Screening Briefs Series

Brief #1: Classification Accuracy

This brief discusses screening assessments and describes classification accuracy. Understanding classification accuracy will help practitioners become more discerning consumers of screening data.

Schools that are implementing an RTI framework have many decisions to make, including decisions about the screening tools that will best fit their school. Schools that are just beginning to implement an RTI framework or schools well into the practice of RTI might benefit from carefully selecting or reviewing their screening assessments for efficiency and accuracy. NCRTI has developed a screening tools chart, which can be found at www.rti4success.org/screeningTools. Although the chart does not recommend specific products, it can help schools become informed about the screening tools that are available.

Schools can also learn more about the role of screening in NCRTI's *Essential Components* document (National Center on Response to Intervention, 2010), http://www.rti4success.org/pdf/rtiessentialcomponents_042710.pdf, and from Module One (Screening) of the NCRTI Implementer Series http://www.rti4success.org/resourcetype/rti-implementer-series-modules.

All of the Screening Briefs in this series are available for download from NCRTI's website, www.rti4success.org.

Classification Accuracy

A valid screening system requires a screener that accurately predicts whether or not students are at risk given their current performance. To understand a screening measure's accuracy, predicted student performance data (i.e., screening data) must be compared with actual performance data. When these sets of data are compared in a setting where interventions are in place, measures should be taken to ensure that the effects of interventions are controlled in the student data. Controlling student data to account for the effects of interventions (or other variables) requires techniques that are beyond the scope of this brief, but a general treatment of the topic can be found in Thorndike and Thorndike-Christ's 2008 book *Measurement and Evaluation in Psychology and Education.*

Four words can be combined to describe a screening measure's classification accuracy: *true, false, positive,* and *negative.* As an example, scores on a screener that is used to predict outcomes on a future assessment (e.g., an end-of-year test or state assessment) can be categorized as *true positive* (correct prediction of being at risk for academic failure) or *true negative* (correct prediction of not being at risk for academic failure). Actual performance outcome scores would then be categorized as *fail* (student failed the outcome assessment) or *pass* (student passed the outcome assessment). In this example, the combination of measurement prediction with actual outcome performance creates the four diagnostic outcomes presented in Figure 1.

Figure 1. Diagnostic Outcomes

		Actual performance (based on outcome assessment such as a state or district assessment)	
		Fail	**Pass**
Predicted performance (based on screening assessment)	At risk	True positive	False positive
	Not at risk	False negative	True negative

Notice that the matrix has two cells of correct, or true, classifications and two cells of false classifications (classification errors).

Examine first the "at risk" row. This row represents students who scored below the cut point on the screening assessment and thus were predicted to be at risk for failing the outcome measure. Students who were predicted to be at risk and who failed the outcome measure are the *true positives*—students who are truly at risk. Students who were predicted to be at risk but who passed the outcome measure are the *false positives*—students who were incorrectly identified as at risk by the screener.

Now consider the "not at risk" row. This row represents students who were considered not at risk because they scored above the cut point on the screening measure. Unfortunately, some of these students failed the outcome measure even though they were predicted to pass. These students are referred to as *false negatives*. Students who were predicted to pass the outcome measure and did pass it are referred to as *true negatives.*

When examining assessments, practitioners should favor those with a high percentage of correct classifications—that is, a high number of true positives and true negatives. False positives and false negatives represent classification errors and should be minimized in screening measures. False positive rates represent errors that result in students being unnecessarily identified as requiring intervention, and false negative rates represent at-risk students who were incorrectly identified as making satisfactory progress. When reviewing assessments, all of the classification rates should be considered and examined in the context of educational need.

Assessing the Classification Accuracy of Screening Assessments

Researchers commonly assess the quality and accuracy of a screening assessment using the terms sensitivity and specificity. *Sensitivity* is the extent to which a screening measure accurately identifies students at risk for the outcome of interest or detects true positives. It is expressed as the percentage of students taking the assessment that the screening assessment accurately identifies as being at risk.

Specificity is the extent to which a screening measure accurately identifies students not at risk for the outcome of interest. It is expressed as the percentage of the not-at-risk students taking the assessment that the assessment accurately identifies as not at risk.

A perfect assessment (and cut point) would have a sensitivity of 100, meaning that the assessment would accurately identify all (or 100 percent) of the students taking the assessment who are truly at risk—the true positives—and would have a specificity of 100, meaning that it would accurately identify all (or 100 percent) of the students who are not at risk—the true negatives.

