. Too many children scored zero at that time, so PSF was replaced with Initial Sounds Fluency (ISF), also a measure of phonemic awareness yet developmentally easier in the beginning of kindergarten. Further research on ISF indicated that by the middle of kindergarten most children were able to perform the task easily, so the goal and the measure for the latter half of the kindergarten year were increased to capture student progress. Because the intention of the measures is to identify and provide support to students who are struggling, the benchmark goals increase over time.

Myth 4: *The standardized procedures for administering DIBELS are inappropriate.*

Fact: *DIBELS measures are standardized so that a child's progress can be evaluated over time.*

DIBELS procedures, including administration and scoring, are standardized, that is, given and scored the same way each time for all students. This procedure ensures that the assessment is delivered uniformly across students. When all students are offered the same opportunity to respond to a particular task, educators can interpret scores relative to benchmark

Table 1

DIBELS Benchmark and Progressive Goals and Probability of Need for Support

Need for Support	Beginning of Year			Middle of Year			End of Year		
Measure	Intensive	Strategic	Benchmark	Intensive	Strategic	Benchmark	Intensive	Strategic	Benchmark
Kindergarten									
ISF	< 4	4 – 7	>= 8	< 10	10 – 24	>= 25			
PSF				< 7	7 – 17	>= 18	< 10	10 – 34	>= 35
NWF				< 5	5 – 12	>= 13	< 15	15 – 24	>= 25
First Grade									
PSF	< 10	10 – 34	>= 35	< 10	10 – 34	>= 35	< 10	10 – 34	>= 35
NWF	< 13	13 – 23	>= 24	< 30	30 – 49	>= 50	< 30	30 – 49	>= 50
ORF				< 8	8 – 19	>= 20	< 20	20 – 39	>= 40
Second Grade									
NWF	< 30	30 – 49	>= 50						
ORF	< 26	26 – 43	>= 44	< 52	52 – 67	>= 68	< 70	70 – 89	>= 90
Third Grade									
ORF	< 53	53 – 76	>= 77	< 67	67 – 91	>= 92	< 80	80 – 109	>=110

Note. ISF = Initial Sound Fluency. PSF = Phoneme Segmentation Fluency. NWF = Nonsense Word Fluency. ORF = Oral Reading Fluency.

goals as well as compare them across students and over time. DIBELS benchmark goals were empirically validated based on the standardized administration and scoring procedures, thus students' scores on DIBELS are only meaningful in comparison to the benchmark goals when they are obtained through standardized administration and scoring, which includes accurate timing. When DIBELS are administered and scored in the same way for all students, differences in scores between students are attributable to actual differences in student performance not to differences in how the measures are administered and scored. In like manner, when the measures are administered and scored the same way each time, differences in an individual student's performance over time are attributable to actual student growth.

DIBELS administration and scoring procedures were developed over a period of years for each measure, with procedures tested and revised repeatedly until those were found that worked with at least 95% of students. With students for whom a standardized administration may not provide an accurate estimate of skills, DIBELS approved accommodations are provided in the DIBELS Administration and Scoring Manual (Good & Kaminski, 2002). DIBELS approved accommodations are those changes in procedures that are unlikely to substantially change the meaning or interpretation of scores on the measures. For example, for students who may not understand the standardized directions, it is permissible on some measures to provide an extra practice example. When DIBELS approved accommodations are used, the regular DIBELS interpretation guidelines apply.

When the DIBELS assessments are administered and/or scored in ways different from either a DIBELS standard administration or the DIBELS approved accommodations, the test administration would be considered unstandardized, and the resulting scores could not be interpreted relative to DIBELS benchmark goals. For example, extended time or untimed administration is not an approved accommodation. Although scores from an untimed administration may provide other information about a child's performance, they would not be comparable or interpretable relative to benchmark goals.

Myth 5: *DIBELS are not appropriate for diverse learners.*

Fact: *DIBELS are appropriate for all students for whom learning to read in English is a current instructional goal, with a few exceptions.*

DIBELS are appropriate for all students for whom a goal is learning to read in English *with a few exceptions:* a) students who are deaf; b) students who have fluency-based speech disabilities, e.g., stuttering, oral apraxia; c) students who are learning to read in a language other than English; d) students with severe disabilities.

