Universal Screening as a Component of Response to Intervention

Evelyn Johnson, Boise State University
Daryl F. Mellard, University of Kansas

A supplemental resource to:
Responsive Instruction: Refining Our Work of Teaching All Children
Virginia's "Response to Intervention" Initiative

January 2009

Virginia Department of Education
Office of Special Education and Student Services

If you have questions about this document, please contact the Virginia Department of Education, Office of Student Services at 804-786-0720. This document complements and extends information disseminated by the Virginia Department of Education in an earlier document entitled, **Responsive Instruction: Refining Our Work of Teaching All Children, Virginia's "Response to Intervention" Initiative.** The earlier document can be accessed the following Web site:
http://www.doe.virginia.gov/VDOE/studentsVCS/RTI/guidance_document.pdf.
The Virginia Department of Education does not mandate or prescribe a particular curriculum model or lesson plans. The information contained herein is provided only as a resource that educators may find helpful and use at their option.

Superintendent of Public Instruction
Dr. Patricia I. Wright

Assistant Superintendent for Special Education and Student Services
H. Douglas Cox

Office of Student Services
Dr. Cynthia A. Cave, Director

As school staffs work to predict which students might develop later academic or behavioral difficulties, they will rely on indicators of those potential difficulties. The logic is that the sooner the staff can know who might be having difficulty or which students could be predicted as having difficulty, then the sooner steps could be taken to help the student and ameliorate the situation. This prediction question is one that screening assessments can address (Mellard & Johnson, 2008). The assessments and concepts of screening have improved extensively over time, especially in the area of predicting reading problems, behavioral difficulties, and school dropout. In this paper we review concepts that are important to understanding the value and limitations of screening assessments. While some of the concepts involved in evaluating screening instruments can become quite complicated, we chose to provide a more basic overview, to help school staff become better consumers of screening and the results. We trust that readers interested in more details will readily access more detailed information relevant to their questions.

What is Screening?

An important first step in any prevention approach is schoolwide screening of students to accurately identify those who are at risk for learning or behavioral difficulties. Screening is a type of assessment characterized by providing quick, low-cost, repeatable testing of age-appropriate critical skills (e.g., identifying letters of the alphabet) or behaviors (e.g., tardiness).

The basic question for a screening measure is whether the student should be judged as "at risk" of encountering difficulties in the target behavior. For a primary school-age student, the target behavior might be predicting reading acquisition. For a secondary school-age student, the target behavior might be predicting school drop out. Students identified as at risk for reading problems, for example, are then referred for a more in-depth assessment of their reading ability.

For a screening measure to be useful, the measure must achieve an appropriate balance of accuracy and efficiency (Jenkins, 2003). Each of these features is described in more detail below.

Figure 3.1. The ideal screen.
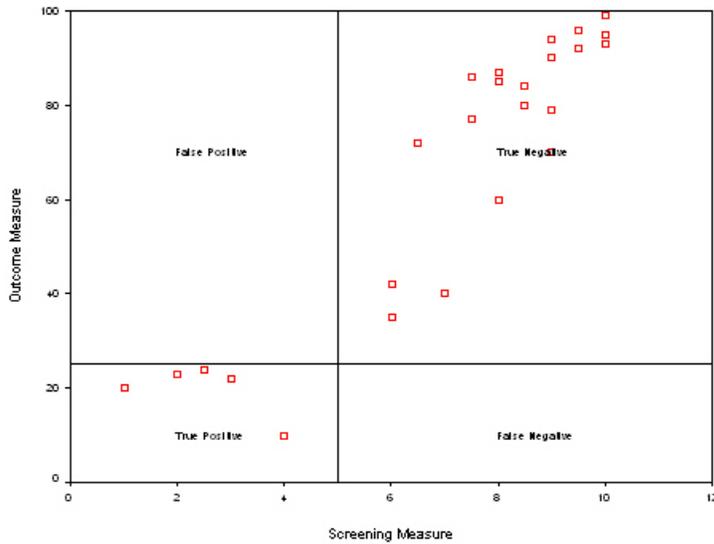
Accuracy. A critical feature of a screening tool is its ability to accurately classify students as at risk or not at risk. Figure 3.1 depicts results of an ideal screen that correctly identified 100 percent of students who are not at risk for reading failure and do not later develop reading problems and 100 percent of students who are at risk and later develop problems. As represented by the vertical line in Figure 3.1, the screening measure had a cutoff score of 5. That is, students who scored below 5, were judged as at risk of difficulties. Students who scored above 5 were judged as not at risk. Those students scoring below 5 are of concern and warrant careful attention. We can assign this at-risk status because we have their scores on the outcome measure (e.g., reading achievement). The outcome measure for adequate performance is indicated by the horizontal line. The school had determined that students scoring below the 25th percentile were not making adequate progress.

