

Martin J. Ikeda

Eric Neessen

Iowa Department of Education

Joseph C. Witt

Louisiana State University and iSTEPP Learning

OVERVIEW

Universal screening is the systematic assessment of all children within a given class, grade, school building, or school district, on academic and/or social-emotional indicators that the school personnel and community have agreed are important (Deno, 2003; Iowa Department of Education, 2006; see Shinn, chapter 14, vol. 2). Universal screening assessment activities provide data to (a) help determine if there are enhancements needed in the core curriculum, instruction and/or general educational environment and (b) guide decisions about supplemental or intensive instruction for those students who may require instructional support beyond what is provided through core programming. Characteristics of universal screening are depicted in Table 1.

Universal screening practices are important for school psychologists for several reasons. Universal screening provides global indicators of the overall academic or social-emotional “health” of a system and the individuals within the system. Universal screening helps identify the prevalence and severity of problems in the system, so that resources can be applied efficiently. Importantly, the nature of universal screening, gathering data at several points in time, allows judgments to be made about effectiveness of supports provided, as well as the need to add or reduce supports provided to groups or individuals in that system. In addition, data can be used with parents as a more objective indication of how their child is doing in relation to grade-level standards and expectations, how their child’s performance

compares with other children in the same district, and how their child’s performance has changed over the course of the school year.

Universal screening is not a panacea. As described in this chapter, when not aligned to curriculum and instruction, and when not used in the context of general school improvement efforts, universal screening is essentially worthless. However, when part of a thoughtful discussion on how best to align resources to improve achievement, universal screening tests provide school personnel with a time- and cost-efficient means for identifying how effectively the general system is meeting the needs of all students within the system as well as identifying students potentially at risk for failure. Universal screening tests are not intended to diagnose academic problems, nor provide a direct link to intervention strategies.

Universal screening practices for education have received considerable press with the acceptance of response-to-intervention (RTI) practices in the 2004 reauthorization of the Individuals with Disabilities Education Act. The importance of universal screening is that students who are at risk for academic failure can be identified early and proactively. This dramatically improves the chances for successful academic outcomes for students over the traditional wait-to-fail process whereby a student is referred and intervention is initiated only after the achievement gap is large (Deno, 2003). Depending upon the results of the screening, preventative strategies occur with large or small groups of students with similar needs or with individual students.

Table 1. Characteristics of Universal Screening

What universal screening is	What universal screening is not
Administered to an entire classroom/grade/school/district, using either individual or group format	Administered to an isolated group
Used to inform instruction	Program placement or tracking
An indicator that there is a potential problem in need of further investigation	Diagnostic: Intended to tell you what the problem is
An answer to the question: Which and how many students are potentially in need of additional support?	Does not answer the question: What program does the student need?
Quantitative (gives a number for a score)	Qualitative (anecdotal)
One piece of information required to determine additional instructional needs	Intended to be used as a sole indicator to identify students as candidates for any program, such as Title I, special education, gifted and talented, or English as a second language
Cost effective in terms of teacher time, student time, and dollar cost of the test and scoring	Expensive or labor intensive
Administered, scored, and interpreted in a consistent manner	Administered, scored, and interpreted randomly or subjectively
Relevant for pre-K–12 grades	Limited for use in primary grades
Aligned with instruction, standards, and benchmarks	Arbitrarily selected
A sampling of one point in time	Intended to monitor individual progress
A sample of a skill	Intended to completely measure every skill
Capable of identifying students who exceed or fall below the learning expectations/standards	Used unless there are specific cut-off points
Results in an action or decision: information must be used	Used for accountability purposes only
Easily administered, scored, and interpreted	Time consuming to administer or difficult to score
Used proactively	Used reactively

Note. Source: Iowa Department of Education (2006).

An analogy illustrating the importance of matching the health of an educational system to measurement is that of an annual physical. A physician examines global indicators of health such as weight, temperature, and blood pressure. If the patient’s indicators are in the normal range for his or her age, the physician has little need to see the patient until the next physical. For younger adults, the risk factors for disease are lower, so a physical is needed every 2 or more years. Older adults are at higher risk for diseases like cancer, so more frequent and thorough physicals are needed. In some cases, blood samples are drawn for basic tests such as those that measure sugar, cholesterol, and iron levels. For patients whose tests indicate levels in the normal range, no action is needed. For patients with high sugar, iron, or cholesterol levels, the physician may prescribe more diagnostic testing or may prescribe a specific treatment regimen. Monthly tests may be ordered to monitor the patient until the condition stabilizes.

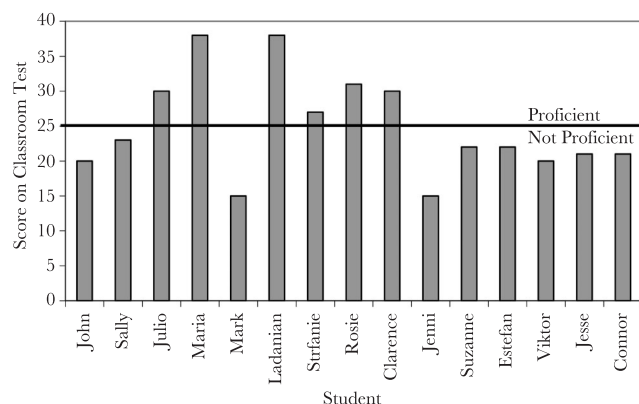
With universal screening of academics, children are given brief measures of academic and/or behavioral competency, and data are analyzed at classroom or school-building levels to understand (a) how many students are responding to the instruction, (b) if the current instruction is effective, (c) how many students are at risk for failure, and (d) which students need

additional assessments. Results of universal screening lead to discussions around the kinds of interventions that can be implemented or which diagnostic tests are needed to better understand why the problems are occurring.

