

Universal Screening for Reading Problems: Why and How Should We Do This?

by [Joseph R. Jenkins, Ph.D.](#) , University of Washington, Seattle and [Evelyn Johnson, Ed.D.](#), Boise State University, Boise, ID

Why Universal Screening?

Response to Intervention (RTI) is a multi-tier instructional and service delivery approach designed to improve student learning by providing high-quality instruction, intervening early with students at risk for academic difficulty, allocating instructional resources according to students' needs, and distinguishing between students whose reading difficulties stem from experiential and instructional deficits as opposed to a learning disability. A central component of all RTI models is early screening of all students to identify those at risk for academic difficulties.

RTI draws on ideas of prevention science. In a prevention approach, schools do not wait for students to fail before coming to their assistance. Instead, they screen all students to identify those who, despite a strong general education program (Tier 1), are on a path to failure. To have any chance of escaping this adverse path, students must obtain immediate help (Tier 2). When RTI is implemented fully, reading, math, writing, and behavior screening is conducted with all students and those determined to be at risk for developing difficulties in one or more of these areas receive targeted evidence-based interventions. Students then move between tiers according to their academic and behavioral rates of progress.

In this article, we use reading screens to illustrate the process of universal screening, its benefits, and its challenges. We chose reading for several reasons. Theory and research on reading development, and on its risk factors and their measurement, is relatively strong. Moreover, reading is the centerpiece of the elementary school curriculum, yet 26% of American 8th graders are unable to read at even a basic level according to the 2007 National Assessment of Educational Progress (NAEP; Lee, Grigg, & Donahue. 2007).

ADDITIONAL ARTICLES

[Screening for Reading Problems in an RTI Framework](#)

[Progress Monitoring Within a Multilevel Prevention System](#)

[The RTI Data Analysis Teaming Process](#)

ADDITIONAL RESOURCES

[RTI Talk: RTI and Universal Screening](#)

[National Center on Response to Intervention Tools Charts](#)

[National Center on Student Progress Monitoring](#)

[Research Institute on Progress Monitoring](#)

Where Does Screening Fit in RTI?

The way that schools identify students for Tier 2 intervention varies according to the type of RTI model that is implemented. In *direct route models*, students identified as at risk by a screening process are immediately provided Tier 2 intervention (e.g., Jenkins, Hudson, & Johnson, 2007; Vellutino et al., 1996; Vellutino, Scanlon, Zhang, & Schatschneider, 2007, October 30). By contrast, in progress monitoring, or *PM route models*, universal screening identifies potentially at-risk students whose progress is then monitored for several weeks. Whether these students enter Tier 2 depends on the level of their performance and rate of growth on PM measures (Compton, Fuchs, & Fuchs, 2007). The PM route yields marginally better identification accuracy than the direct route, but it also postpones intervention during the PM phase. By contrast, the direct route leads to earlier intervention, but without PM to catch screening errors more students are mistakenly identified as being at risk. In both models screening may be a singular event or conducted periodically (e.g., fall, winter, spring).

What Do We Seek in a Screen?

Screening approaches should satisfy three criteria (Jenkins, 2003). First is classification accuracy—a good screen accurately classifies students as at risk or not at risk for reading failure. Second is efficiency—because screening is universal, the procedure must not be too costly, time-consuming, and cumbersome to implement. Good screens can be administered, scored, and interpreted quickly and accurately. Third is consequential validity—overall, the net effect for students must be positive (Messick, 1989). This means students identified as at risk for failure must receive timely and effective intervention, and no students or groups should be shortchanged.

How Screens Are Created—Three Steps

The purpose of screens is to predict an outcome months or years in advance. The first step in creating a screen is to define the future outcome the screen seeks to predict (e.g., unsatisfactory reading ability). *Future* is relative, ranging from several months to several years and marked by specific points in the school curriculum (e.g., end of Grades 1, 4, 8, and 12). Reading screens attempt to predict which students will score poorly on a future reading test (i.e., the criterion measure). Some schools use norm-referenced test scores for their criterion measure, defining poor reading by a score corresponding to a specific percentile (e.g., below the 10th, 15th, 25th, or 40th percentile). Others define poor reading according to a predetermined standard (e.g., scoring below “basic”) on the state’s proficiency test. The important point is that satisfactory and unsatisfactory

reading outcomes are dichotomous (defined by a cut-point on a reading test given later in the students' career). Where this cut-point is set (e.g., the 10th or 40th percentile) and the specific criterion reading test used to define reading failure (e.g., a state test or SAT 10) greatly affects which students a screen seeks to identify.

The second step in creating a reading screen is the identification of early predictors of later reading outcomes. To be effective, a reading screen must be sensitive to different levels of reading development. In kindergarten, children develop phonemic awareness, letter and sound knowledge, and vocabulary. In 1st and 2nd grades, they grow in phonemic spelling, decoding, word identification, and text reading. At higher grades, they gain in ability to comprehend increasingly difficult texts. Thus, screening measures valid for beginning 1st graders (e.g., word identification fluency) differ from those valid for kindergarten (e.g., letter naming fluency) or 2nd grade students (e.g., oral reading skill).