Students who are deaf. For most students who are deaf, the ability to use phonological representations of letters is seriously compromised (Leybaert & Charlier, 1996); yet phonological awareness still appears to a necessary skill for reading in both deaf and hearing children (Nielsen & Luetke-Stahlman, 2002). Because DIBELS were developed based on the research examining the process of learning to read for hearing students, the core competencies assessed by DIBELS, phonemic awareness and alphabetic principle, may not apply for students who are deaf and are learning to read. For children who are deaf and can read orally, oral reading fluency may be used; however the benchmark goals would not apply. DIBELS *would be* appropriate for children with mild to moderate hearing impairments who have residual hearing and who are learning phonemic awareness and phonics skills.

Students who have fluency or oral motor speech disabilities. Speech fluency is compromised in students who stutter or have oral motor speech disabilities, such as oral apraxia. Given that the nature of such disabilities is slow and/or dysfluent speech, (Paul, 2001) the use of fluency-based measures for these students would not be appropriate.

Students who are learning to read in a language other than English. DIBELS are designed to provide information about the progress of children in acquiring literacy skills for reading in English. For children who are learning to read in languages other than English, it would be most meaningful and appropriate to assess their reading skills in the language in which they are being instructed. For English language learners who are learning to read in English, DIBELS are appropriate for assessing and monitoring progress in acquisition of early reading skills. DIBELS have been used successfully with English language learners (Haagar & Windmueller, 2001). In addition, research findings on English learners indicates that children who are non-English speakers can learn to read as well in English as their English-speaking peers (Chiappe, Siegel, & Wade-Wooley, 2002; Geva, Yaghoub-Zadeh, & Schuster, 2000) and, in fact, often outperform their peers in phonemic awareness skills (Lesaux & Siegel, 2003).

Students with severe disabilities for whom a) reading is not on the IEP or b) reading is on the IEP but the long term goal is functional use of environmental print. There are a small number of students for whom learning to read connected text may not be an appropriate goal. For these students, it would be most meaningful and appropriate to use curriculum-based assessment strategies to monitor progress toward individual goals and objectives.

Use of DIBELS is appropriate for all other students, including those in special education for whom reading connected text is an IEP goal. For students receiving special education, it may be necessary to adjust goals and timelines and use out-of-grade level materials for progress monitoring.

Myth 6: *The benchmark goals are not appropriate for diverse learners. It is not realistic to think that children with diverse learning needs and from different cultural/ linguistic backgrounds can achieve these goals.*

Fact: *DIBELS benchmark goals have been researched across all geographic areas of the United States with a wide range of types of schools and students. Students who achieve the DIBELS benchmark goals have a higher probability of becoming readers than students who do not.*

The DIBELS benchmark goals and timelines for achieving the goals are summarized in Table 1. The purpose of the DIBELS benchmark goals is to provide educators with standards for gauging the progress of all students. The benchmark goals represent minimum levels of performance for all students to reach in order to be considered on track for becoming a reader. Benchmark goals for DIBELS measures were based on research that examined the longitudinal predictive validity of a score on a measure at a particular point in time (Good, Simmons, & Kame'enui, 2001). The model of DIBELS benchmark and progressive goals is designed to make explicit a set of parsimonious linkages between earlier and later skills that put the odds in favor of students reaching later, important literacy outcomes.

