Running head: DIBELS ASSESSMENT

Best Practices in Using Dynamic Indicators of Basic Early Literacy Skills (DIBELS®) for Formative Assessment and Evaluation

Ruth Kaminski

Dynamic Measurement Group

Kelli D. Cummings

Kelly A. Powell-Smith

Dynamic Measurement Group

Roland H. Good III

University of Oregon

Best Practices in Using Dynamic Indicators of Basic Early Literacy Skills (DIBELS®) for Formative Assessment and Evaluation

Overview

Expectations for accountability have grown in education over the past decade. Under the auspices of recent federal legislation (e.g. NCLB and IDEA), schools must demonstrate that students with and without disabilities are accessing and benefiting from the general curricula. As educators increasingly are held responsible for student achievement, school personnel struggle to find ways to effectively document student responsiveness to interventions and track progress toward important outcomes. While many educators focus on state-wide and other high-stakes tests as a means of documenting student achievement of important outcomes, other assessment approaches may be better suited to assessing student progress. Teachers may use the results of these other assessment approaches proactively, to differentiate instruction and better support student learning. Assessment that is used to adapt teaching to meet student needs is called *formative assessment*. Because the primary purpose of formative assessment is to support student learning, it may arguably be considered the most important assessment practice in which educators engage.

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. DIBELS were designed for use in identifying children experiencing difficulty in the acquisition of basic early literacy skills in order to provide support early and prevent the occurrence of later reading difficulties. As part of the formative assessment process, DIBELS were designed 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, 1997; Kaminski & Good, 1996), which remain the intended uses of DIBELS to this date (Kaminski & Cummings, 2007). 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 Education Improvement Act (IDEA, 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).

In addition to the primary uses for individual educational decision-making, DIBELS may be used at a systems level. Administrators have access to data on all students in the system by collecting DIBELS data on all their students across a school year. These data can then be used to identify the percentage of students who are on track as well as the percentage of students who are making adequate progress. Aggregation of DIBELS data at the systems level may provide information regarding the effectiveness of the instructional supports within a classroom, school, or district. At the systems level, the process of *formative evaluation* helps administrators adapt resources broadly to meet the needs of all learners within the system.

The purpose of this chapter is to describe the use of DIBELS data within a data-based decision making model. Best practices include the use of DIBELS for formative assessment and evaluation within a multi-tiered system of service delivery and support.

Basic Considerations

DIBELS data can be used to help teachers and administrators in schools to intervene early, focus instruction on Basic Early Literacy Skills, and document progress in an ongoing manner to improve outcomes for all students. The appropriate and effective use of DIBELS data to inform these activities is essential to differentiating instruction for individual students and maximizing outcomes for all students. To use DIBELS data appropriately and effectively, it is important that school psychologists have a clear understanding of: (a) how DIBELS relate to core components of early literacy or Basic Early Literacy Skills, (b) DIBELS as General Outcome Measures, and (c) conceptual and empirical foundations of the decision-making utility of DIBELS benchmark goals.

Basic Early Literacy Skills

Converging and convincing evidence substantiates that reading competence is causally influenced by proficiency with foundational skills in beginning reading (National Reading Panel, 2000, National Research Council, 1998). We refer to these recognized and empirically validated skills related to reading outcomes as Basic Early Literacy Skills. Basic Early Literacy Skills, also known as core components or foundational skills, differentiate successful from less successful readers and, most important, are amenable to change through instruction (Kame'enui, Carnine, Dixon, Simmons, & Coyne, 2002; Simmons & Kame'enui, 1998). In the area of beginning reading, Basic Early Literacy Skills include: (a) phonemic awareness or the ability to hear and manipulate the individual sounds in words; (b) alphabetic principle or the mapping of print (letters) to speech (individual sounds) and the blending of these letter sounds into words; (c) accuracy and fluency with connected text; (d) vocabulary and oral language including the ability to understand and use words orally and in writing; and (e) comprehension (Adams, 1990;

National Reading Panel, 2000; National Research Council, 1998; Simmons & Kame'enui, 1998). An additional DIBELS measure, Letter Naming Fluency (LNF), is included for students in grades K and 1 as an indicator of risk. Unlike the other DIBELS measures, LNF does not measure a Basic Early Literacy Skill. Although letter names comprise a set a teachable skills, teaching letter names does not lead directly to improvements in student reading outcomes in the ways characterized by the foundational skills of early literacy (Adams, 1990). However, because the measure is highly predictive of later reading success, it is included as an indicator for students who may require additional instructional support on the Basic Early Literacy Skills in order to be successful readers. Table 1 provides an overview and brief description of the DIBELS measures and how they relate to the Basic Early Literacy Skills. Current psychometric data on DIBELS is presented in Table 2.

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 as General Outcomes 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 a general 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?"

With GOMs such as DIBELS, student performance on a common task is sampled over time to assess growth and development toward meaningful long-term outcomes. GOMs are deliberately intended *not* to be comprehensive and assess each individual skill related to a domain. Instead, GOMs measure key skills that are representative of and related to an important global outcome such as reading competence. GOMs include multiple alternate forms of approximately equal difficulty that sample these key skills. Also, the administration and scoring of GOMs is standardized so that the assessment procedures are delivered uniformly across

students. GOMs are efficient, generally taking from 1 to 5 minutes to administer and score yet provide data that are highly relevant to instructional planning. Finally, GOMs are highly sensitive to small, but important changes in student performance. Because of these design features, GOMS can be administered frequently over time. Differences in scores are attributable to student growth, not differences in the materials or assessment procedures so educators can compare assessment results over time.

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 brief so that our teaching doesn't have to be.

Decision-making Utility of DIBELS: Benchmark Goals

The DIBELS benchmark goals and timelines for achieving the goals are summarized in Table 3. 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. 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. Numerous additional studies have replicated the predictive utility of these goals 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.

For example, Initial Sound Fluency (ISF) 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. 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, 2005).

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 PSF is 35 by the end of kindergarten and, as indicated in Table 3, a student with a score of 18 on PSF by the *middle* of kindergarten is at the progressive cutoff and is considered on track (i.e., odds > .80) to achieve the benchmark goal of 35 by the end of kindergarten.

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 also were 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.

Terminology used in DIBELS data reports to describe the probability of need for support is provided in Table 4. 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 cut 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."

The DIBELS goals and cut 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 cut 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 3 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.

Best Practices in Use of DIBELS

DIBELS were developed to be inextricably linked to a model of data-based decision making. The Outcomes-Driven Model described here is based on foundational work with a problem-solving model (see Deno, 1989; Shinn, 1995, Tilly, this volume) and the initial application of the problem-solving model to early literacy skills (Kaminski & Good, 1998). The general questions addressed by a problem-solving model include: What is the problem? Why is it happening? What should be done about it? Did it work? (Tilly, this volume). The Outcomes-Driven Model was developed to address these questions, but within a prevention-oriented framework designed to pre-empt early reading difficulty and ensure step-by-step progress toward outcomes that will result in established, adequate reading achievement. The Outcomes-Driven Model accomplishes these goals through a set of five educational decisions: (1) identify need for support, (2) validate need for support, (3) plan support, (4) evaluate and modify support, and (5) review outcomes. A key premise of the Outcomes-Driven Model is prevention for all students. To prevent reading difficulties, educators must have access to a technology that specifies key outcomes and a timeline for achieving those goals. DIBELS data help educators to match the amount and type of instructional support with the needs of individual students to enable all students to reach each benchmark step. Some students will need substantial, intensive, individualized support to reach each step. Others will benefit from general, good, large-group,

classroom instruction. Some students may achieve important reading outcomes regardless of the instruction provided.

Identify System of Support

To use DIBELS data to improve outcomes for all students it is important to have in place a clearly defined system of support that encompasses all students. One such model is a multi-tiered model of support, often referred to as a "three-tier model (see Sugai, Horner & Gresham, 2002; Tilly, this volume). A three-tiered model of support is consistent with an Outcomes-Driven Model in that it is designed to provide a continuum of effective support options to meet the instructional needs of all learners. Additionally, a three-tiered model is a prevention model that is aimed at catching students early and providing the supports needed before students fall behind.

The model consists of having three levels of instruction that increase in intensity to match the needs of students with varying levels of support. Research-based interventions have been developed and validated to provide instructional support for each of the foundational early reading skills for which there is a corresponding benchmark goal. For more information on research-based interventions see Hosp, this volume; and Niebling, Roach, & Rahn-Blakeslee, this volume. Specific research-based interventions may be selected for each level of support, from broad-based programs designed to build the skills for most or all children in the class (tier 1 support) to high-intensity, individual or small group interventions (tier 3 support). With a three-tiered system of support in place, teachers do not need to develop individual plans for all students but can efficiently match student needs with already identified research-based support programs and strategies. When progress is monitored regularly, teachers may then use each student's data to differentiate instruction as needed and re-group as student's skill levels indicate.

Thus, groups should be conceived as being flexible with students moving in or out of groups as their instructional needs change.

The explication of a multi-tiered system of support is an integral component of a Response to Intervention (RTI) approach to identifying students in need of support. Evaluating response to intervention fits within and is, in fact, a key step within an Outcomes-Driven Model of decision-making. When implemented within a decision-making framework such as a problem-solving or outcomes-driven model, RTI is a means to providing early intervention to all children who may be at risk for school failure (Fuchs & Fuchs, 2006).

Identify Need for Support

The first step in the Outcomes-Driven Model is to identify children early who may need additional instructional support to meet a benchmark goal. To identify the need for support, DIBELS benchmark assessment is administered to all children in the school three times per year -- at the beginning, middle, and end of the school year. Periodic screening of all children helps to ensure that all children are making progress and to allow time to make changes in instruction for children who, in the middle of the year, are not on track to meet end of the year goals. In this way, benchmark assessment represents a prevention-oriented strategy that identifies students before a substantial, pervasive problem is evidenced.

DIBELS benchmark assessment is administered one-on-one and takes between 5 and 10 minutes per student depending upon the grade level and assessment period. The benchmark assessment provides data that are used to identify individual students who may need additional instructional support to achieve the next benchmark goal. The benchmark assessment also provides information regarding the performance of all children in the school with respect to benchmark goals.

Individual students who may need additional instructional support are those whose scores on one or more DIBELS measures falls below the benchmark or progressive goal (strategic need for support). For students whose scores fall below the cutoff, the odds are .20 or less of achieving the next benchmark goal (intensive need for support). The probability of need for support for these students to achieve the benchmark goal is high. For students whose scores fall between the cutoff and the goal, the odds are 50/50 of achieving the next benchmark goal. The probability of need for support for these students is moderate.

Table 5 provides data from a Class List Report showing DIBELS scores at the middle of kindergarten for a group of students. The Class List Report and other reports used for illustration purposes are from the University of Oregon DIBELS Data System (available at http://dibels.uoregon.edu). The reports are provided as examples of ways to summarize the data. Schools may use any of a variety of available data systems or enter school data and generate reports in a spreadsheet/data management program such as EXCEL. The specific data system used is not important; what is important is to have a means for summarizing the data, making it useful for educational decision making.

The Class List Report in Table 5 provides each student's raw score and percentile rank (based on school district norms) for each DIBELS measure administered at that time. In addition, the student's level of need for support (benchmark, strategic, intensive) based on the DIBELS goals and cut points is provided. The overall need for support (formerly called "instructional recommendation") is based on the conditional probability of achieving subsequent goals based on a student's pattern of scores across all measures administered at that time (Good, Kaminski, Smith, Simmons, Kame'enui, & Wallin, 2003).