Screening is a major asset in the early stages of problem analysis. The school-based team, in discussing concerns around individual students, has an option in considering whether an intervention should be student focused or system focused. System-focused strategies examine the *interaction* between instruction, as well as individual student’s and collective students’ response to instruction. Implicit in student-focused dialog is the question: What is wrong with the student? Historically, the most frequent response has been: “The student has a learning disability.” With universal screening, depending on the data, responses vary from: “Because there is a mismatch between our curriculum, instruction, and characteristics of learners in our setting” (collective students) to “Because there is a mismatch between an individual child, instruction, and the curriculum” (individual student).

Universal screening data from a classroom, such as the one in Figure 1, allows for system-focused discussion to more easily occur: “Yes, the student (Suzanne) is not at proficiency, but there are many other students scoring

Figure 1. Data indicating class-wide problem.



below the criterion line as well.” Traditional practices, in the absence of universal screening and ignorant of the classroom context in Figure 1, may have triggered a referral for special education evaluation for Suzanne because she is functioning at a very low level. However, there are system-focused questions about instruction in this situation that must be addressed.

BASIC CONSIDERATIONS

Universal screening is part of a decision-making process to determine which students are at risk for failure and which students are not. As with any decision-making framework, probabilities, rather than absolutes, are addressed. Universal screening tests, depending on the technical characteristics of the test, will accurately classify some students and misclassify other students.

In particular, in the era of No Child Left Behind (NCLB), school administrators and teachers do not have the luxury of waiting until failure occurs and then lamenting, “We should have intervened 2 years ago!” Instead, universal screening tests guide decision making using the *least dangerous assumption*. That is, it is better to identify a student as at risk for failure and provide some level of additional learning supports than it is to classify a student as “fine” when the reality is that the student will not be successful in the absence of additional support.

A crosswalk of two realities is depicted in Figure 2: results of the screening test as either at risk or not at risk of failure and the true state of the individual as needing or not needing intervention support. In the top-left cell, students are identified as needing intervention support and receiving intervention support. This is called a true positive. Test results indicate the students are at risk for failure and in reality the students do need support. In

Figure 2. True and false positives and negatives.

		Test result	
		At risk	Not at risk
Individual's state	Needs support	True positive	False negative
	Does not need support	False positive	True negative

the top-right cell, screening test results indicate that students are not at risk for failure but in reality are in need of support. These are false negatives. The bottom-left cell represents false positives, students identified as at risk for failure but in reality do not need support. In the bottom-right cell, students are screened as not at risk and the reality is that the students do not need support. These students are true negatives. Since a major purpose of screening is to detect students who are at risk, generally false negative errors are the most serious because screening has said there is not a problem when, in fact, a problem does exist. False positive errors can also be problematic in that precious intervention resources may be allocated to students who may not have a serious need for intervention.

False positive errors can also be problematic if not achieving a benchmark initiates actions that inappropriately lead to classification. Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good, Gruba, & Kaminski, 2002), for example, have rigorous benchmarks designed so that a student who achieves benchmark should have a satisfactory score on state tests. A failure to meet benchmarks, however, does not necessarily mean the student has a disability or even a deficit and, in fact, the student may be at grade level when administered a traditional norm-referenced reading assessment.

The supporters of universal screening argue that universal screening is effective for helping schools align resources and that universal screening addresses some shortcomings of teacher referral. There is evidence that, while students identified by teachers tend to have problems, there are often times other students with equally severe problems are not referred (VanDerHeyden & Witt, 2005). Supporters of universal screening argue that although universal screening may result in higher numbers of false positives (students who look like they need support but who do not) relying on teacher referral may result in large numbers of false negatives (students who have problems but are not reported as such).

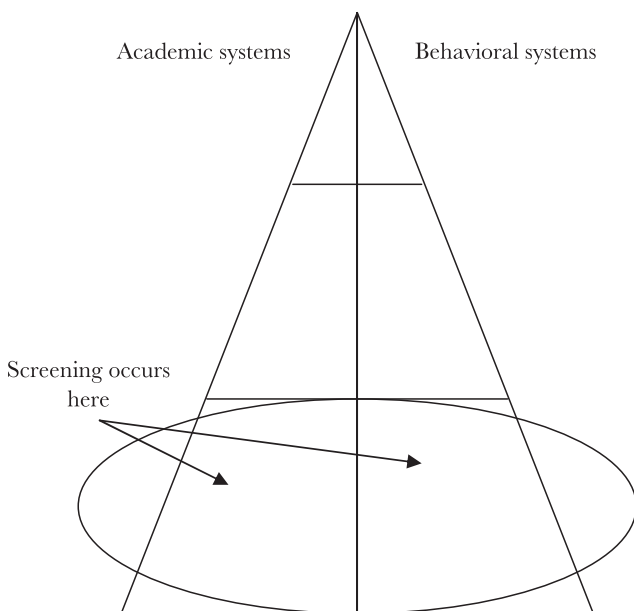
Existence Within a Tiered Framework

Universal screening as part of a tiered system is illustrated in Figure 3. Universal screening activities, both in academic skill areas and in social-emotional areas, are done at the base of the triangle. Universal screening results are an indicator of the overall alignment of instructional frameworks and the skills assessed, as well as an indicator of the overall health of the system.