Unfortunately, achieving such perfect results (as presented in Figure 3.1) with a screen is highly unlikely. Therefore, schools must consider accuracy in relation to the sensitivity and specificity of the measures. *Sensitivity* is a screen's ability to identify "true positives"— those students who perform poorly on the screen and do have reading problems, and, therefore, will require more intense levels of instruction and intervention. *Specificity* refers to the screen's ability to identify "true negatives"—those students who do not perform poorly on the screen and do not have problems.

Because screening does not directly result in a diagnosis, some believe that a screening instrument's cut score is better to err on the side of false positives (identify more students as at risk)—casting a wider net to capture students potentially at risk. However, because identifying more students as at risk requires resources for further assessment and possibly intervention, schools need to maintain data on how well their screen identifies students as at risk. Tracking screen scores over time (e.g., fall, winter, spring scores) as well as alternative measures (e.g., district reading assessment, grade point average, state reading assessment) for each student may aid schools in assessing their screen's accuracy. Several statistics such as correlational analyses of these data may also give an estimate of the screen's predictive validity and help school staff choose among alternative screening assessments.

Factors that can affect a screen's sensitivity and specificity include whether the measure is criterion or norm-referenced and what cut scores distinguish levels of performance. Screens can use either a criterion-referenced or a norm-referenced performance standard. A *criterion-referenced* measure compares a student's performance to a predetermined level or goal. A *norm-referenced* measure compares the screening results to an appropriate target group (e.g., other students in first grade); students scoring below predetermined percentile are considered "at risk".

The "cut scores" used to distinguish students as "at risk" or "not at risk" also affect a screen's accuracy. A cut score is the point that represents the dividing line between students who are not at risk and those who are potentially at risk. Adjusting cut scores changes the screening tool's sensitivity and specificity. If student level data has been collected as described above, a school can plot performance of the screen along with subsequent performance on the targeted skill (e.g., reading achievement). Figure 3.2 shows an example of the distribution of scores on a screen and performance on a state assessment. In this example, the outcome measure has a "proficiency" standard of 400. Students whose scores fall below that standard are considered as not proficient. Figure 3.2 presents only one possible example of a cut score to identify students as at risk for not meeting standard on the outcome measure. Figure 3.3 shows the same distribution of scores; however, in this graph, the cut score has been altered, leading to changes in the sensitivity and specificity of the screening measure. The number of true positives and true negatives changed with the change in the cut score. True positives increased but the number of true negatives decreased. Also, the false positives increased and would be judged as at risk. The number of misses or false negatives decreased. School staff must weigh the consequences of such changes when they are making a decision about cut scores. Although many educators would agree that identifying more students as at risk is better, a negative consequence would be the strain on resources to provide intervention.

Efficiency. A second critical feature of a screening procedure is that it must be brief and easy to implement reliably (Jenkins, 2003). Although increasing the breadth and depth of a screening procedure can help improve its accuracy in correctly classifying students, schools must consider the costs and benefits of such changes.