DIBELS benchmark and progressive goals initially were derived based on data from all schools participating in the DIBELS Data System during the 2000 – 2001 and 2001 – 2002 academic years. Additional studies have replicated the predictive utility of these measures (Rouse & Fantuzzo, 2006; Speece, Mills, Ritchey, & Hillman, 2003) and goals (Good, Baker, & Peyton, in-press) in other, diverse samples. In the original sample, Receiver Operator Characteristic (ROC) curves were examined for each individual measure and evaluated for subsequent benchmark goals. Benchmark goals for each measure and time period were established using a minimum cut point at which the odds were in favor of a student achieving subsequent early literacy goals. For a score to be considered a benchmark goal, at least 80% of students in the sample with that score at that point in time had to achieve the next goal. So, for a child with a score at or above the benchmark goal at a given point, the probability is high for achieving the next goal; the probability of need for additional support to achieve the next goal is low. Using the same analysis, progressive goals were identified at time points between the benchmark goals to help guide intermediate instructional planning within a grade level. For example, the benchmark goal for Initial Sound Fluency (ISF) is 25 by the middle of kindergarten. A student with a score of 8 on ISF at the *beginning* of kindergarten has met the progressive cutoff and is considered on track (i.e., odds > .80) to achieve the benchmark goal of 25 by the middle of the year.

The benchmark goals link progress on the measures to one another. For example, the ISF measure is an indicator of the child's knowledge and awareness of initial sounds in words, an aspect of phonemic awareness desired by winter of kindergarten if the child is on track for reading outcomes. If a child achieves a score of 25 on ISF by winter of kindergarten, the odds are at least .80 of achieving the next benchmark goal of 35 on Phoneme Segmentation Fluency (PSF) by spring of kindergarten. A student who achieves a score of 35 on PSF by the end of kindergarten has a greater than .80 probability of achieving the benchmark goal of 50 on Nonsense Word Fluency (NWF) by winter of first grade and so on. By spring of third grade, adequate progress on measures of ORF is necessary to be on track for high-stakes reading outcomes.

DIBELS benchmark goals represent powerful targets of instructional opportunity and meaningful goals to strive for when monitoring student progress. In addition to these minimum standards, ROC Curve analyses were also used to identify cutoff scores where the odds against achieving subsequent literacy goals would be indicated. In other words, cutoff points were scores at which 20% or fewer students in our sample achieved the subsequent goal. Students with scores at or below these cutoff points are extremely unlikely to meet subsequent early literacy goals unless additional instructional support is provided.

A unique feature of the DIBELS benchmark decision rules is the inclusion of a zone where a clear prediction is not possible. Scores that fall between the benchmark goal and the cutoff score represent patterns of performance where approximately 50% of students achieved subsequent literacy goals. Students with scores in this category require *strategic* planning on the part of educators to determine appropriate strategies to support the students to meet subsequent early literacy goals.

DIBELS data reports utilize specific terminology to describe students' probability of need for support based on their scores (see Table 2). The term used to describe the need for support for the group of students for whom the odds are in their favor (i.e., >.80 probability) of achieving subsequent goals is "benchmark." The descriptor of need for support for students below the cutoff point, for whom the probability is low (i.e., < .20) of achieving subsequent goals, is "intensive." The term used to describe the need for support for the middle group of students for whom it is difficult to make a prediction (.50 probability) is "strategic." (Good, Simmons, & Kame'enui, 2001).

The DIBELS goals and cutoff scores are empirically-derived, criterion-referenced scores. They indicate the probability of achieving the next benchmark goal or the probability of the need for additional instructional support for the student to achieve the next goal. Because the goals and cutoff scores are based on longitudinal predictive probabilities, they are not set in stone. A score at or above the benchmark indicates an 80% probability of achieving the next goal; but it is not a guarantee. Rather, we recommend that educators carefully consider the progress of all their students on all measures administered as they evaluate their instruction. Most students who meet a benchmark goal will need continued, high-quality instruction to hit the next target. However, the odds are that approximately 20% of students who achieve scores at or above the benchmark goal may still need supplemental support to achieve the next goal. Teachers will use additional information that they have about their students, as well as a pattern of performance across all of the DIBELS measures, to plan support for their students.

In this discussion of benchmark goals, it is important to reiterate that the DIBELS benchmark goals summarized in Table 1 represent minimal levels of performance for the lowest student – not goals for the average performing student. Effectively, the benchmark goals represent the minimal level of performance that a child can have, while the odds are still in their favor for becoming a reader. Additionally, each goal represents a steppingstone toward literacy—not a defensible endpoint. The sequence of goals builds upon a recommended skill sequence, with each goal supporting subsequent goals in the context of continued, high-quality differentiated instruction. The first goal is modest: for the student to be confident and fluent with the initial sounds of words, that the word "cat" starts with the sound /k/, for example. The DIBELS benchmark goals are the minimal level students need to achieve to be confident they are on track for literacy outcomes. The ultimate goal is for 100% of children within a school to achieve each benchmark.