In the class depicted in Table 5, the overall level need for support for Jasmine and Tyrell is Benchmark. Both students have met the middle-of-kindergarten benchmark of 25 on ISF, and have met the progressive benchmark for achieving the end of year PSF goal of 35 as indicated by PSF scores of 15 and 13 respectively. Additionally, Jasmine and Tyrell are both beginning to acquire alphabetic principle skills as indicated by scores of 27 and 42 on NWF and are performing in the low-risk range on LNF. Both Jasmine and Tyrell are considered to be on track for the development of early reading skills, and the probability of need for additional support is low. With continued high-quality instruction, the odds are in favor for Jasmine and Tyrell to achieve the next literacy goal. On the other hand, the overall need for support for Samuel, Maria, and Bethany is intensive. All three children scored below 10 on ISF and also scored low on PSF in the middle of kindergarten. Further, all three students scored below the cutoff for NWF. Based on their performance, the probability is high that Samuel, Maria, and Bethany may require intensive instructional support to achieve the end-of-kindergarten benchmark goal.

Three students, Desmond, Hillary, and Lewis have an overall need for support in the strategic range. Desmond scored below 10 on ISF on this assessment; however his scores on PSF indicate that his phoneme segmentation skills are emerging. He is beginning to learn some letter sounds and his score for LNF is in the some-risk range. The probability the Desmond may need additional support to achieve the spring early-literacy goal is moderate. Note that while Hillary and Lewis have an overall need for support in the strategic range, their scores of 3 on PSF are very close to those of Samuel, Maria, and Bethany and indicate intensive need for support on phonemic awareness. Thus, Hillary and Lewis, whose overall need for support is categorized as *strategic*, may need intensive support to reach the goal of 35 of on Phoneme Segmentation Fluency by spring of kindergarten.

Whenever assessment data are collected at one point in time under one set of conditions, alternative reasons for low performance must be considered. For example, students may have a bad day, be ill, be confused by the directions, or be uncomfortable with an unfamiliar examiner rather than have an actual skill deficit. For students whose skills may be a concern as indicated by performance on the benchmark assessment, validation of need for instructional support is the next step.

Validate Need for Support

The next step in the Outcomes-Driven Model is to rule out alternative reasons for a child's poor performance, and be reasonably confident that the student needs additional instructional support. In addition to the alternative reasons for poor performance mentioned previously, we have encountered instances where a score was miscopied, entered incorrectly, or transposed. Before making individual student decisions, it is important to consider additional information beyond the initial data point obtained for benchmark testing purposes. The possibility of inaccurate scores is not unique to the DIBELS or GOMs. Inaccurate scores are possible with any standardized assessment and may be more likely with younger children. We therefore recommend examining a *pattern of performance* rather than a single score before making important, individual decisions about instructional support.

DIBELS are only one piece of information that teachers have access to when making educational decisions. Teachers may always use additional assessment information and knowledge about a student to validate a score and the decision as to whether or not to provide additional support. One option for validating scores is to retest the student using alternate forms of the DIBELS measures. One advantage of DIBELS over other standardized assessments is the ease with which a child's score can be rechecked. When validating the need for instructional

support, an examiner may conduct brief repeat assessments using alternate forms of the DIBELS measures. The progress-monitoring booklet for each measure includes alternate forms that can be used by any trained teacher to retest a child. Because assessments are each about 1 minute, and because each measure has over 20 alternate forms available, repeated measures for students of concern can be conducted with reasonable efficiency. By retesting students whose skills are of concern, confidence is increased that a low score is indicative of low skills rather than a bad day, illness, shyness, etc. Each retest examines a hypothesis about poor performance. If it seems plausible that a child was uncomfortable with an unfamiliar adult, the child can be retested with a familiar teacher or aide. If the child had a bad day or was sick, he or she can be retested another day.

In Ms. Smith's classroom, for example, five students scored below the cutoff on PSF. The scores of all five students were a cause for concern for Ms. Smith. Three of the five children did not meet the middle-of-kindergarten benchmark of 25 on ISF, and all of the children's low scores on PSF indicate that additional support in the domain of phonemic awareness may be necessary. Based on their classroom performance, Ms. Smith was confident of the DIBELS score for two of the students, Samuel and Maria. However, she wanted to make sure that the low performance demonstrated by Hillary, Lewis, and Bethany was not simply because an unfamiliar person tested them right after the winter break. To validate the benchmark results, Ms. Smith personally retested Hillary, Lewis, and Bethany with alternate forms of PSF on two different days the following week. Ms. Smith then examined each student's pattern of performance across the multiple assessments.

Three patterns of performance are possible on repeated assessments. First, if a child's scores are consistently low on two or three retests using different samples of behavior on

different days under different conditions, the teacher can be reasonably confident the child's skills are truly low. Second, if only the initial testing is low, and the child consistently scores in the desired range on subsequent retests, the first assessment may have been inaccurate (i.e., the child was having a bad day). Third, the child may be extremely variable in performance: fine one day, but very low the next. In this case, we are alerted to other factors potentially affecting the child's performance. For example, an ear infection may negatively affect a child's hearing on some days but not others, or a child's motivation and effort may be inconsistent from one day to the next.

Cover graphs from PSF Progress-monitoring booklets in Figures 1, 2, and 3 illustrate scores from repeated assessments for Bethany, Hillary, and Lewis. As depicted in Figure 1, Bethany's scores across the repeated assessments were consistent with her benchmark score. Ms. Smith concluded that Bethany's low score was not just a bad day and that she was in need of additional instructional support.

The pattern of repeated assessment scores for Lewis is presented in Figure 2. Lewis' score of 3 on PSF was of concern to Ms. Smith, however repeated assessments indicated a clear upward trend for Lewis with his second score being above the progressive benchmark goal of 18. Lewis' pattern of performance on the repeated assessments with DIBELS was consistent with his performance during instruction in class and the decision was made that Lewis did not require additional intensive support on phonological awareness skills. While not placing Lewis in the group receiving intensive support on phonemic awareness, Ms. Smith did elect to strategically monitor Lewis on PSF for three weeks to be sure that he was successfully acquiring phonemic awareness skills through the core classroom instruction.

Scores for repeated assessment for Hillary are presented in Figure 3. As with Bethany, the pattern of scores indicated a high probability of need for intensive instructional support to achieve the next goal. However, Ms. Smith was surprised at Hillary's consistent pattern of low scores as her other scores, for example on LNF, were similar to the scores of Lewis. Further, Ms. Smith thought Hillary was learning along with the rest of the class because she sat quietly and was very attentive during reading instruction. Validating scores, by looking at a pattern of performance, is a very powerful indicator of student need.

Plan and Implement Support

The next decision-making step in the Outcomes-Driven Model is to plan instructional support. A plan for providing additional instructional support should include (a) a clear instructional goal and timeline; (b) a focus on basic early literacy skills; (c) a plan for the amount and type of support the student is likely to need; (d) a specification of the logistics of who will teach using what instructional materials, when and where; and (e) a measurement plan to evaluate progress.

Clear Instructional Goal

For each DIBELS measure that is related to a Basic Early Literacy Skill, a benchmark goal has been established and validated (Table 3). Achievement of the goal means that the odds are in the student's favor of achieving the next goal, thus achievement of each benchmark goal is a step toward reading outcomes. Phonemic awareness goals for Bethany and Hillary and the rate of progress necessary for each child to meet the goals are illustrated in Figures 4 and 5. The aimline is the line connecting the child's current performance to the end-of-year goal. The aimline represents the expected rate of progress and can be used to monitor growth and evaluate

the effectiveness of instruction from week to week. Both students will need to demonstrate a rate of progress consistent with their aimlines if they are to meet the end-of-year goal of 35 on PSF.

Some students may have significant skill deficits in reading, and are beyond the specified time period for a particular benchmark goal. For students in grades 2 and beyond whose literacy skills are one or more years below grade level, progress monitoring materials are used that match the child's instructional level. However, if a student is progress monitored below grade level, it is necessary to accelerate progress beyond what would be expected. Therefore, the timeline for achieving the goal and the corresponding aimline need to be adjusted. In other words, a fourth grader may be monitored with second grade materials, yet the goal for the end of second grade (90 words read correct in one minute on DORF) would not be sufficient for the child to get back on track. The actual adjusted timeline for meeting the goal would need to be made individually and take into account other factors such as the student's starting point, pattern of performance, factors influencing the student's learning, and intensity of intervention provided. Procedures for setting ambitious but realistic goals have been developed for CBM and similar procedures may be used with DIBELS (Deno, Fuchs, Marston, & Shin, 2001; Fuchs, 1993; Fuchs & Fuchs, 2004; Speece, 2004)).

Focus on Basic Early Literacy Skills

The DIBELS benchmark goals represent powerful indicators of student progress, and provide information about global targets for instructional intervention. As outlined in Table 1, each DIBELS measure is an *indicator* of a Basic Early Literacy Skill. It is critical that school psychologists help teachers to understand the distinction between what and how to assess and what and how to teach. DIBELS allow educators to assess student's acquisition of Basic Early Literacy Skills <u>efficiently</u>. Teaching should focus <u>broadly</u> on the Basic Early Literacy Skill. For

example, research evidence suggests that NWF works well as an indicator of the alphabetic principle (Good, Baker, & Peyton 2007). However, using nonsense word reading as a means of assessing does not mean that teaching should focus on the reading of nonsense words. Rather, the focus of teaching should be the alphabetic principle, or the knowledge of letter-sound correspondences and the application of that knowledge of letter sounds to reading unfamiliar words (i.e., blending or recoding). Teaching of the alphabetic principle should be integrated with other, meaningful kindergarten and first grade reading activities (e.g., opportunities to apply these skills to reading connected text) as early as possible. As pointed out by Kaminski and Good (1996) "While the DIBELS measures may provide an indicator of a child's acquisition of early literacy skills, teachers should not limit instruction to only those skills that are measured by DIBELS. Teachers should provide a broad range of experiences with print as well as instruction in all of the skills that are known to be facilitative of early reading."

Amount and Type of Support

If a three-tiered system of support is in place, teachers have a variety of choices of programs, materials, and strategies to match the instructional needs of learners. For more information on aligning instruction with assessment, see Niebling, Roach, & Rahn-Blakeslee (this volume) and Hosp (this volume). Sometimes it is necessary to conduct further assessment to match a child's needs with instruction. For more information on curriculum-based evaluation, see the chapters by Howe, Hosp, and Kurn and by Hosp and MacConnell (this volume).

One way to achieve differentiation of instruction is to use DIBELS data to plan for flexible grouping of students based on student needs. When using differentiated, flexible grouping strategies, teachers consider the instructional needs of individual students and choose a

grouping strategy to facilitate optimum learning. It should be noted that, although students with similar learning needs may be grouped together for instruction in a particular content area (e.g., math, reading), flexible grouping is not the same as "tracking." Different groupings are used for different instructional purposes. Additionally, flexible groups are dynamic. The ongoing progress monitoring data provided by DIBELS informs teachers if the instructional support, including the particular grouping strategy, is working. Teachers use data about students' progress to make adjustments in amount and type of support, including group membership and/or grouping strategies to meet students' changing needs.