③ Tilly (see chapter 2, vol. 1) describes the movement toward aligned service delivery systems in which a plethora of instructional options are available to address student performance problems. Additionally, *School Psychology: A Blueprint for Training and Practice III* (Ysseldyke et al., 2006) describes an educational system in which academic competence and identification for the purposes of early intervention are cornerstones rather than afterthoughts. In fact, *Blueprint III* declares that, “all assessment activities should relate to prevention and intervention” (p. 18).

Universal screening is related to prevention and intervention and is a necessary, but not sufficient, component of this kind of aligned service delivery system. An aligned service delivery system consists of complementary curriculum, instruction, and assessment. Each component needs to be studied for adequacy, and change in one component has an impact on the other components. Hence, universal screening, while an

Figure 3. Universal screening in an integrated, three-tier support system.



important part of a system in which curriculum, instruction, and assessment are aligned, in and of itself will not be enough to change levels of achievement.

As a start, school personnel must agree that the school needs to examine academic and/or behavioral health. NCLB has highlighted the importance of examining academic health. In the area of behavior, school personnel need to agree that they, as a faculty, are committed to studying, understanding, and having an impact on the social-emotional health of the student body (Sugai, Horner, & Gresham, 2002). School personnel (including administration) should agree that if data indicate achievement deficits for large groups of students, then the district administration and school board should be willing to engage in conversations about changing or enhancing core programming and general instructional practices. Additionally, school personnel must recognize that current structures for referring students for additional support, and the current mechanisms for delivering supplemental instruction, may need to change.

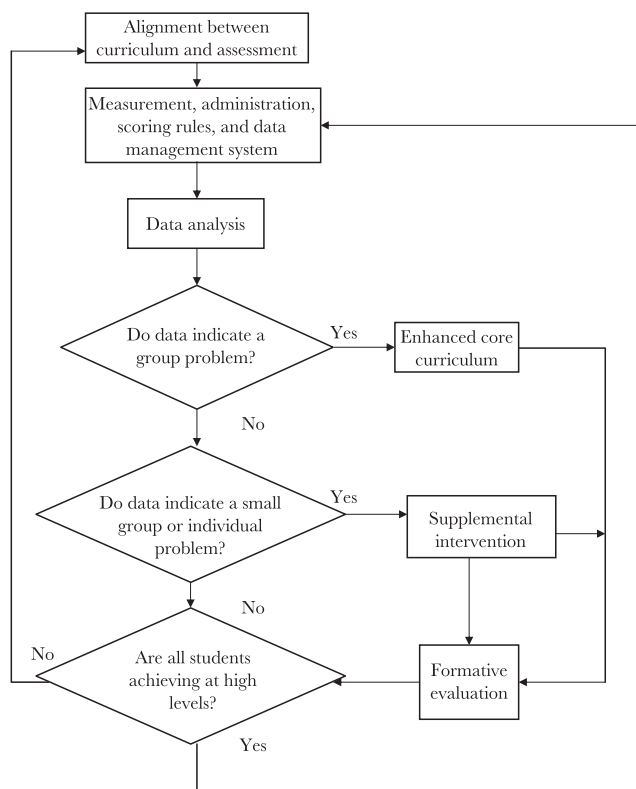
Critical features of an effective universal screening system are depicted in Figure 4. The first step in identifying a universal screening tool is ensuring the alignment between curriculum and assessment. Once aligned, universal screening tests must be evaluated for technical adequacy and efficiency. Means for data analyses must exist, long before ever answering the important questions relevant to universal screening: (a) Do data indicate a large group problem? (b) Do data indicate a small group problem? (c) Do data indicate an individual problem? Importantly, the process is cyclical in nature, with decisions about alignment, measurement, data, and effect continuing even when a system reaches the point of all students achieving at high levels.

Alignment Between Curriculum and Assessment

A three-tiered system rests on the assumption that an effective core curriculum supports about 80% of learners (see Tilly, chapter 2, vol. 1). It is important for school systems to understand the level of academic or social-emotional success achieved through the core curriculum so that decisions about resource alignment can occur. For example, if 80% or more of the learners meet or exceed grade-level expectations, 20% of the students would require additional group or individual support according to their needs. However if data indicate that fewer than 80% of students are meeting expectations,

④

Figure 4. Flowchart of characteristics of universal screening.



then the core instruction and curriculum should be reevaluated and strengthened.

Reasonable linkages to the universal screening measure and what is being taught, how skills are taught, and what is assessed must be present (see Niebling, [5][6] Roach, & Rahn-Blakeslee, chapter 65, vol. 4). For example, an approach could be taken that because curriculum-based measures (CBM; Shinn, 1989) have been validated for use in universal screening, CBMs can be applied in any school district in the United States. However, this approach would be incorrect, because if the construct being assessed by the CBM was not being taught by the district, screening decisions would be expected to be unreliable (Ardoin et al., 2004).