Myth 7: *DIBELS are linked to a particular curricula or approach to teaching reading.*

Fact: *DIBELS measures are neutral with regard to particular programs, curricula, or instructional approaches. DIBELS were designed with an unremitting focus on the outcomes—children becoming skilled and fluent readers by the time they leave third grade.*

DIBELS have been designed to assess "basic early literacy skills" or the core components of reading: a) phonemic awareness, b) alphabetic principle, c) accuracy and fluency reading connected text, d) comprehension, and e) vocabulary/oral language (NRP, 2000). Yet DIBELS are neutral with respect to specific programs, curricula, and/or approaches to teaching reading. DIBELS may be used with *any* curriculum or program using any particular approach or strategy provided that students are being taught the component skills with sufficient depth, breadth, and integration to become successful readers. DIBELS have

Table 2

Probabilities of Meeting Goals and Need for Support with Corresponding DIBELS Descriptors

Probability of achieving subsequent goals	> .80	.50	< .20
Probability of need for support	Low	Moderate	High
Terminology used to describe need for support	Benchmark	Strategic	Intensive

been validated for this specific purpose. The authors are advocates of an approach to improving student learning where outcomes drive the decisions. In other words, "if it isn't broken, don't fix it." However, the same rule applies to situations where outcomes are unsatisfactory, "if you don't like what you've been getting, it's time to change what you've been doing." The authors of DIBELS do not advocate for any particular program but DO advocate for teaching the components of reading in such a way that all students become successful readers.

Myth 8: *There is no research on DIBELS.*

Fact: *Research on DIBELS has been ongoing since 1988 and continues to this day. This research is grounded in programs of research continuing for over 30 years, beginning with the validation of Curriculum-Based Measures procedures.*

Initially called CBM-P, DIBELS measures have been extensively evaluated through a series of studies at the University of Oregon (e.g., Good & Kaminski, 2002; Good, Kaminski, Smith, & Bratten, 2001; Good, Kaminski, Shinn, Bratten, Shinn, Laimon, Smith, & Flindt, 2004; Good, Simmons, & Kameenui, 2001; Good, Wallin, Simmons, Kameenui, & Kaminski, 1997; Kaminski & Good, 1996; Kaminski, Good, Shinn, Smith, Laimon, Shinn, & Bratten 2004; McKenna & Good, 2003). Since the early 2000s, research on the DIBELS measures has become even more widespread, and the measures are now regularly used in research and investigated elsewhere (Allor & McCathren, 2004; Barger, 2003; Buck & Torgesen, 2003; Hintze, Ryan, & Stoner, 2003; Rouse & Fantuzzo, 2006; Shaw & Shaw, 2002; Speece, Mills, Ritchey, & Hillman, 2003). Studies have included students from a range of schools including high and low income, rural, urban, inner city, suburban schools, and encompassing high and low proportions of ELL students, and schools with high and low proportions of students from diverse ethnic/racial backgrounds. Current psychometric data on DIBELS are presented in Table 3.

In addition to the psychometric properties of the DIBELS measures, there are substantial data being collected on an ongoing basis regarding the linkages between this type of formative assessment and both school and student achievement. For example, at the time of this publication, at least nine states have documented odds of .90 or greater of passing a third-grade state test for students who also read at least 110 words correct per minute by the spring of third grade (Castillo & Powell-Smith, 2003; Shapiro et al., 2006). Research on DIBELS is an ongoing effort to improve the technical adequacy of the measures as well as to increase the decision-making utility of the measures.

Misuses of DIBELS

In this paper we have described the most common myths associated with DIBELS, and the evidence that refutes those myths. In addition to noting the myths, however, we believe it is important to highlight two gross misuses of DIBELS that stem directly from these myths and then to discuss appropriate uses of the measures.