As educators determine appropriate instructional groups, it is our recommendation that they look beyond the overall support recommendations and consider student performance across DIBELS measures (i.e., the performance on each measure). The rationale for this thinking is threefold. First, as seen in Figure 6, students with the same overall support recommendation may have different instructional needs. Figure 6 shows a scatterplot depicting individual student scores on PSF and NWF in the beginning of first grade. Each circle represents an individual's scores on both measures. For example the lowest left-most circle represents the scores for a student who achieved a score of zero on PSF and a zero on NWF. The circles are color coded as to the overall support recommendation with white circles for intensive support, striped circles for strategic support, and black circles for benchmark support. As is evident from Figure 6, some students with strategic overall recommendations have scores more similar to students with intensive overall recommendation and it may be more efficacious to group these students together. At least one student whose overall recommendation for support is benchmark has scores much more similar to students in the strategic range. Thus, overall statements about need

for support are guidelines only, and may mask instructionally useful information about individual students.

Second, the measures that we attend to for instruction may differ at key time points throughout the school year. For instance, PSF has a threshold effect, where once students reach the goal of 35 correct sounds per minute their scores may actually decrease slightly as they begin to focus on acquiring new skills such as the alphabetic principle. Also, NWF and DORF both measure skills in the alphabetic principle domain, however children can begin to acquire skills as measured by DORF (i.e., reading sight words) yet continue to struggle with the alphabetic principle. This situation is a problem because students with underdeveloped alphabetic principle skills often experience fluency problems eventually as the complexity of text increases. For this reason, it is important to consider performance on both measures rather than assuming that a child is on-track based on an isolated skill set. Finally, even students with scores in a particular range may have different instructional needs. For example, a student who earns a score of 35 on NWF but is going through the probe sound-by-sound likely has different needs than a student who earns a score of 35 but has 10 words re-coded or read as whole words. The bottom line is that educators know more than the computer about planning support. Thus, while the overall need for support recommendations provide a good starting point for attending to these issues, educators should bring the knowledge and expertise they have to bear when planning groups and identifying skills to teach within those groups.

Evaluate and Modify Support

A key part of RTI and a key step in an Outcomes-Driven Model is a feedback loop in which the effectiveness of instructional support is evaluated and modified as needed for individual students. The goal of the feedback loop in an Outcomes-Driven Model is to find a

match between the support provided and the child's needs. Modifications in support are provided until the student's progress is sufficient to achieve the benchmark goal. A key premise at this stage is that a good plan is a powerful starting place for instruction, but an individual child's response to a particular curriculum or strategy, even a research-based curriculum or strategy, is unpredictable. No matter how good the plan, program, or strategy, if it is not supporting the student's progress toward the goal, it needs to be modified. Within a three-tier model, the instructional support becomes more intensive as a student moves through the tiers.

The first step in evaluating support is to establish a data collection plan for progress monitoring. In general, for students who are receiving substantial, intensive instructional support (i.e., tier 3), weekly monitoring of progress is recommended. For students who need less support (i.e., tier 2), less frequent monitoring may be sufficient – every other week or once a month. Of course, data collection plans need to be considered within the context of the school and the resources available. In a school in which a large number of students require intensive support, it may not be feasible to monitor progress on a weekly basis for all students in need of intensive support. In such a case, resources may be focused on enhancing the core program for all students to decrease the number of students who need intensive support.

The next step is to establish decision rules to use to evaluate the data. Broadly speaking, there are two general types of approaches for operationalizing the RTI construct. One approach is a strategy in which a standard protocol is implemented for a fixed duration of time (e.g., 10-15 weeks) after which a decision is made as to the "responsiveness of the student" to the intervention (e.g., McMaster, Fuchs, Fuchs, & Compton, 2005). Others advocate a problem-solving approach to identify and implement intervention strategies (e.g., Ikeda & Gustafson, 2002). Methods used to evaluate responsiveness across approach type may include performance

within a normative range on a norm-referenced test, performance above a benchmark on a criterion-referenced test, performance above the mean or median of other students receiving the same support, and/or slope of progress equal to or better than the mean of either a normative group or others receiving the same support (Fuchs & Fuchs, 2006).

We recommend using a goal-oriented rule for evaluating a student's response to intervention that is straightforward for teachers to understand and use. Decisions about a student's progress are based on comparisons of DIBELS scores that are plotted on a graph and the aimline, or expected rates of progress. We suggest that educational professionals consider instructional modifications when student performance falls below the aimline for *three consecutive points*. This recommended decision rule is based on early work with Curriculum-based Measurement (Fuchs, 1988, 1989) and precision teaching (White & Haring, 1980) and allows for a minimum of three data points to be gathered before any decision is made. As when validating a student's need for support, a pattern of performance is considered before making individual student decisions.

In addition to being a more reliable indicator of student skill level, ongoing progress monitoring data places an upper limit on the amount of time an ineffective intervention is allowed to continue. For example, if data are gathered weekly, and a student is flatlining, educators will have information useful for deciding how to modify instructional support within three weeks, a relatively short period of time. It is important to note, however, that modifying support does not necessarily mean discontinuing an intervention. Especially if a school is implementing an evidenced-based program, important alterable instructional variables ought to be considered before adopting an entirely different approach. These modifications may include,

but are not limited to, assessing fidelity of implementation, decreasing group size, spending more time on the content, or providing additional, explicit instructions and opportunities to practice.

As with all recommendations regarding the use of DIBELS, these decision rules should be considered guidelines. If, for example, a student displays three data points below the aimline but the data indicate a steep upward trend, the teacher may choose to continue to collect progress monitoring data for one or two more weeks before making a change. When modifications are considered, the teaching team should always use multiple sources of information to make important decisions about a student's instructional plan, including additional assessment information regarding the severity of the problem (e.g., end-of-unit tests, curriculum-based assessments), observations of the student during instruction, and fidelity of implementation of the intervention.

Figures 7 and 8 illustrate the step of evaluating and modifying instructional support for Bethany and Hillary. Because Ms. Smith wanted to be sure that the support was effective for Bethany and Hillary, she assessed their skills weekly on PSF. Using the progress-monitoring booklet for PSF, Ms. Smith administered a PSF probe to both students using alternate forms every week and plotted each student's performance on the graph on the front of that child's progress monitoring booklet.

Small-group instructional support was begun for both students at the beginning of February. In the school that Bethany and 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 for each student. Bethany's performance showed progress sufficient to achieve the goal by the end of the year, thus continued extra small

group practice was recommended for Bethany. She continued to receive support and be monitored once a week through the spring benchmarking period.

On the other hand, 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.

In general, it is recommended that support be continued until a student achieves at least 3 data points at or above the goal. If a decision is made to discontinue support, it is recommended that progress monitoring be continued weekly for at least one month to ensure that the child is able to maintain growth without the supplemental support. The frequency of progress monitoring will be faded gradually as the child's progress continues to be sufficient.

Review Outcomes

In the Outcomes-Driven Model, the bottom line is achievement of essential literacy outcomes. The purpose of the reviewing-outcomes step is to review the structure of supports the school has in place to achieve outcomes at both an individual-student level and at a systems

level. For individual students, the teacher must decide if the student has achieved the benchmark and no longer requires additional instructional support. At this stage, teachers will again review benchmark assessment data for each goal period using the same procedures as when identifying need for instructional support.

For both Bethany and Hillary, the final data points in Figures 7 and 8 correspond to their end-of-kindergarten benchmark scores of 41 and 38 respectively. The scores at the end of kindergarten indicate they have achieved the benchmark goal. Both students will continue to be assessed in the fall, winter, and spring of first grade to ensure that they stay on track to achieve the subsequent benchmark goals at the middle and end of first grade.

Reviewing Outcomes at the Systems Level

At a systems level, school- and/or district-based data on basic early literacy skills can shape and define the application of research-based programs, practices, and materials used within a three-tiered model of instructional support (Baker & Smith, 2001; Good, Kaminski, Simmons and Kame'enui, 2001; Good Simmons & Kame'enui, 2000). As with decision making for individual students, when used at the systems level DIBELS data should be used *formatively* to *identify needs for support* at a school level. By using an Outcomes-Driven Model to analyze school/district-based data on Basic Early Literacy Skills, a school team can evaluate the various components of their system of support in beginning reading including core and supplemental curricula and programs, instructional strategies and materials, and professional development.

Similar to formative assessment of individual students, systems-level DIBELS data is 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. There are several organizing questions for reviewing outcomes at the systems level:

1. What proportion of students are meeting benchmark goals?
2. How effective is our core curriculum and instruction (tier 1) in supporting students who are on track to achieve benchmark goals?
3. How effective is our system of supplemental programs and instruction (tier 2) in supporting students who need strategic support to achieve benchmark goals?
4. How effective is our system of intensive intervention (tier 3) in supporting students who need intensive support to make adequate progress toward benchmark goals?
5. How do the early literacy skills displayed by students at each grade and point in the year compare to literacy skills displayed by students the previous year?

As described earlier, the specification of a multi-tiered system of support is an integral component in effectively using DIBELS data within an Outcomes Driven Model and/or within an RTI approach implemented to improve outcomes for all students. When DIBELS benchmark data are collected on all children, it is possible to know how many students receiving different levels of support are achieving goals at each step. It is our view that an essential component of an RTI model is an evaluation or the effectiveness of the support or intervention being provided across the continuum. The analysis of the system-wide DIBELS data allows for evaluation of the effectiveness of each component of the three-tiered system including core and supplemental programs.

The DIBELS Summary of Effectiveness Report assists a district in reviewing outcomes at the systems level by providing a clear, vivid, bottom line assessment of outcomes. The overarching question answered is, “What proportion of students are meeting benchmark goals?”

The report further addresses the questions about effectiveness of curriculum and instruction by describing the number and proportion of students at each tier of support who achieve subsequent benchmark goals and/or make adequate progress toward achieving those goals.

Based on the current research literature regarding system sustainability (McIntosh et al., 2006) and effective core reading programs,a rough guideline for interpreting system-level data is that approximately 95% of the children who enter a grade at benchmark should stay at benchmark. Of the children who are identified as in need of strategic support, 80% should achieve the benchmark goal. Of the children who are identified as in need of intensive support, 80% should make *adequate progress*. Adequate progress for students receiving instruction on grade level is defined as moving from an intensive recommendation to either strategic or benchmark. It should be noted that the Summary of Effectiveness Report provides information about adequate progress for students receiving instruction at grade level. For students whose instructional level is one or more levels below grade level (e.g., a fourth grade student who is receiving instruction and is monitored at a second grade level) adequate progress will need to be examined individually. Adequate progress for these students is defined as moving at least two recommendation levels (e.g., intensive to benchmark) within that instructional grade level or one grade level (e.g., strategic at second grade level to strategic at third grade level).

Each Summary of Effectiveness Report covers the time period from one benchmark assessment period to the next, for example, from the beginning to middle of kindergarten, middle to end of kindergarten, beginning to middle of first grade etc. The report provides information about the progress of students receiving different levels of support from the beginning of the year to the middle or from the middle to the end of a school year. Table 6 shows a Summary of Effectiveness Report for the first semester of first grade. The leftmost column provides the

names of the school district and schools. The three columns to the right show the outcomes for students at each tier of support (i.e., intensive, strategic, and benchmark). Each row contains information about either the district or an individual school.

In this report, there are a total of 188 first grade students in the district. At the beginning of the year 120 of these 188 students were at benchmark, 48 students needed strategic support, and 20 students needed intensive support. Of the 120 students in the district who began first grade at benchmark, 87% achieved the middle of year benchmark goal on NWF. Sixteen percent of the students who began the year at benchmark achieved a score that placed them as needing strategic support and two students (2%) achieved mid-year scores placing them as needing intensive support. Of the students in the district who began the year in the strategic and intensive ranges, approximately one half achieved the middle of year benchmark goal and/or made adequate progress toward the goal (54% of strategic; 55% of intensive).