This issue may be magnified in the area of mathematics, where popular instructional frameworks in math, and constructs assessed by CBM digits-correct fluency probes, may not be well aligned (see Kelley, [7] chapter 24, vol. 2). District personnel must discuss standard and benchmark alignment of their district, alignment with assessment measures, and the research base on construct and predictive validity of measures in question. Decisions about screening measures that are most appropriate given district standards can be made

after having thoughtful conversations and studying professional research.

Measurement, Administration, Scoring Rules, and Data Management System

Once the alignment between curriculum, instruction, and assessment has been established, the efficiency and technical adequacy of the screening measure must be examined. Criteria against which universal screening measures are judged are (a) can identify potential problem areas, (b) answers questions about efficacy of the core program, (c) can be disaggregated and used by teachers, (d) can be easily administered to groups of students, (e) are brief in administration time, (f) are repeatable over time, and (g) have published reliabilities of at least .80 (these are summarized in Table 2).

For example, while published, standardized, norm-referenced tests of achievement could be administered and distribution of scores to analyze could be created, such an approach would be problematic because (a) such measures are not easily administered to groups of students; (b) administration, scoring, interpreting, and reporting takes substantial time; and (c) such measures are not easily repeated during the course of a school year. Such measures may be helpful for problem analysis but are not appropriate for initial universal screening.

CBMs, because of their technical characteristics and ease of administration, scoring, and interpretation, are efficient for use as universal screening *if the measures align to district or state defined benchmarks*. CBM Oral Reading Fluency (ORF) measures have an established research base on the robustness of the measure for assessing general reading competence and an emerging research base on the utility of CBM ORF in predicting performance on large-scale assessments. Hence, issues of construct validity and alignment may be smaller for ORF than other CBM measures where the research base is not as established. The bigger issue is that consumers of the assessments must understand the merits and demerits of what questions the assessments can answer and the efficiency with which assessments can be administered and scored.

Universal screening measures are most efficient when measures can be administered multiple times in a school year. Typical large-scale assessments may not have the ease of administration and scoring; however, these assessments can help determine how many and which students are at risk for not achieving adequately. Supplemental measures will be needed to assess growth

Table 2. Judging Sufficiency of Universal Screening Measures

Element/criteria	Meets	Emerging	Does not meet
Purpose	<ul style="list-style-type: none"> Identifies potential academic and/or behavioral concerns in need of additional assessment 	<ul style="list-style-type: none"> Does not accurately and consistently identify potential academic and/or behavioral concerns 	<ul style="list-style-type: none"> No screening exists
Questions answered	<ul style="list-style-type: none"> How is each student responding to core instruction? How many students are at risk for failure? Is the current core instruction effective? Which students are in need of additional assessments? What levels of resource support might be needed to promote criterion-level performance? 	<ul style="list-style-type: none"> Only a few questions are answered 	<ul style="list-style-type: none"> No questions are answered
Data use	<p>All of the following are observed:</p> <ul style="list-style-type: none"> Teachers get the data Teachers use the data to differentiate instruction District has expectation that all teachers use data to align instructional resources. 	<p>Some of the following are observed:</p> <ul style="list-style-type: none"> Teachers get the data Teachers use the data to differentiate instruction District has expectation that all teachers use data to align instructional resources Most students are included in screening 	<p>None of the following is observed:</p> <ul style="list-style-type: none"> Teachers get the data Teachers use the data to differentiate instruction District has expectation that all teachers use data to align instructional resources Most students are excluded from screening
Student participation	<ul style="list-style-type: none"> An entire classroom/grade/ school/district, using either individual or group format 	<ul style="list-style-type: none"> Most students are included in screening 	<ul style="list-style-type: none"> Most students are excluded from screening
Psychometric standards	<ul style="list-style-type: none"> Has reliability of .80>All of the following are observed: The test scores distinguish the proficiency levels of students; the test identifies which and how many students differ from the standard Items are sufficient to accurately reflect changes in student performance Sources of bias have been eliminated People are adequately trained to administer, score, and interpret the test results 	<ul style="list-style-type: none"> Has reliability of .70-80.Most of the following are observed: The test scores distinguish the proficiency levels of students; the test identifies which and how many students differ from the standard Items are sufficient to accurately reflect changes in student performance Sources of bias have been eliminated People are adequately trained to administer, score, and interpret the test results 	<ul style="list-style-type: none"> < .70 or no evidence of reliability None of the following is observed: The test scores distinguish the proficiency levels of students; the test identifies which and how many students differ from the standard Items are sufficient to accurately reflect changes in student performance Sources of bias have been eliminated People are adequately trained to administer, score, and interpret the test results

over time and to provide additional benchmark assessments so that resources can be reallocated if needed.

Methods for Data Analyses

After alignment between curriculum-instruction-assessment is established, and the universal screening measure is evaluated for technical adequacy and efficiency, data can be obtained on all students in the school and on any measure for which alignment and technical sufficiency was established. There is a need, then, for means of analyzing the data to assess the effectiveness of the core programming.

Performance of students in a mock classroom, sorted from low to high, is depicted in Figure 5. Computer software like Microsoft Excel, readily accessible to educators, can be used to generate histograms as depicted in several figures in this manuscript. Each bar represents a student's performance. The solid line crossing the y axis is the criterion line. Students above the line are deemed not at risk and those below the line are deemed at risk.

The criterion line is established using one of several means. First, test publishers often provide benchmarks against which proficient performance is judged. Second, researchers may provide guidelines for cut-points that predict proficiency. In the absence of the first two methods, school personnel will need to seek technical assistance from universities, educational intermediate units, state departments of education, or private consultants on visual and statistical methods for setting cut-points.