The third step in designing a reading screen is to determine a cut-point on the screening measure(s) that identifies students at risk for failing the future criterion test. To identify a cut-point, researchers work backwards—first selecting students who failed the criterion (later) reading measure, then identifying the score on the screening measure that best distinguishes those students from the students who passed the criterion measure. This cut-score is then used to screen subsequent groups.

How Do We Evaluate the Accuracy of a Screening Process?

A perfect screen would distinguish every student who needs intervention from every student who doesn't—a simple dichotomy with no hedging of bets. Unfortunately, because the perfect screen doesn't exist, schools have to weigh the trade-offs of over- and underidentifying students as at risk. In the counterintuitive language of screening, students who score below the cut-point on the screening measure are labeled "positives," meaning they are at risk for the problem condition, (i.e., future reading failure). Students who score above the cut-point on the screening test are labeled "negatives," meaning they are not at risk for future reading failure.

Screens can be correct (or true) in two ways: a) "True positives" are individuals who fail the screening measure (the predictor) and the later outcome measure (the criterion); b) "true negatives" are individuals who pass both the screen and the later criterion measure. Screens can also be incorrect (or false) in two ways: a) "False positives" are individuals who fail the screen but pass the later criterion measure; b) "false negatives" are individuals who pass the screen but fail the later criterion measure.

Two statistics, sensitivity and specificity, are used to gauge a screen's accuracy in classifying students. *Sensitivity* gauges the screen's ability to identify individuals who will pass the criterion measure. Specificity is calculated by dividing the number of true negatives by the total number of individuals who perform successfully on the outcome measure.

Sensitivity increases as the screen correctly identifies more and more of the students who have later reading difficulties, whereas specificity increases as the screen correctly identifies more and more of the students who read satisfactorily. Sensitivity is easily manipulated by adjusting cut-scores. For example, if we raised the cut-score on a kindergarten letter naming screen from 6 letters correct to 12 letters correct, we would likely identify all or nearly all of the children who will eventually fail the criterion test at the end of 1st grade. The increased sensitivity would be offset by decreased specificity because raising the cut-score means the screen will overidentify many students who are not really on the path to reading failure (i.e., increasing false positives).

The Consequences of Screening Mistakes

When screening mistakenly overidentifies many students as at risk (false positives), schools spend precious intervention resources on students who don't need the extra help. This may result in spreading intervention resources so thinly that they are insufficient for students who really need them. Therefore, screens should strive to correctly identify 80% or more of the students who are not at risk.

Overidentifying students as at risk taxes school resources, but an even more serious problem is overlooking students who can't succeed without Tier 2 assistance. In an RTI model, accurately identifying all or nearly all of the truly at-risk students is more important than is accurately identifying students who are truly not at risk. Screens should strive to identify correctly at least 90% of the students who will later exhibit reading failure.

How Screening Errors Play Out

Using the previously mentioned NAEP results, one-quarter of students in a typical class of 28 first graders are on the path to reading failure. Whereas a perfect screen would identify all seven of the true positives in this class, a very good screen, one with 90% sensitivity and 80% specificity, would miss one (1) of the truly at-risk students and (over) identify 4 students who are not really at risk. If a school has several 1st grade classes, the number qualifying for Tier 2 services multiplies accordingly. These numbers mount further if the school screens all K–6 students. For example, in an average size elementary school of 400 students, a screen with 90% sensitivity and 80% specificity will identify 150 students as at risk, missing 10 truly at-risk

students and overidentifying 60 students who are not really at risk. It is easy to see the importance of screening accuracy.

Some Key Ideas In Choosing A Procedure For Universal Screening

Multiple Measures Are More Accurate Than a Single Measure

Researchers who use a screening battery (multiple measures) obtain better classification accuracy (Jenkins & O'Connor, 2002). For example, O'Connor and Jenkins (1999) distinguished at-risk and typically developing kindergarteners better by using a combination of measures (Letter Name Fluency, Phonemic Segmentation, and Syllable Elision) than any single measure. Working with other age groups, Compton et al. (2007) and Foorman et al. (1998) also reported better screening accuracy for a battery of measures than for single measures. Adding teachers' ratings of child attention and behavior to the screening battery can also enhance its accuracy (Davis, Lindo, & Compton, 2007; Ritchie & Speece, 2004).

Screening Measures Should Address Both Print and Comprehension Skill

Reading is a multidimensional ability and screens should reflect this. Gough and Tunmer's (1986) analysis of reading ability into two broad components (the ability to read words and comprehend language) is helpful in thinking about potential screening measures. Screens used in the primary grades typically focus on print skills (letter and word reading). Such screens have proven valid for identifying students who eventually perform poorly on tests of word reading and reading fluency, but they miss some students with limited language comprehension skills whose print skills are in the normal range (Catts, Fey, Zhang, & Tomblin, 1999). It is equally critical to identify this latter group of students for Tier 2 intervention. Thus, screens should include measures (e.g., vocabulary) that predict later appearing reading comprehension problems (Davis et al., 2007).