The scores for each school show patterns of relative strength across the tiers of support. Washington school, for example, has a strength in its core program in the first half of first grade. Of the students who enter first grade at benchmark, 95% achieved the benchmark goal on NWF at the middle of first grade. Washington also has a strength in its intensive support program. Eighty percent of the students receiving intensive support at Washington School made adequate progress (i.e., moved from intensive to strategic or benchmark). Relative to the core and intensive support program, Washington's strategic support program may need to be evaluated and enhanced. Of the students who begin first grade in the strategic range at Washington, only two thirds (67%) achieve the benchmark goal in the middle of first grade. Approximately one fifth of the students remain in the strategic range and 13% score in the intensive range.

The pattern displayed by Washington is not an uncommon one. When implementing system improvements, schools frequently begin with adoption of a research-based core and/or program for students most in need. It is not unusual for tier 2 students to "slip between the cracks." One way to enhance the strategic support at Washington school may be to provide greater differentiation of support for students scoring within the strategic range. As described previously, there is not a "one size fits all" support that is effective for students in the strategic range. A score in the strategic range means that the probability of achieving the next benchmark goal for these students is .5 or 50/50. Some of these students' scores are close to the benchmark, and for these students the core curriculum may be sufficient. Some of these students' scores are close to the intensive range, or may actually *be* in the intensive range, in one or more Basic Early Literacy Skills. In other words, some of these students whose overall need for support falls in the strategic range may, in fact, benefit from intensive support in one or more areas. Further, students who score in the strategic range may not only need a different intensity of instruction but a focus on different content. For example, some students scoring in the strategic range on NWF may need instruction on learning letter-sound correspondences while others may know their letter sounds and need instruction on phonological recoding (i.e., blending). Some students scoring in the strategic range on ORF may need fluency building while others may need to work on decoding in the context of reading connected text and others may need work on building a sight word vocabulary.

Garfield School, on the other hand, has a relative strength in its strategic support program. Eighty percent of the students in the strategic category at the beginning of the year at Garfield achieve the benchmark goal at the middle of first grade. However, both the benchmark and intensive programs at Garfield school may need to be enhanced. At Garfield school only

79% of students who begin the year at benchmark achieve the middle of the year benchmark; 60% of students who begin the year with intensive needs score in the strategic/benchmark range at the middle of the year. If a research-based core is in place, it may be that greater fidelity of instruction is required in the core. Alternatively, the core may be weak in some aspect of early phonics skills and need to be supplemented for all students. Likewise, it may be that greater fidelity of instruction is required for the intensive support program. On the other hand, perhaps greater attention to differentiation of core instruction for the lowest performing students may increase the effectiveness of that instruction.

At both Garfield and Washington, the number of students at benchmark increased from the beginning to the middle of the year, 8% and 15% respectively. Likewise, the number of students with strategic and intensive needs decreased at both schools, 8% and 17% respectively. This movement is in the right direction – some adjustments to the system of support may further accelerate this change from the middle to end of the year.

At McKinley school, the overall system of support should be evaluated. At McKinley, only two thirds (67%) of the students who begin the year at benchmark achieve the benchmark goal at the middle of kindergarten. Of the students who begin the year in the strategic range, only 8% achieve the benchmark goal--while 50% of students in the intensive range make adequate progress or achieve the benchmark goal. At McKinley, the proportion of students at benchmark decreased from 60% at the beginning of the year to 44% at the middle of the year. This pattern at McKinley is also not an uncommon one. It is seen most often when a school adopts an intensive support program in the context of a generally ineffective and/or weak core.

The final question to be asked when reviewing outcomes is, “how are we doing this year compared to last year?” A number of DIBELS reports provide this comparison information

including Histograms, Cross-Year Box Plots, and Summary Reports. Ultimately, schools and districts may use the DIBELS data to establish school- or district-wide goals for literacy growth. The first year of DIBELS data provides a baseline from which subsequent growth can be measured and ambitious and realistic goals set. Although our ultimate goal is to have all children achieving literacy goals, it is unrealistic for a school to progress from a baseline of 40% of students at benchmark to 95% of students at benchmark in one year. Data on a large number of schools and districts indicate that it is not uncommon to experience a 10% to 20% growth in the number of students achieving benchmark goals in kindergarten and first grade over the first two years of collecting DIBELS data. A gradual increase in scores across years is illustrated by the Cross-Year Box Plot in Figure 9.

The Cross-Year Box Plot in Figure 9 shows annual data on PSF from the mid- and end-of-year benchmark assessments for the school years 2003-2004 through 2006-2007. Each box shows the range of scores on PSF for that time period and year. The box represents the average scores, that is the range of scores from the twentieth to eightieth percentiles. The vertical line at the top of the box corresponds to scores in the above-average range, i.e., the eightieth to ninetyfifth percentiles. The vertical line at the bottom of the box corresponds to scores in the below-average range, or the fifth to twentieth percentiles. The horizontal line in the middle of the box represents the average score for that time period and year. Horizontal lines on the graph at the mid- and end-of-year show the benchmark and progressive benchmark goal levels for PSF.

A review of the Cross-Year Box Plot shows that from year 1 to year 2, mid-year scores did not go up; however at each successive year mid-year scores increased to 2006, when all but the lowest performing students met the progressive benchmark by mid-year. Likewise, scores on PSF increase at each end-of-year benchmark from 2003 to 2006, when almost all students met

the end-of-year benchmark. Thus, the district depicted in the Cross-Year Box Plot in Figure 9, progressed from 65% of students at benchmark at the end of kindergarten in their first year of collecting DIBELS data to 95% of students at benchmark at the end of kindergarten over a four-year period.

Summary

This chapter has described best practice in assessment of early literacy skills using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) within an Outcomes-Driven Model of educational decision making. The model is prevention-oriented and is designed to preempt reading difficulties and support all children to achieve adequate reading outcomes by the end of third grade. 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 (DIBELS) to be used to document growth toward outcomes, as well as a set of steps for using the data generated by the measures at both the individual and systems levels. At the individual level, DIBELS may be used to evaluate response to intervention within an RTI approach. At the systems level, DIBELS data may be used to evaluate the overall effectiveness of the system of support that is being implemented.

The Outcomes-Driven Model is intended to be a continuous, recursive model. At an individual level, the data are used formatively to monitor each student's progress toward an important goal, make modifications to the intervention when indicated, and evaluate the effectiveness of the support being provided. 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 school 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 increase outcomes are maintained; changes that decrease outcomes are abandoned. Because data are collected on an ongoing basis documenting growth toward important benchmark skills, instructional and or curricular modifications can be made in a timely fashion to ensure all children achieve the goal of becoming established readers.

ANNOTATED BIBLIOGRAPHY

Good, R. H., Kaminski, R. A., Shinn, M., Bratten, J., Shinn, M., Laimon, L., Smith,S., & Flindt, N. (2004). *Technical adequacy and decision making utility of DIBELS*. Technical Report. Eugene, OR: University of Oregon.

This technical report presents reliability and validity data on DIBELS from the *Early Childhood Research Institute on Measuring Growth and Development.* Available from: http://www.dibels.org.

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*(3), 257-288.

This article describes the empirical and theoretical foundations for the DIBELS benchmark goals. A series of linked, short-term longitudinal studies provide data on the predictive validity of the measures from kindergarten through third grade with the Oregon Statewide Assessment-Reading/Literature as a high-stakes reading outcome.

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

This article describes DIBELS as General Outcomes Measures to be used within a formative assessment process. Web-based resources are provided.

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

This article describes the initial DIBELS research on LNF and PSF, providing an overview of the underlying rationale for development of the measures.

DIBELS Data System Web Site. http://dibels.uoregon.edu

All DIBELS materials are available for free downloading from this website at the University of Oregon. The data system is a fee for service providing a data entry and reporting system for DIBELS data.

DIBELS Online Community. http://dibels.wgen.net

This website provides an online networking community for DIBELS users.

REFERENCES

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Baker, S., & Smith, S. (2001). Linking school assessments to research-based practices in beginning reading: Improving programs and outcomes for students with and without disabilities. *Teacher Education and Special Education, 24,* 315-322.

Castillo, J. M., & Powell-Smith, K. A. (2005, March). *Predicting outcomes on statewide reading assessments using reading fluency*. Paper presented at the National Association of School Psychologists Annual Convention, Atlanta, GA

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

Deno, S. L. (1989). Curriculum-Based Measurement and special education services: A fundamental and direct relationship. In M. R. Shinn (Ed.), *Curriculum-Based Measurement: Assessing special children* (pp. 1-17). New York: Guilford.

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., Fuchs, L. S., Marston, D., and Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review, 30* (4), 507-524.

Deno, S.L. & Mirkin, P.K. (1977). *Data-based program modification: A manual*. Reston VA: Council for Exceptional Children.

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.

Fuchs, L. S. (1993). Enhancing instructional programming and student achievement with curriculum-based measurement. In J. J. Kramer (Ed.), *Curriculum-based measurement* (pp. 65-103). Lincoln, NE: Buros Institute of Mental Measurements.

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

Fuchs, L. S. (1988). Effects of computer-managed instruction on teachers' implementation of systematic monitoring programs, student achievement, and student awareness of learning. *Journal of Educational Research*, 81, 294-304.

Fuchs, L. S. (1989). Evaluating solutions: Monitoring progress and revising intervention plans. In M. Shinn (Ed.), *Curriculum-based measurement: Assessing special children*. New York: Guilford.

Fuchs, L. S., & Fuchs, D. (2004). What is scientifically-based research on progress monitoring. **Available: http://www.studentprogress.org/weblibrary.asp**

Good, R.H., Baker, S.K., & Peyton, J.A. (2007). Making sense of nonsense word fluency. Determining adequate progress in early first grade reading. *Reading and Writing Quarterly*.

Good, R.H., & Kaminski, R.A. (1997). Assessment for instructional decisions: Toward a proactive/prevention model of decision-making for early literacy skills. *School Psychology Quarterly*

Good, R. H., Kaminski, R. A., Shinn, M., Bratten, J., Shinn, M., Laimon, L., Smith,S., & Flindt, N. (2004). *Technical adequacy and decision making utility of DIBELS*. Technical Report. Eugene, OR: University of Oregon.

Good, R.H., Kaminski, R.A., Simmons, D., & Kame'enui E. (2001). Using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in an outcomes-driven model: Steps to reading outcomes. *Oregon School Study Council (OSSC) Bulletin, 44*(1), 2-24.

Good, R.H., Kaminski, R. A., Smith, S.B., Simmons, D.R., Kame'enui, E., & Wallin, J. (2003). Reviewing outcomes: Using DIBELS to evaluate kindergarten curricula and interventions. In S. Vaughn & K.L. Briggs (Eds.) *Reading in the classroom: Systems for the observation of teaching and learning.* Baltimore, MD: Brookes.

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*(3), 257-288.

Good, R. H., Simmons, D. C., & Smith, S. (1998). Effective academic interventions in the United States: Evaluating and enhancing the acquisition of early reading skills. *School Psychology Review, 27,* 45-56.

Ikeda, M., & Gustafson, J.K. (2002). *Heartland AEA 11's problem solving process: impact on issues related to special education* (Research Report No. 2002-01). Johnston, IA: Heartland Area Education Agency 11.

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

Kame'enui, E. J., Carnine, D. W., Dixon, R. C., Simmons, D. C., & Coyne, M. D. (2002). *Effective teaching strategies that accommodate diverse learners (2nd ed.).* Upper Saddle River, NJ: Prentice Hall.