In Figure 5, 85% of students are performing at or above the criterion line. The few students below the criterion line are exhibiting performance well below the line. One interpretation of Figure 5 is that core

instruction and curriculum are meeting the needs of most students. There are few students for whom performance was not adequate. The school would want to reassess those students or review other records to determine that test results were accurate and then use a data team (see Kovaleski & Pedersen, chapter 6, vol. 2) to further analyze the data and determine courses of action. Typically, in situations depicted in Figure 5, small group or individual interventions are implemented rather than modifying core programming.

In contrast, Figure 6 depicts the situation in which 65% of students are at or above the criterion line. There are two distinct subgroups of students whose performance is below the goal line. The first subgroup (Juan through Manuel) are almost at criterion. The second subgroup (Jane through Alex) perform well below the criterion line. In this example, because there are fewer than 80% of students achieving at the criterion, the data team should first reevaluate and strengthen the core instruction and curriculum as well as implement small group or individualized interventions for those students who need it.

In examining core programming, alignment would be a reasonable first step. Another tool for examining the instructional practices within schools in reading is the Planning and Evaluation Tool for Effective School-Wide Reading Program—Revised (Kame'enui & Simmons, 2003). This tool can be used with teachers to help understand if elements such as goals/objectives/priorities, assessment, instructional practices, instructional time, differentiated grouping, administration, and professional development are adequate for maximizing achievement.

Figure 5. Data indicating small group problem.

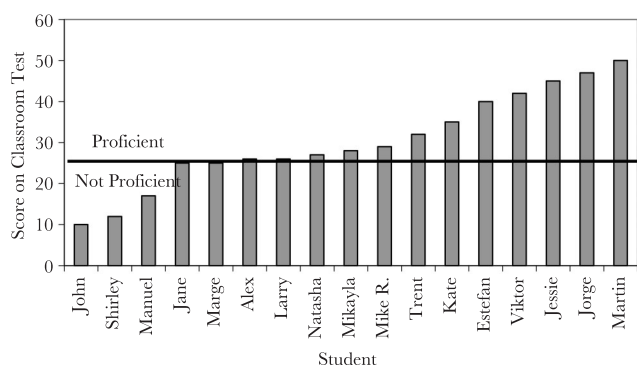
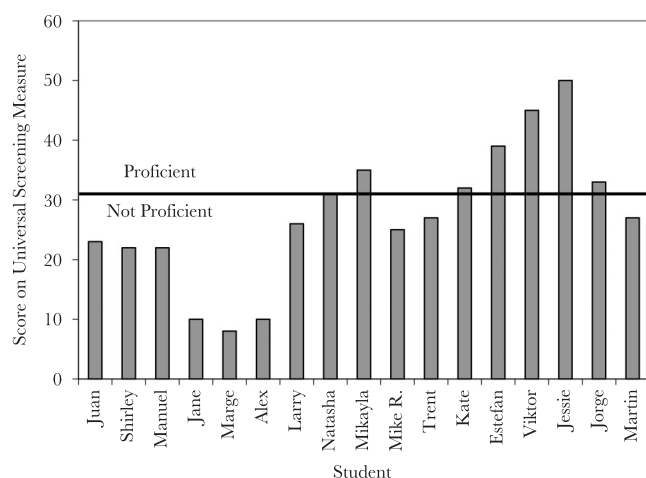


Figure 6. Data indicating class-wide problem.



It can be useful to aggregate screening data by NCLB subgroups because this can provide information about whether Tier 1 is effective for all subgroups. For example, screening may reveal that upward of 80% of students in a school are above a benchmark standard but that only 20% of students from the ethnic subgroup “Hispanic” are meeting the standard. This may lead to an examination of whether the core curriculum is meeting the needs of this subgroup so that enhancements to core curriculum can be made. Such use of data for resource alignment is particularly helpful in predicting student achievement on state assessments as well as avoiding disproportionate representation of minority groups in special education.

Universal screening one or two times per year may be sufficient to determine how best to align resources. However, because it can never be predicted ahead of time what achievement levels will occur year to year and cohort to cohort, the least dangerous assumption is to create a universal screening system to assess performance two to four times per year. Multiple assessment periods are important for several reasons.

First, growth over time can be assessed. Second, students whose status changes over the course of the school year, from at risk to not at risk, or from not at risk to at risk, can be provided the supports they need (or no longer need) throughout a given school year rather than waiting to change supports from one school year to the next school year (see Powell-Smith, chapter XX, vol. X). Hence, school personnel will want to select universal screening tests or tools with several parallel test forms in order to avoid practice effects.

9

In addition, for individual students in need of additional support, teachers need to create formative evaluation measures so that instructional effect can be assessed over shorter time periods (weeks or months rather than quarters or semesters). When formative measures are the same as universal screening measures, continuity of performance over time is assessed. This continuity makes interpretation of growth over time easier to assess (see Fuchs & Fuchs, chapter 136, vol. 6). Frequency of formative evaluation depends on the students’ level of need. For students with more intensive needs, teachers monitor progress more frequently. For students achieving in the core curriculum, assessing performance three times per year is sufficient.