Table 3

Technical Adequacy of DIBELS Measures

DIBELS Measure	Grade	Reliability: Single Probe	Reliability: Multi-Probe	Reliability: Slope	Validity: Concurrent	Validity: Predictive	Sensitivity – Slope per week: Mean	Sensitivity – Slope per week: SD	Study
ISF	K	.61	.91	.35†	.36[a], .47[b]	.37[c], .36[d]	1.34†	0.81†	Good et al., 2004; †Good et al., 2000
PSF	K	.74, .88‡	.99‡	.88‡	.54[a], .65‡,[e]	.60[c], .58[f], .52[g]	1.12†	1.06†	Good et al., 2004; †Good et al., 2000; ‡Kaminski & Good, 1996
LNF	K	.89, .93‡	.99‡	.70‡	.70[a], .77‡,[e]	.66[c], .72[g]	.	.	Good et al., 2004; ‡Kaminski & Good, 1996
NWF	1st	.83	.94	.71†	.	.67[c], .75[h]	2.08†	1.62†	Good et al., 2004; †Good et al., 2000
WUF	1st	.65	.90	.55	.55[i], .71[j]	.	0.46	1.24	Kaminski et al., 2004
RTF	3rd	.	.	.	.45[d], .50[k]	.	.	.	McKenna, 2003

Note. Multi-probe reliability based on the mean of 5 probes for ISF and WUF and the mean of 3 alternate forms for PSF, NWF, and LNF. Reliability of slope estimates based on weekly repeated assessments for 10–11 weeks. When possible, median validity coefficients are reported. All data are reported from Good et al., 2004, unless otherwise noted.

[a]Woodcock-Johnson Psycho-Educational Battery Readiness Cluster score. [b]DIBELS PSF. [c]Woodcock-Johnson Psycho-Educational Battery Total Reading Cluster score. [d]CBM Oral Reading Fluency. [e]Metropolitan Readiness Test. [f]Winter 1st Grade NWF, median. [g]Spring 1st CBM ORF, median. [h]Overall median coefficient from CBM ORF from Winter of 1st to Spring of 2nd. [i]Test of Language Development. [j]Language Sample. [k]Oregon State Assessment Test.

Misuse 1: *DIBELS data are used for* high-stakes *decisions, both for individual students (e.g., labeling, tracking, grades, retention) and at a systems level (e.g., funding, teacher evaluation and job retention).*

Appropriate Use: *DIBELS were developed and validated for the purposes of providing effective support to individual students and by way of formative review of programs and student response to those programs.*

It has never been the intention of the developers of DIBELS that the data be used for labeling, tracking, or grading students or for high-stakes decisions such as retention. At a systems level, DIBELS were not intended to be used to evaluate individual teachers or be used for other systems-level high-stakes decisions, such as funding (Kaminski & Good, 1996).

The reasons against use of DIBELS for high-stakes decisions for individual children or at a systems level are three-fold. First, DIBELS were not validated for such uses. It is important to remember that although DIBELS have demonstrated technical adequacy for the purposes of screening and progress monitoring, they are one-minute measures administered at a single point in time. Many factors can impact a child's score and reliability of the measures is increased by repeated assessment over time (i.e., validating student need for support, progress monitoring). Best practice in assessment deems that assessment tools should be used for the purposes for which they were validated (AERA, 1999; APA, 1999).

Second, DIBELS data do not provide a comprehensive evaluation of the many aspects of a program that need to be considered when making high-stakes decisions. DIBELS are deliberately intended *not* to assess a wide range of individual skills related to a domain nor the many aspects of a school's program related to school success. Instead, DIBELS were designed to be *indicators* of five key early literacy skills that are predictive of later reading achievement. When aggregated, the data provide a broad snapshot of general program functioning and are an *indicator* of systems-wide successes/needs.

Finally, use of any single indicator of competence to make important decisions, such as child retention, teacher evaluation, or funding, violates professional standards of measurement (AERA, 1999; APA 1999). The importance of using other relevant information, including multiple forms of assessment, and viewing assessment results within the context of the school cannot be overstated.