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. (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: Guilford.

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

McIntosh, K., Horner, R. H., Chard, D. J., Boland, J. B., & Good, R. H., III (2006). The use of reading and behavior screening measures to predict nonresponse to school-wide positive behavior support: A longitudinal analysis. *School Psychology Review, 35* (2), 275-291.

McMaster, K.L., Fuchs, D., Fuchs, L.S., & Compton, D.L. (2005). Responding to nonresponders: an experimental field trial of identification and intervention methods. *Exceptional Children, 71*, 445-463.

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: Reports of the subgroups*. Bethesda, MD: National Institute of Child Health and Human Development.

National Research Council. (1998*). Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Shinn, M. R. (1995). Best practices in curriculum-based measurement and its use in a problem-solving model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology III* (pp. 547-567). Washington, DC: National Association of School Psychologists.

Shinn, M.R. (Ed.) *Curriculum-based measurement: Assessing special children.* New York: Guilford.

Simmons, D. C., & Kame'enui, E. J. (Eds.). (1998). *What reading research tells us about children with diverse learning needs: Bases and basics*. Mahwah, NJ: Lawrence Erlbaum Associates.

Speece, D. (2004). How progress monitoring assists decision making in a response-to-intervention framework. **Available:** http://www.studentprogress.org/weblibrary.asp

White, O.R., & Haring, N.G. (1980). *Exceptional teaching* (2nd ed.). Columbus OH: Merrill.

Table 1

Overview of DIBELS Measures

DIBELS Measure	Basic Early Literacy Skill	Description	Benchmark Assessment Times	Benchmark Goal
Initial Sound Fluency	Phonemic Awareness	The examiner presents four pictures to the child, names each picture, and then asks the child to identify (i.e., point to or say) the picture that begins with the sound produced orally by the examiner. The child is also asked to orally produce the beginning sound for an orally presented word that matches one of the given pictures.	K: F, W	25 by middle of K
Phoneme Segmentation Fluency	Phonemic awareness	The examiner says words, the student is asked to say the individual phonemes for each word.	K: W, Sp 1: F, W, Sp	35 by end of K
Nonsense Word Fluency	Alphabetic principle	The student is presented an 8.5" x 11" sheet of paper with randomly ordered VC and CVC nonsense words (e.g., sig, rav, ov) and asked to read the words.	K: W, Sp 1: F, W, Sp 2: F	50 correct letter sounds with at least 15 words correctly recoded by middle of 1
Oral Reading Fluency	Accuracy and fluency reading connected text	The student is presented with a reading passage and asked to read aloud.	1: W, Sp 2 – 6: F, W, Sp	40 by end of 1, 90 by end of 2, 110 by end of 3, 118 by end of 4, 124 by end of 5, 125 by end of 6
Word Use Fluency	Vocabulary and oral language	The examiner says a word and the student is asked to use the word in a sentence.	K – 3: F, W, Sp	Use local percentiles
Retell Fluency	Reading comprehension	After reading a passage the student is asked to tell everything he/she can about the passage read.	1: W, Sp 2 – 6: F, W, Sp	ORF => benchmark for grade and number of words in retell => 25% words read

Table 2

Technical Adequacy of DIBELS Measures

Technical Adequacy of Dynamic Indicators of Basic Early Literacy Skills

	Reliability			Validity		Sensitivity – Slope per week		
DIBELS Measure	Single Probe	Multi-Probe	Slope	Concurrent	Predictive	Mean	SD	Study
ISF	.65	.90	.35	.36[a], .48[b]	.36[a], .45[c]	1.34	0.81	Good et al., 2004; Laimon, 1994
PSF	.88	.96	.71	.54[b], .65[d]	.62[d], 68[be]	1.12	1.06	Good et al., 2004; Kaminski & Good, 1992; 1996
NWF	.92	.98	.71	.59[b]	.66[a], .82[c]	2.08	1.62	Good et al., 2004
WUF	.62	.90	.50	.48[f], .72[g]		.47	1.24	Kaminski et al., 2004
RTF				.59[h]	.50[i]			McKenna & Good, 2003
LNF	.93	.98		.70[a], .77[d]	.65[a], .76[e], .81[c]			Good et al., 2004; Kaminski & Good, 1992; 1996

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.

[a]Woodcock-Johnson Psycho-Educational Battery Total Reading cluster score. [b]DIBELS PSF. [c]CBM Oral Reading Fluency. [d]Metropolitan Readiness Test. [e]Stanford Diagnostic Reading Test. [f]Test of Language Development. [g]Language Sample. [h]Oral Reading Fluency. [i]Oregon State Assessment Test.

Table 3

DIBELS Benchmark and Progressive Goals and Probability of Need for Support

Measure	Need for Support								
	Beginning of Year			Middle of Year			End of Year		
	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.

Table 4

Probabilities of Meeting Goals and Need for Support and 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

Table 5

Sample Middle-of-Year Kindergarten DIBELS Scores from Class List

	Initial Sound Fluency			Phoneme Segmentation Fluency			Nonsense Word Fluency			Letter Naming Fluency			
Student	Score	Percentile	Need for Support	Score	Percentile	Need for Support	Score	Percentile	Need for Support	Score	Percentile	Level of Risk	Overall Need for Support
Samuel	9	4	Benchmark	1	7	Intensive	0	9	Intensive	8	13	At	Intensive
Maria	7	2	Intensive	1	7	Intensive	0	9	Intensive	11	19	At	Intensive
Hillary	14	12	Strategic	2	9	Intensive	5	36	Strategic	29	46	Low	Strategic
Lewis	19	22	Strategic	3	11	Intensive	5	36	Strategic	35	59	Low	Strategic
Bethany	9	4	Intensive	3	11	Intensive	4	32	Intensive	24	35	Some	Intensive
Tyrell	42	86	Benchmark	13	31	Strategic	27	83	Benchmark	48	85	Low	Benchmark
Desmond	5	1	Intensive	14	33	Strategic	5	36	Strategic	21	28	Some	Strategic
Jasmine	38	75	Benchmark	15	35	Strategic	43	94	Benchmark	37	66	Low	Benchmark

Table 6

Sample Summary of Effectiveness Report

Beginning to Middle of First[1]	Beginning of Year Need for Support								
	Intensive			Strategic			Benchmark		
District									
Beginning of year total[2]	20 students (10%)			48 students (25%)			120 students (65%)		
Need for support at middle of year	Intensive	Strategic	Benchmark	Intensive	Strategic	Benchmark	Intensive	Strategic	Benchmark
# in each category	9	9	2	4	18	26	2	19	104
% in each category	(45%)	(45%)	(10%)	(8%)	(38%)	(54%)	(2%)	(16%)	(87%)
Middle of year total[3]	15 students (8%)			46 students (24%)			132 students (70%)		
Garfield									
Beginning of year total[2]	5 students (10%)			12 students (24%)			34 students (67%)		
Need for support at middle of year	Intensive	Strategic	Benchmark	Intensive	Strategic	Benchmark	Intensive	Strategic	Benchmark
# in each category	2	2	1	0	2	10	0	7	27
% in each category	(40%)	(40%)	(20%)	(0%)	(17%)	(83%)	(0%)	(21%)	(79%)
Middle of year total[3]	2 students (4%)			11 students (22%)			38 students (75%)		
McKinley									
Beginning of year total[2]	10 students (18%)			12 students (22%)			33 students (60%)		
Need for suport at middle of year	Intensive	Strategic	Benchmark	Intensive	Strategic	Benchmark	Intensive	Strategic	Benchmark
# in each category	5	4	1	1	10	1	1	10	22
% in each category	(50%)	(40%)	(10%)	(8%)	(83%)	(8%)	(3%)	(30%)	(67%)
Middle of year total[3]	7 students (13%)			24 students (44%)			24 students (44%)		
Washington									
Beginning of year total[2]	5 students (6%)			24 students (28%)			58 students (67%)		
Need for support at middle of year	Intensive	Strategic	Benchmark	Intensive	Strategic	Benchmark	Intensive	Strategic	Benchmark
# in each category	1	3	1	3	5	16	1	2	55
% in each category	(20%)	(60%)	(20%)	(13%)	(21%)	(67%)	(2%)	(3%)	(95%)
Middle of year total[3]	5 students (6%)			10 students (11%)			72 students (82%)		

Notes: [1] Beginning to Middle of First = Overall need for support at beginning of first to need for support (based on NWF) at middle of first. Beginning of year total = number of students (% of all) at each level of support at beginning of year. 2 Middle of year total = number of students (% of all) at each level of support at middle of year.

Figure Captions

Figure 1. Validating need for support for Bethany.

Figure 2. Validating need for support for Lewis.

Figure 3. Validating need for support for Hillary.

Figure 4. Goal and aimline for Bethany.

Figure 5. Goal and aimline for Hillary.

Figure 6. Scatterplot showing variability in need for support.

Figure 7. Evaluating support for Bethany.

Figure 8. Evaluating support for Hillary.

Figure 9. Cross-year boxplot for Kindergarten PSF.

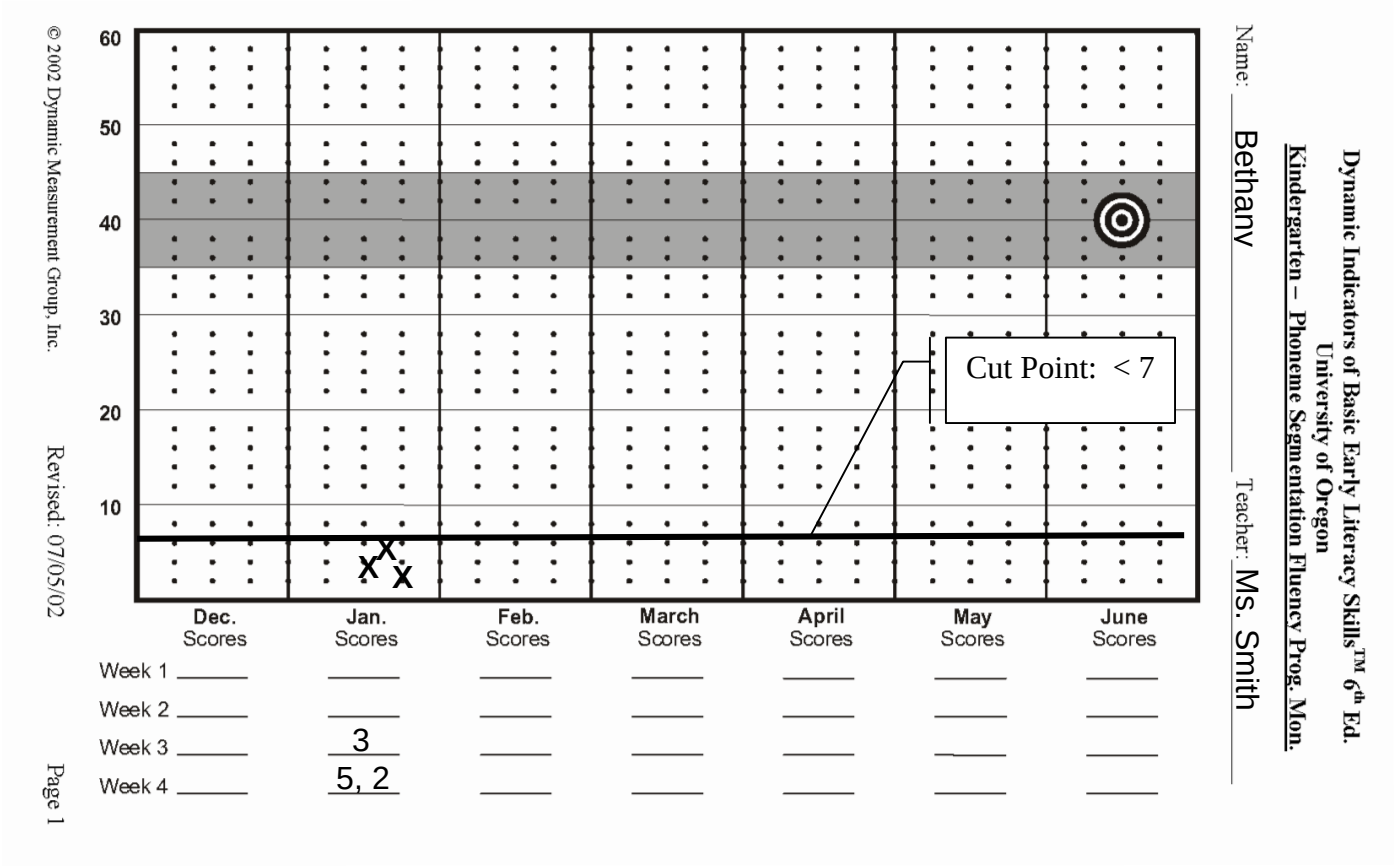
Dynamic Indicators of Basic Early Literacy Skills™ 6th Ed.
University of Oregon
Kindergarten – Phoneme Segmentation Fluency Prog. Mon.
Name: Bethany Teacher: Ms. Smith
60
50
40
30
20
10
Cut Point: < 7
Dec. Scores
Jan. Scores
Feb. Scores
March Scores
April Scores
May Scores
June Scores
Week 1
Week 2
Week 3
Week 4
3
5, 2
© 2002 Dynamic Measurement Group, Inc. Revised: 07/05/02 Page 1