10

Questions

Identifying how students are responding to benchmark-aligned core instruction is the ultimate goal of academic

universal screens. A data analysis team should be able to look at the data and answer the following questions:

- How is each student responding to instruction?
- How many students are at risk for failure?
- Is the current core instruction effective?
- Which students may need additional assessments?
- What levels of resource support might be needed to promote criterion-level performance?

An effective universal screening measure addresses all questions. A less effective universal screening measure will address only certain questions. At the very least, universal screening measures need to address how many students are achieving at criterion levels, how many students are below criterion, and how far below criterion are students. Importantly, universal screening tools should be able to discern if the students whose performance is significantly *above* the criterion line are also receiving appropriately aligned supplemental or intensive support (e.g. gifted students).

Assumptions in Universal Screening

One of the most fundamental assumptions with universal screening is that student achievement (or lack of achievement) lies in the alignment between curriculum (what is taught), instruction (how it is taught), and assessment (how we know it was learned). Other important assumptions for successful implementation of universal screening are (a) problems are contextually defined, (b) culture determines what is imperative to screen, (c) direct measures of student performance are more reliable predictors of student need than are inferential measures of student performance, and (d) school psychologists need to understand that screening will help shape their work.

Contextually Defined Problems

Deno (1989) described special education as a system designed to address the problems of general education. Students who are low performers display a mismatch between what behaviors they exhibit and what the system demands. Hence, special education becomes a resource available to general education to solve problems of students not succeeding in general education.

An important assumption in a three-tiered system and universal screening is that most students displaying low achievement are displaying low achievement in relation

to that specific school district. A disability becomes defined in the context of a school system rather than in the context of some inferred trait being displayed by the student. Hence, it is likely that, in a contextually defined system, some students can be “cured” of a disability merely by changing schools. Conversely, some students may “become disabled” merely by moving from a moderately achieving school to a high achieving school.

Cultural Imperatives Versus Cultural Electives

Universal screening is designed to have an impact on skills that allow access to life. These access skills are locally defined. Most people would agree that being emotionally healthy and proficient in reading, math, and science provides equal opportunity for students to pursue life, liberty, and happiness. Reading, math, and science proficiency are *cultural imperatives*. Music, athletics, and art are also important, but in most communities would be considered *cultural electives*.

Schools are not expected to screen all students for music, athletic, or artistic performance because proficiency in these areas is not needed to have access to living, learning, and working. Value judgments need to be made when deciding what skills to universally assess. An extreme example is that some communities may decide to conduct universal screening on body mass index. The question community leaders must address becomes: Is school the appropriate vehicle in which to conduct such screening?

Direct Measures of Performance

Another assumption inherent with universal screening is that measures of student performance are more reliable predictors of student need than are inferential measures of student performance, such as IQ tests. Universal screening measures are directly linked to important achievement outcomes and provide objective data to school personnel about skills that need instruction or additional support. Rather than focusing on assessing constructs and inferring a disabling condition, universal screening focuses on academic and behavioral indicators for measurement because it assumes a contextual basis and because it is linked to curriculum and instruction. The emphasis on direct assessment is important, because empirical research has found that often-assumed measurement issues such as performance scatter and aptitude limiting achievement levels are untrue (Speece & Case, 2001).

Impact on Work of School Psychologists

Universal screening is important for school psychologists because results of universal screening shape the work of the school psychologist in that building for that school year. In some cases the school will modify the core curriculum and the school psychologist will support implementation of the changes and help teachers assess effect. In other cases, the school will implement small group interventions with demonstrated effect and the school psychologist will assist with implementation and evaluation. Whatever the situation, the school psychologist will need to be prepared to assist in selecting, administering, and interpreting instructionally relevant assessment data.

To contribute to a school’s effort to improve achievement for all students, through a study of the alignment between curriculum, instruction, and assessment, a school psychologist must have basic understanding in several areas. School psychologists must have experience with direct measures of performance (such as CBM). School psychologists must have experience with interventions, both for individual students and for groups of students. School psychologists should have an understanding of effective instructional principles, such as errorless learning, modeling, prompting, and reinforcing academic and social behaviors. In addition, school psychologists must stay current with professional literature, including the research base for interventions in both academic and behavioral domains. Psychologists must have the knowledge to interpret these results, including an understanding of effect size.

BEST PRACTICES

Universal screening has an emerging research base. In the literature to date, CBMs are most commonly used as the universal screening tool. Results from several studies will be summarized demonstrating the efficacy of using an aligned curricular–instructional–assessment framework to improve learning. The studies also reflect real-life applications of universal screening, in some cases decision making reflected at the building level.

Witt and colleagues have engaged in an extensive research agenda around universal screening and have studied a process called System to Enhance Educational Performance (STEEP; VanDerHeyden & Witt, 2005; VanDerHeyden, Witt, & Gilbertson, 2007; VanDerHeyden, Witt, & Naquin, 2003). Schools engaged in STEEP use CBM in reading, math, and writing, to identify the extent to which the core

curriculum is effective for sufficient numbers of students and to identify the students for whom additional resources are provided. VanDerHeyden et al. (2007) reported that school psychologists conducted fewer psychoeducational evaluations when STEEP was implemented school-wide compared to evaluations conducted prior to STEEP, at a substantial cost savings. VanDerHeyden et al. also reported that special education placements at STEEP sites rose initially and then reduced over time.