It is our concern that the use of DIBELS data for high-stakes decisions will compromise instructional practices. Use of DIBELS data for high-stake decisions such as continuation of funding or employment also may encourage "cheating" in test administration and scoring. For example, "teaching the test" in ways that raise test scores but do not focus on needs of students and promote broader learning of critical skills (see also Misuse #2). Such practices defeat the intended purposes of DIBELS.

We support the need for accountability and the use of *multiple* measures of students' and schools' achievements and success for making decisions that may have serious consequences for teachers and schools. DIBELS' value is as a formative assessment and evaluation tool to help teachers and administrators to identify needs for support, whether those needs are at the individual student level or the system level. We recommend the following practices related to the use of DIBELS data as one piece of data within an accountability and/or evaluation *system*:

a. Establishment of system-wide goals/outcomes and alignment of DIBELS with goals/outcomes
b. Adequate training on administration, scoring, and interpretation of DIBELS data
c. Ongoing monitoring of test administration and scoring
d. Collection of formative data on program implementation and programmatic variables that impact student success
e. Ongoing and integrated professional development on Big Ideas of early literacy, instruction/intervention practices, interpretation and use of data

Misuse 2: *Teaching the test and/or artificially raising DIBELS scores without teaching the critical skills.*

Appropriate Use: *DIBELS are designed as indicators of an underlying basic early literacy skill. The DIBELS materials should never be used for practice or instructional purposes. The focus of instruction should be on the basic early literacy skill, not the test.*

Because DIBELS subtests are *indicators* of important skills, it is important to teach the skills *not* the test. DIBELS work well as a test, but the DIBELS test materials have not been designed for and should not be used for instruction or to practice. There are many effective research-based curricula and programs as well as fun and engaging activities that parents and teachers can use to teach phonemic awareness, the alphabetic principle, vocabulary, and comprehension in ways that will help a child become an accurate and fluent reader. A child who learns the underlying skills through integrated and meaningful early literacy activities does well on DIBELS and, more importantly, is on a trajectory to becoming a skilled reader. Teaching or practicing the test may raise DIBELS scores, but to do so artificially without changing the underlying skills may be harmful to children by preventing them from getting the support they need to be successful readers.

Concluding Comments

The bottom line is to teach the basic early literacy skills broadly, using a variety of curricula, programs, instructional strategies and approaches. Use DIBELS to assess efficiently. Keep doing what is working; change what is not, and let the outcomes drive the decisions. The purpose of DIBELS is to provide students with sufficient, appropriate, effective instruction that results in meaningful gains in performance to make all children readers.

References

Allor, J., & Mccathren, R. (2004). Learning Disabilities Practice The Efficacy of an Early Literacy Tutoring Program Implemented by College Students. *Learning Disabilities Research and Practice*, 19(2), 116-129.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Buck, J., & Torgesen, J. *The Relationship Between Performance on a Measure of Oral Reading Fluency and Performance on the Florida Comprehensive Assessment Test (FCRR Technical Report #1).* Tallahassee, FL: Florida Center for Reading Research.

Barger, J. (2003). *Comparing the DIBELS Oral Reading Fluency indicator and the North Carolina end of grade reading assessment (Technical Report).* Asheville, NC: North Carolina Teacher Academy.

Castillo, J. M., & Powell-Smith, K. A. (2005). The Across-Year Predictive Validity of Three Reading Fluency Measures on a Statewide Reading Accountability Test. University of South Florida.

Chiappe, P., Siegel, L. S., & Wade-Woolley, L. (2002). Linguistic diversity and the development of reading skills: A longitudinal study. *Scientific Studies of Reading*, 6(4), 369-400.

Corcoran Nielsen, D., & Luetke-Stahlman, B. (2002). Phonological awareness: one key to the reading proficiency of deaf children. American Annals of the Deaf, 147(3), 11 – 19.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children 52*(3), 219-232.