Figure 1

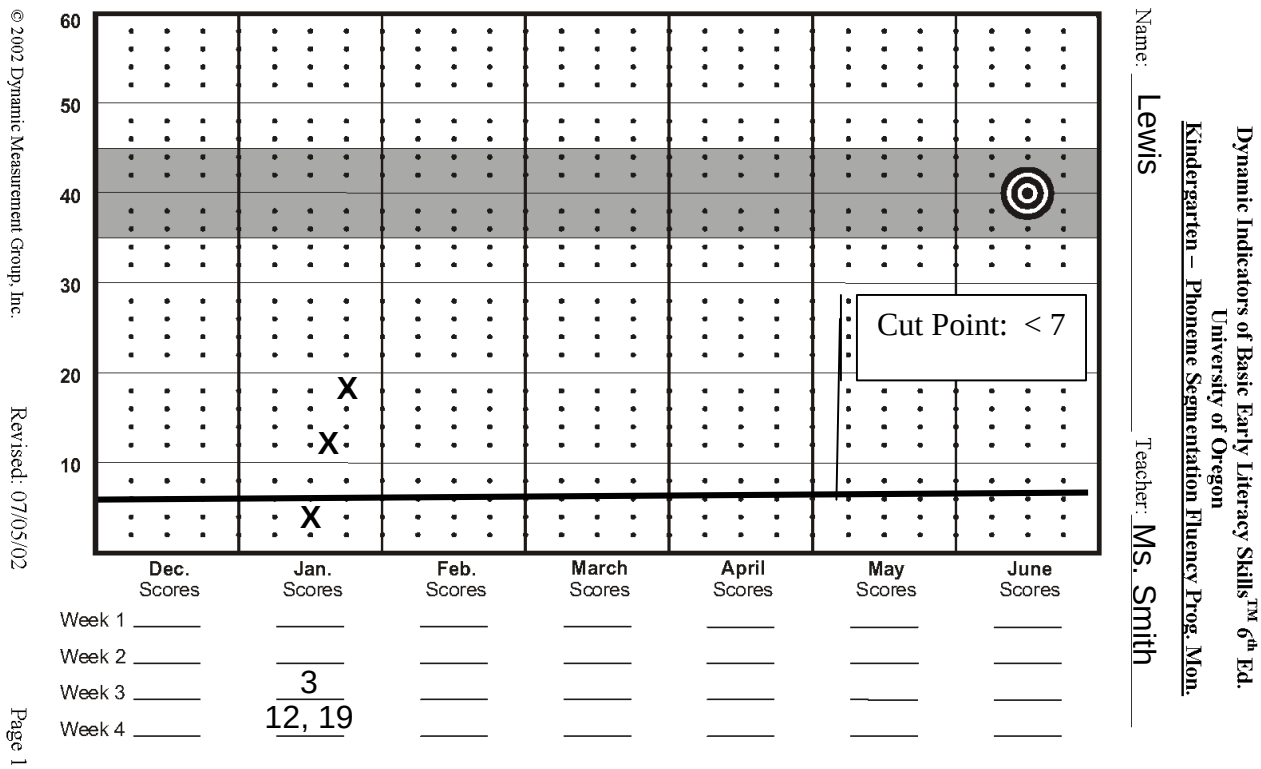
Dynamic Indicators of Basic Early Literacy Skills™ 6th Ed.
University of Oregon
Kindergarten – Phoneme Segmentation Fluency Prog. Mon.
Name: Lewis
Teacher: Ms. Smith
60
50
40
30
20
10
Cut Point: < 7
X
X
X
Dec. Scores
Jan. Scores
Feb. Scores
March Scores
April Scores
May Scores
June Scores
Week 1
Week 2
Week 3
Week 4
3
12, 19
© 2002 Dynamic Measurement Group, Inc.
Revised: 07/05/02
Page 1

Figure 2

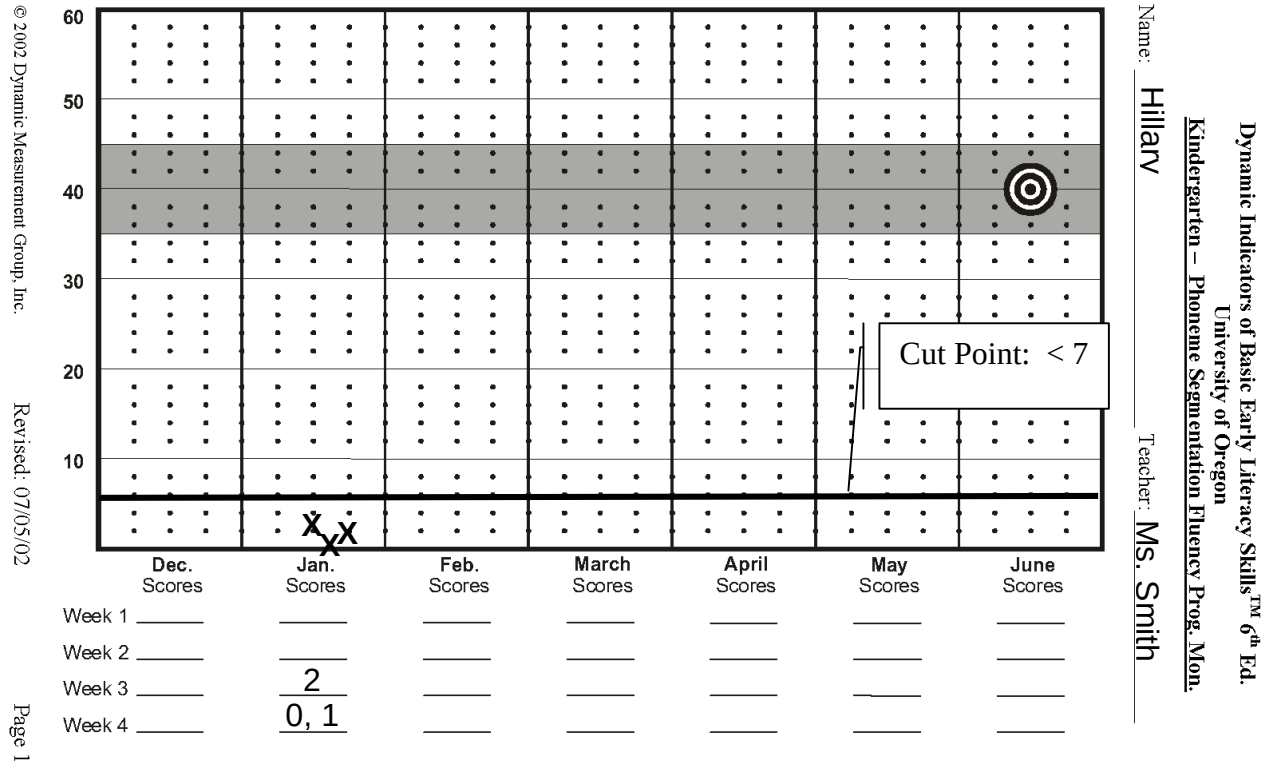
Dynamic Indicators of Basic Early Literacy Skills™ 6th Ed.
University of Oregon
Kindergarten – Phoneme Segmentation Fluency Prog. Mon.
Name: Hillary
Teacher: Ms. Smith
60
50
40
30
20
10
Cut Point: < 7
Dec. Scores
Jan. Scores
Feb. Scores
March Scores
April Scores
May Scores
June Scores
Week 1
Week 2
Week 3
Week 4
2
0, 1
© 2002 Dynamic Measurement Group, Inc.
Revised: 07/05/02
Page 1