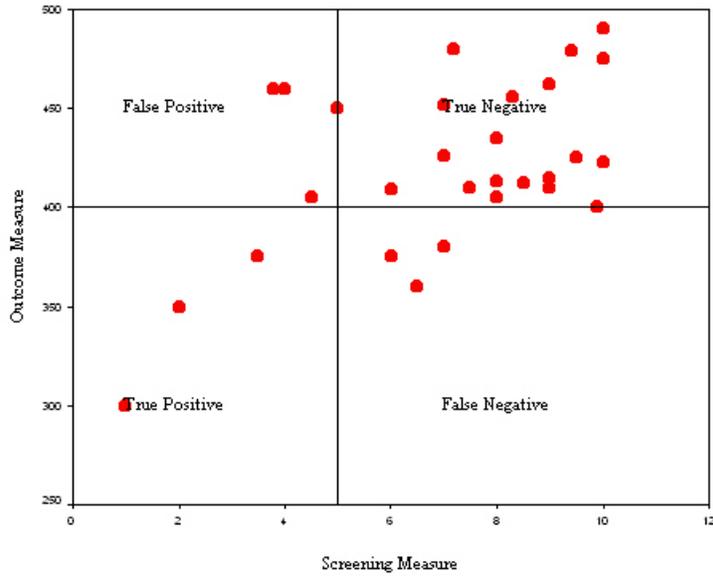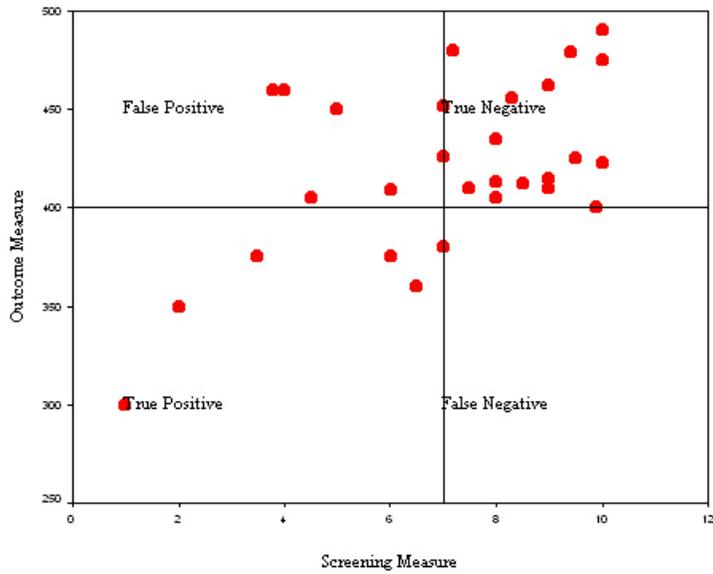*Figure 3.2.* Screening and outcome measures with cut scores.

Figure 3.3. Changing the cut score changes who is judged as "at risk."

Implementing academic and behavioral screening poses several challenges including

administrative issues like scheduling and record keeping. The greater challenges, however, are

associated with ensuring that the staff has the knowledge to use the screening results in curricular

decisions regarding their class and individual students. Screening measures can help inform instruction, but the measures themselves should not drive instruction.

What Is the Role of Screening within a RTI model?

In RTI, proactive screening assessments are best employed at least three times per academic year (beginning, middle, end) and are used as general screening procedures for all students. Screening results can be used to target students who may be at risk by comparing their performance relative to a criterion or normative index of performance.

Screening is important as it represents the first gate or point of entry into subsequent tiers of RTI instruction (e.g., Tier Two, secondary interventions, and Tier Three, tertiary intervention). Screening is not a one-time event but an iterative process taking place during the school year and across grade levels. During the course of general instruction (Tier One), the school uses schoolwide screening in essential academic areas to identify each student's level of proficiency (usually three times per year). The screening data are organized to allow for comparison of both group (e.g., class) and individual performance. Comparisons of group performance can provide feedback about class performance to school leadership to identify when a teacher may require additional support, for example. Individual performance helps identify students who are potentially at risk for not acquiring the academic skill or experiencing behavior difficulties.

As mentioned, ideally, screening should be conducted at least three times per year. One time screenings at the beginning of the school year yield more false positive errors than is generally acceptable (Fuchs & Fuchs, 2006). Research examining standards for screening suggests that one way to help make the screening process more efficient is to combine screening

with five to six weeks of supplemental progress monitoring for students identified as at risk on the initial screen (Compton, Fuchs, Fuchs & Bryant, 2006).

A second point to consider is the extent to which the screening is effectively "schoolwide." The omission of students for any reason distorts the picture of students' skills and abilities from the school profile. Thus, while the results may not have a major consequence for any one student, the school's profile is distorted and that distortion could have unintended consequences regarding decisions about the effectiveness of curricular decisions, the adequacy of instructional practices, and the magnitude of behavioral difficulties.

Standards for Judging High-Quality Screening

Screening tools must be aligned with the requirements of the school district, school site, and the curriculum. The process outlined in this chapter can help a school develop screening measures that reach the optimal balance between accuracy and efficiency to correctly identify those students whose performance warrants intervention. Table 3.1 presents standards for judging high-quality screening that are based on the research in this area and that were used as part of a national effort to identify model RTI sites (Mellard, Byrd, Johnson, Tollefson, & Boesche, 2004). The checklist is formatted so that you can indicate current and planned implementation. If the practice has been implemented, indicate that with a checkmark. If the practice is being developed, rank its priority of focus: 1 = highest priority, 3 = lowest priority.

Table 3.2 *Standards for Judging High-Quality Screening*

Standard
Screening is schoolwide, meets accepted psychometric standards[1], and has evidence of documented reliability[2] and concurrent[3] and predictive validity[4] within the particular school setting.