Research on universal screening in reading was thoroughly reviewed by Jenkins (2003). While the data supported many measures as candidates for use in universal screening, to date no study or bodies of studies have validated any measure or set of measures as proven for use in universal screening. Jenkins concluded that fluency measures like CBM were good candidates for use in universal screening through the second grade, but added that the research base on screening measures is not well developed beyond grade 2. DIBELS are fluency-based indicators of skills such as initial sound fluency, letter naming fluency, phoneme segmentation fluency, nonsense word fluency, and oral reading fluency (Good et al., 2002). DIBELS are frequently cited as useful for universal screening of early literacy development. (For research on technical adequacy of DIBELS, see Web Resources at the end of this chapter for the DIBELS website.)

In the area of behavior, positive behavior supports (PBS; Sugai, Horner, & Gresham, 2002) is viewed by some as an exemplar of a universal screening system. PBS does not use universal screening data per se in decision making. Office referrals are used as a proxy for estimating whether or not a school has an issue with behavior that is more like a core behavioral programming problem or if the issue is with groups of students.

In all cases, the intervention practices are the same in PBS sites: (a) develop and implement core behavioral expectations applicable to all students; (b) teach, model, and reward appropriate behaviors on a frequent and ongoing basis; (c) develop interventions that emphasize rewarding appropriate behavior over punishing inappropriate behavior, for students needing additional support; and (d) use RTI data to determine students for whom more extensive analysis of behavior is warranted (Sugai, Horner, & Gresham, 2002).

PBS has an emerging research base that supports implementation of PBS on more distal measures like office referrals. PBS does not utilize a universal screening of social-emotional behavior in decision making. Because PBS is implemented on a school-by-school

basis, results in the literature tend to be of a case study nature. As such, an aggregation of successful implementation and effect creates the foundation for building a compelling case of effect.

Other work in universal screening in behavior has been summarized by Severson, Walker, Hope-Doolittle, Kratochwill, and Gresham (2007) and by Elliott, Huai, and Roach (2007). Severson et al. described several published tools validated for use in universal screening in the area of behavior. Characteristics of effective universal screening in behavior involved repeated measuring of subsets of students identified through initial screening as at risk, teacher evaluation of all students using the same measure, teacher nomination of students with behavioral difficulty, and additional teacher ratings of students' behaviors. The gated procedure for identifying students helps reduce false positives. The process described by Severson et al. is consistent with the three-tiered, preventative approach described throughout this edition of *Best Practices*.

Elliott et al. (2007) describe direct and indirect measurement of key academic and social behaviors for use in universal screening. Use of curriculum-based academic measures and teacher rating scales of behavioral competence were proposed as a comprehensive model for screening at-risk behavior in preschoolers. The research line appears promising, but conclusions about effective practice cannot yet be drawn from the data to date.

SUMMARY

Universal screening involves systematic assessment of all children within a system on important indicators like academic or behavioral competence. Universal screening tools must be aligned with schools' curriculum and instruction and must allow school personnel to understand the general academic and behavioral health of students in that system. Many CBMs, including DIBELS and behavior-rating scales, in general *may be* appropriate for use in universal screening if the construct assessed is relevant for student learning. Other measures may have utility depending upon the degree of alignment of those measures to district standards, as well as the time it takes to administer and score the measures. Measures that can be administered more than once over the course of the school year are preferable to measures that can be administered only one time.

Importantly, universal screening is useful only to the extent that the data are used in decision making. Schools

need to use data teams to analyze data and ask important questions such as: What is the impact of our core instruction on overall student achievement?

School personnel may be better off changing or enhancing core curricula prior to deciding whether or not additional resources (e.g. Title I, special education) should be explored for students not achieving.

Universal screening is but one component of an integrated academic and behavioral system. It provides data for school personnel to use in judging the overall effectiveness of their current system. In addition to data, schools need instructional strategies to meet the diverse needs of students. Research on the effectiveness of universal screening practices on achievement is beginning to appear in the professional literature. While current data are promising, more research on applications of universal screening in school systems is needed.