Deno, S.L. & Fuchs, L.S. (1987). Developing curriculum-based measurement systems for databased special education problem solving. *Focus on Exceptional Children, 19*(8), 1-15.

Deno, S. L., & Mirkin, P. K. (1977). The Contextual Framework. In S. L. Deno & P. K. Mirkin (Eds.), *Data-Based Programs Modification: A Manual* (pp. 5-174). Reston, VA: Council for Exceptional Children.

Fuchs, L.S. & Deno, S.L. (1994). Must instructionally useful performance assessment be based in the curiuclum? *Exceptional Children, 6*(1), 15-24.

Fuchs, D., & Fuchs, L.S. (2006). Introduction to response to intervention: what, why, and how valid is it? *Reading Research Quarterly, 41*(1), 93-99.

Geva, E., Yaghoub-Zadeh, Z., & Schuster, B. (2000). Understanding individual differences in word recognition skills of ESL children. *Annals of Dyslexia 50*, 123-154.

Good, R. H., Baker, S. K., & Peyton, J. A. (in press). Making sense of nonsense word fluency: Determining adequate progress in early first grade reading. *Reading and Writing Quarterly.*

Good III, R. H., Gruba, J., & Kaminski, R. A. (2001). Best Practices in Early Literacy Assessment: Dynamic Indicators of Basic Early Literacy Skills in an Outcomes-Driven Model. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology IV* (pp. 679-700). Washington, D.C.: National Association of School Psychologists.

Good III, R. H., & Kaminski, R. A. (2002). *Dynamic Indicators of Basic Early Literacy Skills: Administration and Scoring Guide*. Eugene, OR: University of Oregon.

Good, R.H., Kaminski, R.A., Bratten, J., & Smith, S. (2000). *Reliability, Validity, and Sensitivity of DIBELS.* Unpublished raw data.

Good III, R. H., Kaminski, R. A., Shinn, M. R., Bratten, J., Shinn, M., Laimon, D., Smith, S., Flindt, N.

(2004). *Technical Adequacy of DIBELS: Results of the Early Childhood Research Institute on measuring growth and development (Technical Report No. 7).* Eugene, OR: University of Oregon.

Good III, R. H., Kaminski, R. A., Smith, S., & Bratten, J. (2001) *Technical Adequacy of Second Grade DIBELS Oral Reading Fluency Passages (Technical Report #8).* Eugene, OR: University of Oregon.

Good III, R. H., Wallin, J., Simmons, D. C., Kameenui, E. J., & Kaminski, R. A. (2002). *System-wide Percentile Ranks for DIBELS Benchmark Assessment (Technical Report #9).* Eugene, OR: University of Oregon.

Gruba, G. G. (1997). *Evaluating Dynamic and Static Measurement Sensitivity to the Effects of a Phonological Awareness Intervention for Kindergarten Children.* University of Oregon, Eugene, OR.

Haager, D., & Windmueller, M. (2001). Early Reading Intervention for English Language Learners At-Risk for Learning Disabilities: Student and Teacher Outcomes in an Urban School. *Learning Disability Quarterly, 24*(4), 235-250.

Hintze, J. M., Ryan, A. L., & Stoner, G. (2003). Concurrent Validity and Diagnostic Accuracy of the Dynamic Indicators of Basic Early Literacy Skills and the Comprehensive Test of Phonological Processing. *School Psychology Review, 32*(4), 541-556.

Howell, K. & Nolet, V. (2000). *Curriculum-based evaluation: Teaching and decision making* (3rd edition). Stamford, CT: Wadsworth Publishing.

Hudson, R., F., Lane, H. B., & Pullen, P., C. (2005). Reading fluency assessment and instruction: What, why and how? *The Reading Teacher, 58*(8), 702-714.

Individuals with Disabilities Education Improvement Act of 2004, Pub. L. 108-466.

Johnson, D. S. (1996). *Assessment for the Prevention of Early Reading Problems: Utility of Dynamic Indicators of Basic Early Literacy Skills for Predicting Future Reading Performance.* University of Oregon, Eugene, OR.