Figure 3

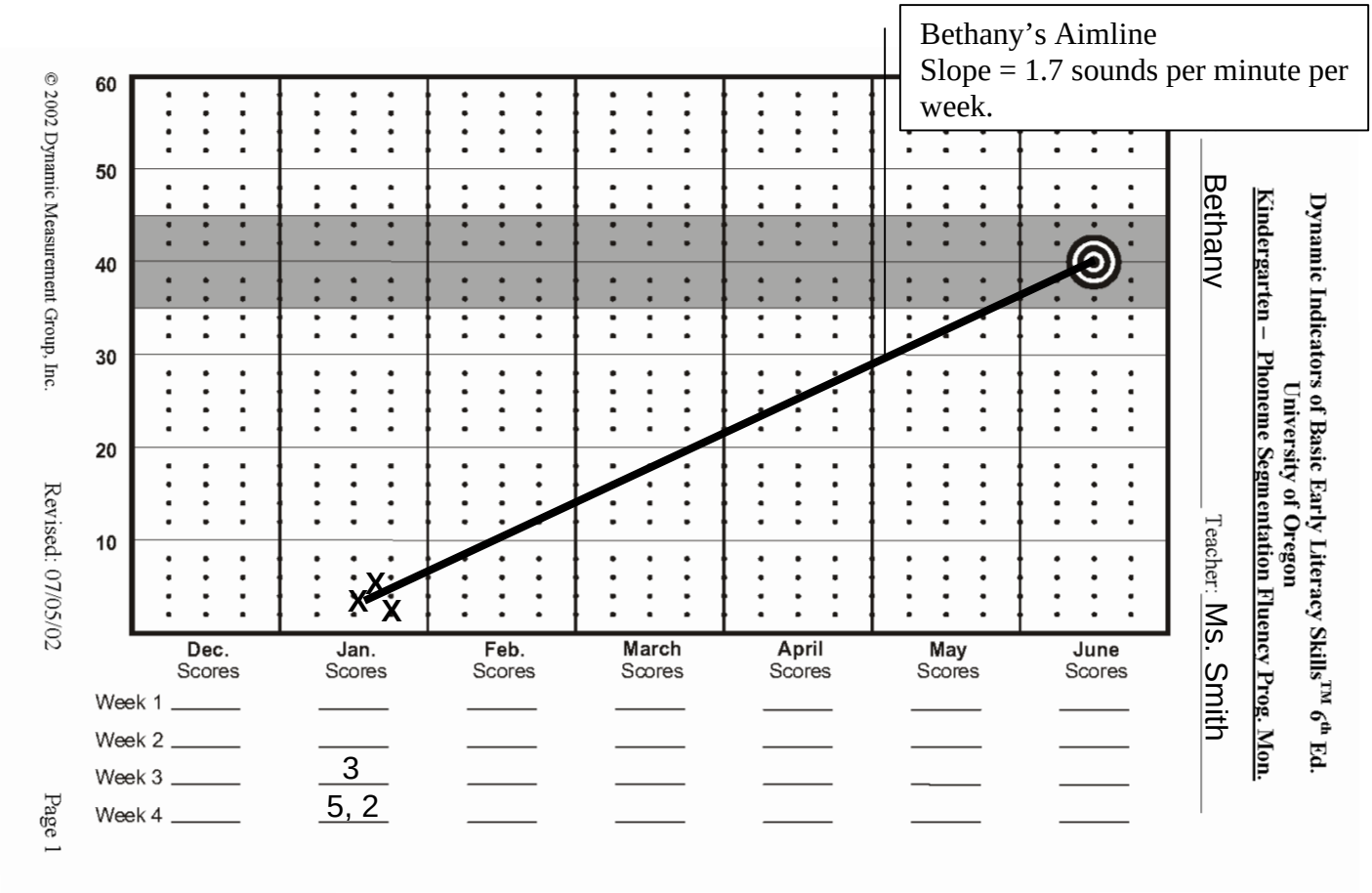
Bethany's Aimline
Slope = 1.7 sounds per minute per week.
Dynamic Indicators of Basic Early Literacy Skills™ 6th Ed.
University of Oregon
Kindergarten – Phoneme Segmentation Fluency Prog. Mon.
Bethany
Teacher: Ms. Smith
60
50
40
30
20
10
Dec. Scores
Jan. Scores
Feb. Scores
March Scores
April Scores
May Scores
June Scores
Week 1
Week 2
Week 3
Week 4
3
5, 2
© 2002 Dynamic Measurement Group, Inc.
Revised: 07/05/02
Page 1

Figure 4

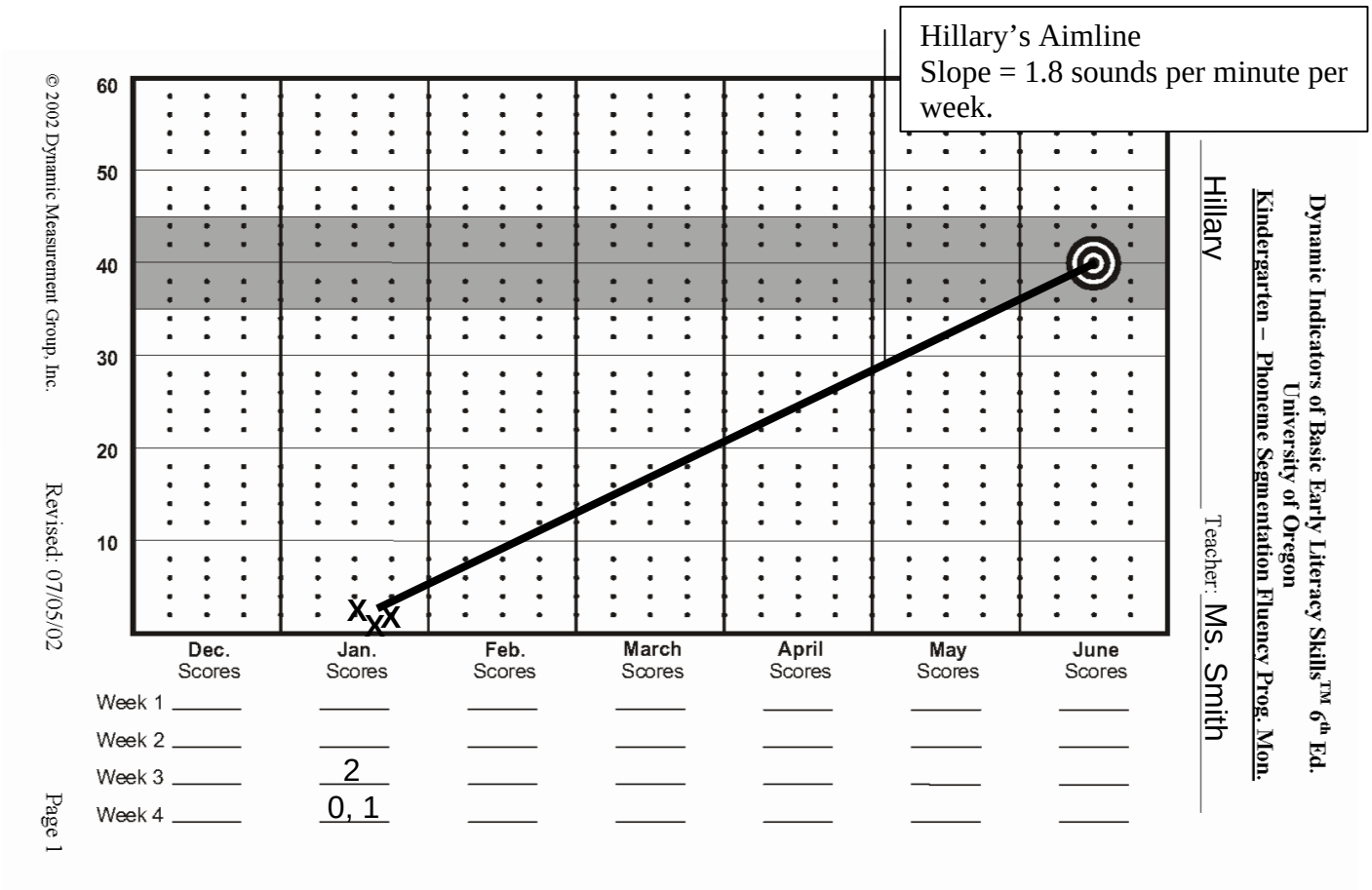
Hillary's Aimline
Slope = 1.8 sounds per minute per week.
Dynamic Indicators of Basic Early Literacy Skills™ 6th Ed.
University of Oregon
Kindergarten – Phoneme Segmentation Fluency Prog. Mon.
Hillary
Teacher: Ms. Smith
60
50
40
30
20
10
Dec. Scores
Jan. Scores
Feb. Scores
March Scores
April Scores
May Scores
June Scores
Week 1
Week 2
Week 3
Week 4
2
0, 1
© 2002 Dynamic Measurement Group, Inc.
Revised: 07/05/02
Page 1