Much of the work of Compton, Fuchs, Fuchs, and Bryant (2006) and Compton et al. (2010) has focused on determining effective screening assessments and practices that accurately identify students at risk for reading disabilities. They suggest that an effective reading screening assessment would have

- A sensitivity, or true positive, rate greater than 90 percent (i.e., 90 percent of the truly at-risk students would be identified as such)

- A specificity, or true negative, rate greater than 80 percent (i.e., at least 80 percent of the students who are not at risk would be identified as such)

Recommendations for Practitioners
- Aim for screening assessments and cut scores that result in a balance between ideal sensitivity (greater than or equal to 90 percent) and specificity (greater than or equal to 80 percent). Note that these percentages are very rare and thus are very optimistic goals.
- Base the selection or the evaluation of predictor assessments and cut scores on the selected outcome measure of interest.
- Review the NCRTI Screening Tools Chart (www.rti4success.org/screeningTools) to view the classification accuracy, reliability, and validity of reviewed tools.
- For tools not on the NCRTI Screening Tools Chart, review the technical manual or contact the vendor for information about classification accuracy, reliability, and validity.

Using Multiple Screening Measures to Improve Classification Accuracy

Research articles have frequently reported the advantages of using more than one screening assessment for reading, including greater screening accuracy (increased rates of sensitivity and specificity) from using multiple measures than from using a single measure (Catts, Fey, Zhang, & Tomblin, 2001; Compton, Fuchs, Fuchs, & Bryant, 2006; Compton et al., 2010; Fletcher et al., 2002; Francis et al., 2005; Jenkins, Hudson, & Johnson, 2007).

Compton and his colleagues (2006) completed extensive research to determine screening assessments and analysis techniques that would increase the accuracy of identifying first grade students at risk for reading problems. They concluded that the practice of using multiple screening measures (i.e., adding five weeks of progress monitoring with word identification) in the fall of first grade showed significant improvement in classification accuracy, with sensitivity at 90 percent and specificity at 82.7 percent.

> **Recommendations for Practitioners**
> - To identify first graders at risk for reading difficulties, use multiple measures. This will result in predictions that are more accurate (i.e., increased sensitivity and specificity) than those based on a single assessment.
> - For first graders, consider adding data from five weeks of progress monitoring with word identification fluency to improve correct classification of students.

Other researchers, too, have commented on adding progress monitoring to the screening process. Jenkins, Hudson, and Johnson (2007) referred to the use of progress monitoring as part of the screening process as "the progress monitoring route" (as opposed to "the direct route," which involves no progress monitoring). Although they acknowledged that the progress monitoring route shows better at-risk identification accuracy, they expressed concerns that including progress monitoring delays intervention; on the other hand, they noted that, while the direct route leads to earlier intervention, it may result in students being incorrectly identified as being at risk for not meeting proficiency on an outcome assessment.

Using a Two-Stage Screening Process

Compton et al. (2010) concluded that a two-stage screening process resulted in high sensitivity and specificity values, with the added advantage of using minimal administration time. The first stage focused on a single measure (phonemic decoding efficiency) administered to all first-grade students to efficiently identify all students who were at low risk for developing reading difficulties—the true negatives. The second stage involved assessing students who, on the basis of the results of the first stage screen, might be at risk. These students (all but the true negatives) were given the more time-consuming assessment battery, which included word identification fluency with progress monitoring.

> **Recommendations for Practitioners**
> - For first grade, consider a two-stage screening process to increase efficiency and effectiveness:
>   - Stage 1: Screen all students by using a single-measure assessment of phonemic decoding efficiency to eliminate as many true negatives as possible.
>   - Stage 2: Assess all students other than the true negatives (those correctly classified as not at risk).

References

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (2001). Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools, 32,* 38–50.

Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., et al. (2010). Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology, 102,* 327–340.

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.

Fletcher, J. M., Foorman, B. R., Boudousquie, A., Barnes, M. A., Schatschneider, C., & Francis, D. J. (2002). Assessment of reading and learning disabilities: A research-based intervention-oriented approach. *Journal of School Psychology, 40,* 27–63.

Francis, D. J., Fletcher, J. M., Stuebing, K. K., Lyon, G. R., Shaywitz, B. A., & Shaywitz, S. E. (2005). Psychometric approaches to the identification of LD: IQ and achievement scores are not sufficient. *Journal of Learning Disabilities, 38,* 98–108.

Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for at-risk readers in a response to intervention framework. *School Psychology Review, 36,* 582–600.

Thorndike, R. M., Thorndike-Christ, T. (2008). *Measurement and evaluation in psychology and education.* New Jersey: Prentice Hall.

About the National Center on Response to Intervention

Through funding from the U.S. Department of Education's Office of Special Education Programs, the American Institutes for Research and researchers from Vanderbilt University and the University of Kansas have established the National Center on Response to Intervention. The Center provides technical assistance to states and districts and builds the capacity of states to assist districts in implementing proven response to intervention frameworks.

National Center on Response to Intervention
http://www.rti4success.org

National Center on Response to Intervention

1000 Thomas Jefferson Street, NW

Washington, DC 20007

Phone: 877–784–4255

Fax: 202–403–6844

Web: http://www.rti4success.org

AIR
AMERICAN INSTITUTES FOR RESEARCH®

IDEAs that Work
U.S. Office of Special Education Programs