Kaminski, R. A. (1992). *Assessment for the Primary Prevention of Early Academic Problems: Utility of Curriculum-Based Measurement Prereading Tasks.* University of Oregon, Eugene, OR.

Kaminski, R.A. & Cummings, K. D. (2007). Assessment for learning: using general outcomes measures. *Threshold*, Winter, 2007, 26-28. **Available**: http://ciconline.org/threshold.

Kaminski, R. A., & Good, R. H., III. (1998). Assessing early literacy skills in a Problem-Solving model: Dynamic Indicators of Basic Early Literacy Skills. In M. R. Shinn (Ed.), *Advanced applications of Curriculum-Based Measurement* (pp. 113-142). New York, NY: Guilford Press.

Kaminski, R. A., & Good, R. H., III. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review, 25*(2), 215-227.

Kaminski, R. A., Good III, R. H., Shinn, M. R., Smith, S. R., Laimon, D., Shinn, M., Bratten, J. (2004). *Development and Research on DIBELS Word Use Fluency Measure for First through Third Grades (Technical Report No. 13).* Eugene, OR: University of Oregon.

Koehler, K. M. (1996). *The Effects of Phonological Awareness and Letter Naming Fluency on Reading Acquisition for First-Graders Experiencing Difficulty Learning to Read.* University of Oregon, Eugene, OR.

Laimon, D. E. (1994). *The Effects of a Home-Based and Center-Based Intervention on At-Risk Preschool Children's Early Literacy Skills.* University of Oregon, Eugene, OR.

Lesaux, N. K., & Siegel, L. S. (2003). The Development of Reading in Children Who Speak English as a Second Language. *Developmental Psychology, 39*(6), 1005-1019.

Leybaert, J., Charlier, B. (1996). Visual Speech in the Head: The Effect of Cued-Speech on Rhyming, Remembering, and Spelling. *Journal of Deaf Education*, 1(4), 234-248.

Marston, D. B. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. R. Shinn (Ed.), *Curricul*um-based measurement: Assessing special children (pp. 18-78). New York, NY: Guilford Press.

McKenna, M. K., & Good III, R. H. (2003). *Assessing Reading Comprehension: the Relation Between DIBELS Oral Reading Fluency, DIBELS Retell Fluency, and Oregon State Assessment Scores (Technical Report).* Eugene, OR: University of Oregon.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction* (NIH Publication No. 00-4769). Washington, D.C.: U.S. Government Printing Office.

Otterstedt, J. R. H. (1993). *The Reliability and Validity of Rhyming and Onset Recognition Tasks as Measures of Phonological Awareness.* University of Oregon, Eugene, OR.

Paul, R. (2001). *Language Disorders from Infancy through Adolescence: Assessment & Intervention.* St. Louis: Mosby, Inc.

Pikulski, J., Chard, D., (2005). Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher, 58*(6), 510-518.

Rouse, H. L., & Fantuzzo, J.W. (2006). Validity of the Dynamic Indicators for Basic Early Literacy Skills an an indecator of early literacy for urban kindergarten children. *School Psychology Review, 35*(3), 341-355.

Shapiro, E. S., Keller, M. A., Lutz, J., Santoro, L. E., & Hintze, J. M. (2006). Curriculum-Based Measures and Performance on State Assessment and Standardized Tests: Reading and Math Performance in Pennsylvania. *Journal of Psychoeducational Assessment, 24*(1), 19-35.

Shaw, R., & Shaw, D. (2002). *DIBELS Oral Reading Fluency-Based Indicators of Third Grade Reading Skills for Colorado State Assessment Program (CSAP).* Eugene, OR: University of Oregon.

Shaywitz, S. (2003). *Overcoming Dyslexia: A New and Complete Science-Based Program for Reading Problems at Any Level.* New York: Alfred A. Knopf.

Shinn, M.R. (1989). *Curriculum-Based Measurement: Assessing Special Children.* New York: Guilford Press.

Speece, D.L., Mills, C., Ritchey, K.D., & Hillman, E. (2003). Initial evidence that letter fluency tasks are valid indicators of early reading skill. *The Journal of Special Education, 36*(4), 223-233.