Figure 5

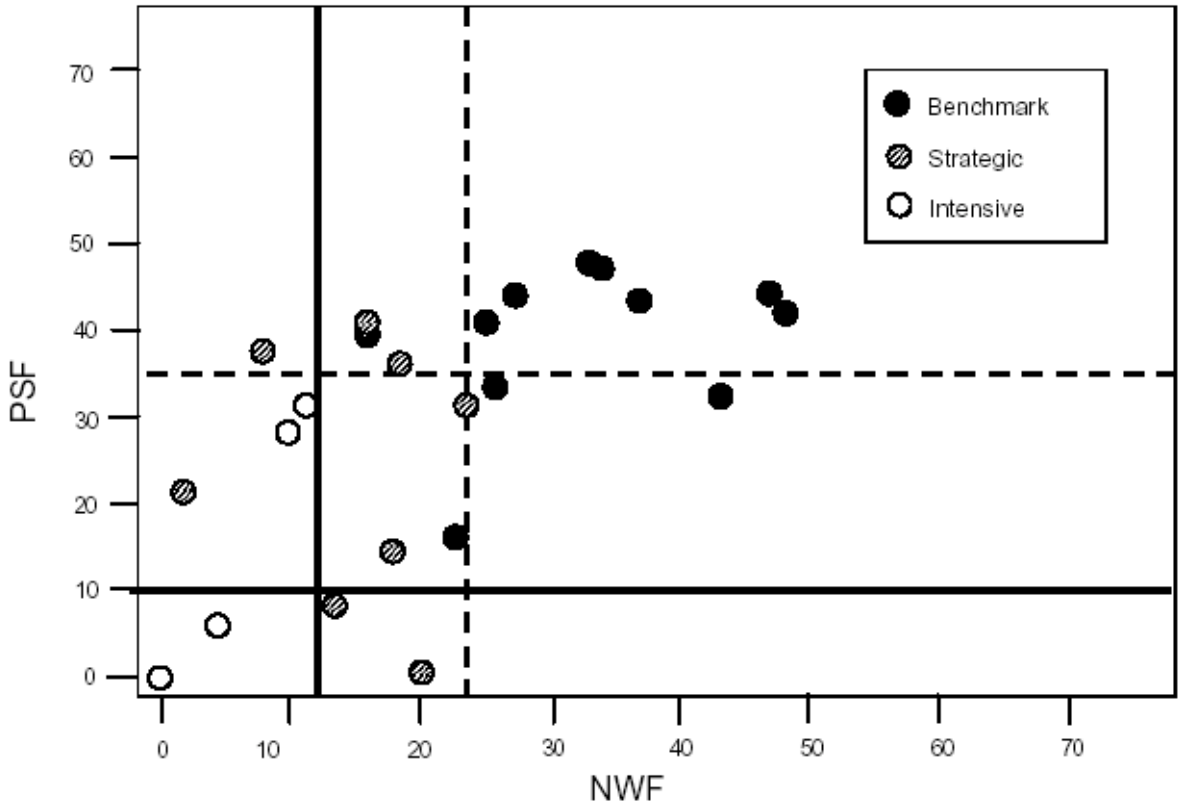

Figure 6

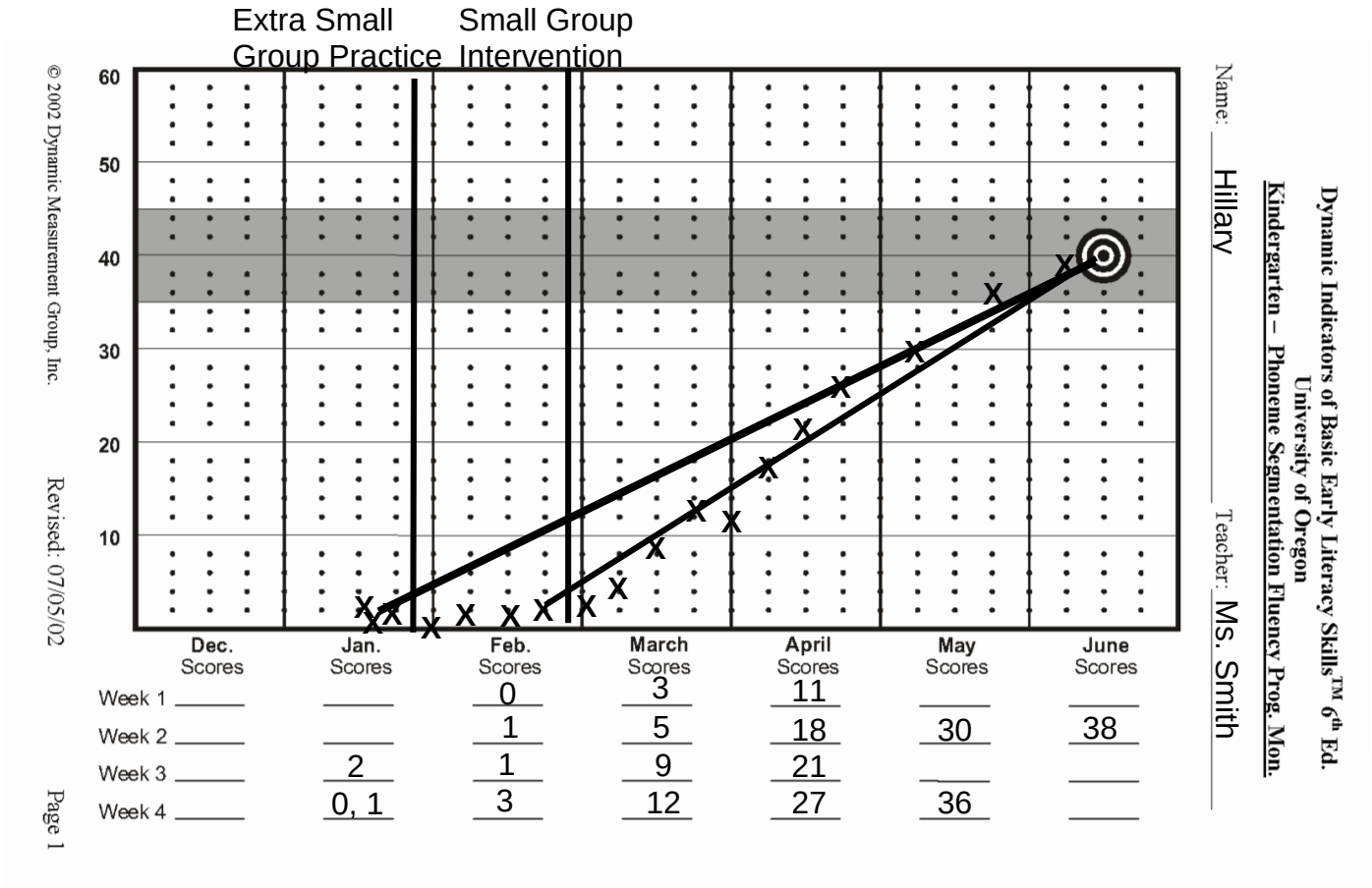
Extra Small Group Practice
Small Group Intervention
Dynamic Indicators of Basic Early Literacy Skills™ 6th Ed.
University of Oregon
Kindergarten – Phoneme Segmentation Fluency Prog. Mon.
Name: Hillary
Teacher: Ms. Smith
© 2002 Dynamic Measurement Group, Inc.
Revised: 07/05/02
Page 1
60
50
40
30
20
10
Dec. Scores
Jan. Scores
Feb. Scores
March Scores
April Scores
May Scores
June Scores
Week 1
Week 2
Week 3
Week 4
0
1
3
11
2
1
5
18
30
38
9
21
0, 1
3
12
27
36

Figure 7

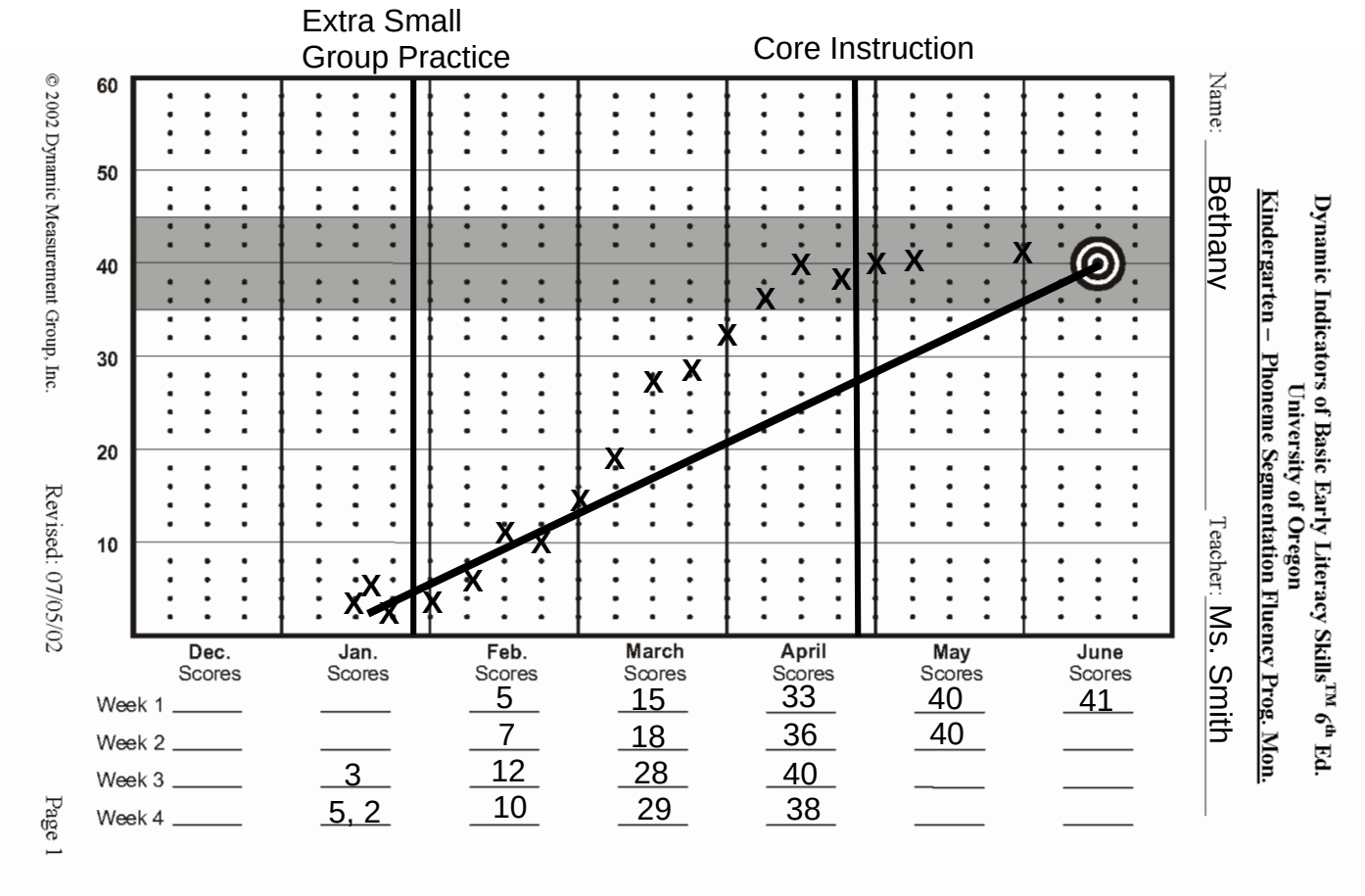
Extra Small
Group Practice
Core Instruction
Dynamic Indicators of Basic Early Literacy Skills™ 6th Ed.
University of Oregon
Kindergarten – Phoneme Segmentation Fluency Prog. Mon.
Name: Bethany Teacher: Ms. Smith
© 2002 Dynamic Measurement Group, Inc. Revised: 07/05/02 Page 1
60
50
40
30
20
10
Dec. Scores
Jan. Scores
Feb. Scores
March Scores
April Scores
May Scores
June Scores
Week 1 5 15 33 40 41
Week 2 7 18 36 40
Week 3 3 12 28 40
Week 4 5, 2 10 29 38

Figure 8

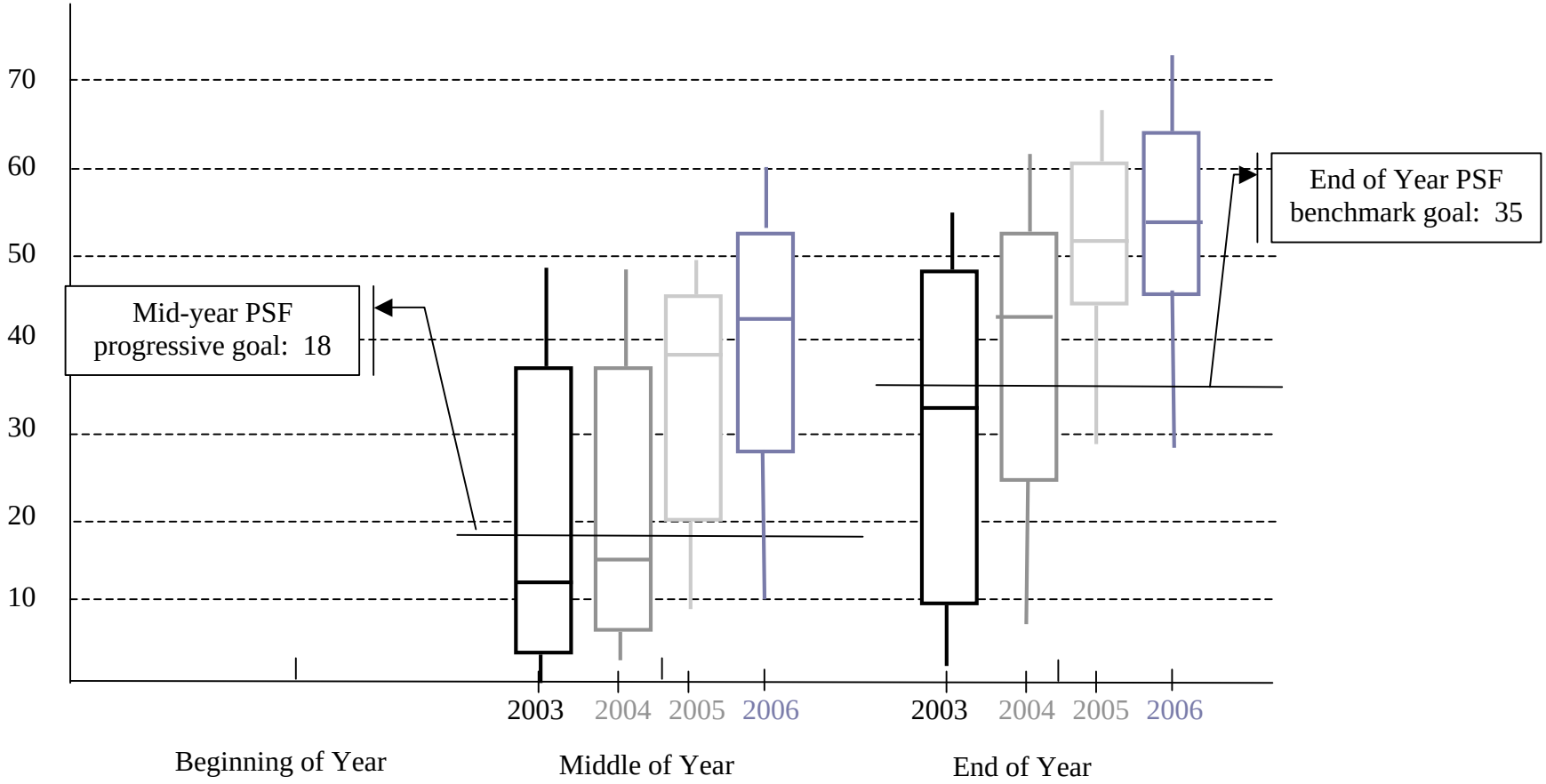
70
60
50
40
30
20
10
Mid-year PSF
progressive goal: 18
End of Year PSF
benchmark goal: 35
2003
2004
2005
2006
2003
2004
2005
2006
Beginning of Year
Middle of Year
End of Year

Figure 9