Standard
Individuals involved in administration, scoring, and interpretation of screening assessments are appropriately trained.
The site obtains screening data following a designated, fixed schedule.
Students' screening results are documented and analyzed to refine the process.
An established data-management system allows ready access to students' screening data.
Cut points are reviewed frequently and adjusted as necessary.
Explicit consideration is given to the costs and benefits of classification errors in screening (e.g., false positive versus false negative errors) when setting cut scores.
A rationale is provided for the cut points and decision rules (e.g., normative or specific criteria reference).

[1] Psychometric standards are the theoretical approaches and procedures used to measure the difference between individuals' knowledge, attitudes, abilities, and personality traits.

[2] Documented reliability is the extent to which a measurement yields consistent results over repeated testing of the same measure under identical conditions.

[3] Concurrent validity occurs when a new measurement or test correlates well with a previously validated measure.

[4] Predictive validity is the extent to which performance on one measure predicts performance on a later, related measure.

Changing Structures and Rules

As with most elements in RTI, implementation of schoolwide screening procedures necessitates a closer collaboration among general education and specialist staff. Thus, when planning for the implementation of schoolwide screening, school leaders must include both the

acquisition of resources and the time needed to administer screening. Schools must develop a standard procedure for identifying students as at risk. Additionally, the procedure will need to be adjusted based on existing data, so initial implementation also requires the development of a database that can accurately record screening, progress monitoring and outcome data for students so that cut scores and criteria can be adjusted as necessary. This is an iterative, continual process.

Challenges to Implementation

Universal screening in academic skills and behavior provides the information that determines which students need more intense assistance. That assistance might occur in the general education classroom as part of modifications, accommodations, or extra assistance. The low screening score might also be the signal that the student should enter Tier Two in the RTI process and receive interventions (Fuchs & Fuchs, 2006). Therefore, accuracy of the screening measure is paramount. The screening scores are intended to indicate who needs the added assistance. Additionally, because this assessment is conducted schoolwide, screening needs to be efficient. As RTI moves to curricular areas beyond early reading, screening measures that have appropriate discriminant and predictive validity are required for areas such as mathematics, writing, and later reading. Discriminant validity refers to the accuracy with which scores represent different knowledge, skills, and ability.

A further point to consider as a wise consumer involves the time interval between administering the screening assessment and judging whether the expected criterion was achieved. For example, some reading screening measures administered in kindergarten might predict reading achievement at the end of first grade. Other instruments might predict reading achievement at the end of second grade. As you can imagine, predicting an event two years later is likely to have more predictive errors than just one year later. Also, as a rule, predicting

kindergartners' reading achievement is more difficult than predicting first graders' reading achievement. Catts (2006) provides examples of these difficulties as he reviews alternative screening measures.

Summary

When RTI is implemented with fidelity and rigor, all students should benefit. An initial step in the RTI process is ensuring that students who are at risk for academic or behavioral difficulties are identified as early as possible. Early identification avoids the added complications students encounter through repeated failure including negative changes in self-concept and efficacy. Schoolwide screening provides the initial closer examination at students' learning and performance and those screening results can be used for indicating those students needing closer monitoring and more intense interventions and supports than available in the Tier One of general education.

References

Catts, H. W. (2006, April). *School-wide screening*. Presentation at the National SEA Conference on Responsiveness to Intervention: Integrating RTI within the SLD Determination Process, Kansas City, MO. Retrieved July 12, 2006, from http://www.nrcld.org/sea/presentations_worksheets/screening/Catts_screening.pdf

Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology*, *98*, 394-409.

Fuchs, L. S., & Fuchs, D. (2006). Implementing responsiveness-to-intervention to identify learning disabilities. *Perspectives on Dyslexia*, *32*(1), 39-43.

Jenkins, J. R. (2003, December). *Candidate measures for screening at-risk students*. Paper

    presented at the NRCLD Responsiveness-to-Intervention Symposium, Kansas City, MO.

    Retrieved April 3, 2006, from http://www.nrcld.org/symposium2003/jenkins/index.html.

Mellard, D.F., & Johnson, E.S. (2008). *RTI: A practitioner's guide to implementing response to*

    intervention. Thousand Oaks, CA: Corwin Press.

Mellard, D., Byrd, S., Johnson, E., Tollefson, J., & Boesche, J. (2004). Foundations and research

    on identifying model responsiveness to intervention sites. *Learning Disability Quarterly,*

    27, 1-14.