REFERENCES

- Ardoin, S. P., Witt, J. C., Suldo, S. M., Connell, J. E., Koenig, J. L., Resetar, J. L., et al. (2004). Examining the incremental benefits of administering a maze and three versus one curriculum-based measurement reading probes when conducting universal screening. *School Psychology Review, 33*, 218–233.
- 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. 1–18). New York: Guilford Press.
- Deno, S. L. (2003). Developments in curriculum-based measurement. *Journal of Special Education, 37*, 184–192.
- Elliott, S. N., Huai, N., & Roach, A. T. (2007). Universal and early screening for educational difficulties: Current and future approaches. *Journal of School Psychology, 45*, 137–161.
- Good, R. H., III, Gruba, J., & Kaminiski, R. (2002). Best practices in using Dynamic Indicators of Basic Early Literacy Skills (DIBELS). In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 699–720). Bethesda, MD: National Association of School Psychologists.
- Iowa Department of Education. (2006). *Instructional decision making*. Des Moines, IA: Author.
- Jenkins, J. R. (2003, December). *Candidate measures for screening at-risk students*. Paper presented at the National Research Center on Learning Disabilities Responsiveness-to-Intervention Symposium, Kansas City, MO.
- Kame'enui, E. J., & Simmons, D. C. (2003). *Planning and Evaluation Tool for Effective School-Wide Reading Programs—Revised (PET-R)*. Eugene, OR: Institute for the Development of Educational Achievement, University of Oregon. Retrieved January 10, 2007, from <http://oregonreadingfirst.uoregon.edu/downloads/PET-R.pdf>
- Severson, H. H., Walker, H. M., Hope-Doolittle, J., Kratochwill, T. R., & Gresham, F. M. (2007). Proactive, early screening to detect behaviorally at-risk students: Issues, approaches, emerging innovations, and professional practices. *Journal of School Psychology, 45*, 193–223.
- Shinn, M. R. (Ed.). (1989). *Curriculum-based measurement: Assessing special children*. New York: Guilford Press.
- Speece, D. L., & Case, L. P. (2001). Classification in context: An alternative approach to identifying early reading disability. *Journal of Educational Psychology, 93*, 735–749.
- Sugai, G., Horner, R., & Gresham, F. (2002). Behaviorally effective school environments. In M. Shinn, H. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approach*. Bethesda, MD: National Association of School Psychologists.
- VanDerHeyden, A. M., & Witt, J. C. (2005). Quantifying context in assessment: Capturing the effect of base rates on teacher referral and a problem-solving model of identification. *School Psychology Review, 34*, 161–183.
- VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multiyear evaluation of the effects on an RTI model on identification of children for Special Education. *Journal of School Psychology, 45*, 225–256.
- VanDerHeyden, A. M., Witt, J. C., & Naquin, G. (2003). Development and validation of a process for screening referrals to special education. *School Psychology Review, 32*, 204–227.
- Ysseldyke, J., Burns, M., Dawson, P., Kelley, B., Morrison, D., Ortiz, S., et al. (2006). *School psychology: A blueprint for training and practice III*. Bethesda, MD: National Association of School Psychologists.

ANNOTATED BIBLIOGRAPHY

Donovan, M. S., & Cross, C. T. (2002). *Minority students in special education and gifted education*. Washington, DC: National Academies Press.

An important book for understanding the need and rationale for RTI in general and screening in particular. It is a product that came out of an extensive study by a distinguished National Academy of Sciences panel on minority over/under representation in special and gifted programs. Although the focus is on minorities, the book is relevant to all students. The great strength of this work is the depth and breadth of the research presented, which details the need for screening and early detection of students who are at risk for learning problems.

Shinn, M. R. (1989). *Curriculum-based measurement: Assessing special children*. New York: Guilford Press.

Provides fundamentals of CBM. Of particular relevance to universal screening is chapter 4, Identifying and Defining Academic Problems: CBM Screening and Eligibility Procedures.

The technical properties of CBM are summarized, and the appendix contains instructions for administration of CBM probes in typical academic content areas.

VanDerHeyden, A. M., Witt, J. C., & Naquin, G. (2003). Development and validation of a process for screening referrals to special education. *School Psychology Review*, 32, 204–227.

Part of a research line that describes use of CBM in universal screening and alignment of instructional supports. Impact of using data to determine how best to support students not meeting criteria is summarized. This study is important in establishing an emerging body of evidence of the validity of CBM in universal screening as well as the rationale behind universal screening.

WEB RESOURCES

AIMSweb: www.aimsweb.com

Contains comprehensive measurement tools and reports as well as frameworks to guide decision making. For a fee, schools obtain access to measures, reporting tools, and other utilities to help implement school-wide models for promoting academic achievement.

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

Includes description of measures, video clips of measures being administered, technical reports, sample reports for disaggregating information, and measures available for download.

Intervention Central: www.interventioncentral.org

Contains intervention ideas, tools for educators, curriculum-based measurement probes, topical downloads, and other resources for using data to support learning.

National Center on Student Progress Monitoring: www.studentprogress.org/default.asp Information on progress monitoring, including reviews of tools, research, training opportunities, and an online forum.

OSEP Technical Assistance Center on Positive Behavioral Interventions and Supports: www.pbis.org/main.htm

Information on school-wide PBS. Presentations, research, and tools for successful implementation of PBS are available for download.

Oregon Reading First: http://oregonreadingfirst.uoregon.edu/downloads_gen.php

Contains information on comprehensive reading programs, supplemental and intervention reading programs, forms for evaluation of current implementation of efforts in reading, and implementation checklists. The assessment section contains information on screening and diagnostic measures.

Research Institute on Progress Monitoring: www.progressmonitoring.net

Contains information on tools used for progress monitoring, literature reviews, probe development, and background research.

Special Connections: www.specialconnections.ku.edu/cgi-bin/cgiwrap/speconn/main.php?cat=assessment§ion=teachertools

Information from the University of Kansas on data-based decision making. Contains information about instruction, behavior plans, collaboration, assessment, and standards. Each area contains teacher tools, research, case studies, and an online forum. In the area of assessment, teacher tools of CBM, direct observation methods, test construction, accommodations, and grading are summarized.

STEEP: www.isteep.com

Provides information on STEEP, which is an RTI model that uses a standard protocol to integrate screening, intervention, and progress monitoring.

Authors Queries

Journal: **Best Practices in School Psychology V**

Paper:

1 Title: **Best Practices in Universal Screening**

Dear Author

During the preparation of your manuscript for publication, the questions listed below have arisen. Please attend to these matters and return this form with your proof. Many thanks for your assistance

Query Reference	Query	Remarks
1	This has references to other chapters	
2	Chapter number	
3	Chapter number	
4	Chapter number	
5	Chapter number	
6	Chapter number	
7	Chapter number	
8	Chapter number	
9	Chapter number	
10	Chapter number	