Once Is Not Enough

Screening should occur every year across the elementary grades. To allow early intervention, schools should screen early in the year so that they can allocate instructional resources intelligently. Students with scores below the screening cut-point should be directly assigned to Tier 2 (the direct route model) or targeted for progress monitoring (the PM model). Compton et al. (2006) were successful in reducing classification errors for beginning 1st graders by combining screening results with five (5) weeks of progress monitoring. Alternatively, if schools use a direct route model for immediate assignment to Tier 2, they should rescreen periodically (e.g., in December and March) to catch false positives and identify students who were missed on the first screen (Vellutino et al., 2007). Progress monitoring and/or periodic within-year rescreening is especially important for students who score near the screening cut-point, where measurement errors have their

greatest effect on decision making.

Screening Cut-Points Will Vary

Different states, districts, and schools designate unsatisfactory reading in different ways (e.g., a criterion score below the basic level on a state reading test). This means that a school or district must select its screening cut-point in accord with the specific criterion test it will use later to measure reading ability. Most important, as Tier 1 instruction improves schools will have to revisit their screening cut-points because improved Tier 1 instruction will raise long-term outcome and affect students' performance on screening tests.

How Should We Screen?

Because it is user-friendly, the DIBELS assessment system is a frequent choice for a screening and progress-monitoring tool for RTI. Unfortunately, sensitivity and specificity levels for DIBELS are far from the ideal of 90% and 80%, respectively, for predicting reading outcomes measured by standardized tests. For example, Schatschneider (2006) reported sensitivity and specificity levels of 52% and 85% for 1st grade students, and Riedel (2007) reported 68% for both statistics at this grade level. Better screens are available. (See Jenkins et al., 2007, for a review of screening accuracy.) Below, we identify several screening measures that have achieved the 90% / 80% criterion for sensitivity and specificity.

Kindergarten

The most successful screening measures in kindergarten have used various combinations of Letter Naming Fluency, Letter Sound Identification, blending onset-rimes, phoneme segmentation, and sound repetition (Foorman et al., 1998; O'Connor & Jenkins, 1999).

1st Grade

The most successful screening measures for 1st grade students have used various combinations of Word Identification Fluency, Letter Naming Fluency, Letter Sound Identification, phoneme segmentation, sound repetition, vocabulary, and word identification fluency (Compton et al., 2007; Foorman et al., 1998; O'Connor & Jenkins, 1999).

2nd Grade and Beyond

There are surprisingly few studies of screening accuracy beyond Grade 2. In many states, annual results of district- or statewide achievement and standards tests can be used to identify at-risk students. These should result in reasonably good predictions, given that spring–spring and fall–spring achievement correlations are typically strong. Schools should consider a combination of oral reading fluency and more comprehension-oriented assessments, like the Scholastic Reading Inventory (n.d.) and the 4Sight Benchmark Assessments (Slavin & Madden, 2006) that are designed for within-year periodic progress monitoring/rescreening.

No One-Best Approach

Whether we can settle on one-best approach for screening is another matter. Local preferences for criterion measures, the criterion performance level that designates unsatisfactory reading, and tolerance for under- and overidentification rates will lead to different choices of screening measures and cut-points. What is critical is ensuring that screens select all or nearly all students in a school who require secondary intervention.

References

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal study. *Scientific Studies of Reading*, 3, 331–361.

Compton, D. L., Fuchs, D., & Fuchs, L. S. (2007). *The course of reading and mathematics disability in first grade: Identifying latent class trajectories and early predictors*. Manuscript submitted for publication.

Davis, G. N., Lindo, E. J., & Compton, D. L. (2007). Children at risk for reading failure: Constructing an early screening measure. *Teaching Exceptional Children*, 39(5), 32–37.

Foorman, B. R., Fletcher, J. M., Francis, D. J., Carlson, C. D., Chen, D., Mouzaki, A., et al. (1998). *Technical report: Texas Primary Reading Inventory* (1998 Edition). Houston: Center for Academic and Reading Skills, University of Texas Health Science Center at Houston and University of Houston.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7, 6–10.

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

Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for service delivery in an RTI framework: Candidate measures. *School Psychology Review*, 36, 560–582.

Jenkins, J. R., & O'Connor, R. E. (2002). Early identification and intervention for young children with reading/learning disabilities. In R. Bradley, L. Danielson, & D. P. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 99–150). Mahwah, NJ: Erlbaum.

Johnson, E. S., Mellard, D. F., Fuchs, D., & McKnight, M. (2006). *Response to intervention: How to do it*. Lawrence, KS: National Research Center on Learning Disabilities.

Lee, J., Grigg, W., & Donahue, P. (2007). *The Nation's Report Card: Reading 2007* (NCES-2007-496). Washington, DC: National Center for Education Statistics, Institute for Education Sciences, U.S. Department of Education.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: Macmillan.

O'Connor, R. E., & Jenkins, J. R. (1999). The prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading*, 3, 159–197.

Riedel, B. W. (2007). The relationship between DIBELS, reading comprehension, and vocabulary in urban first-grade students. *Reading Research Quarterly*, 42, 546–567.

Ritchie, K. D., & Speece, D. L. (2004). Early identification of reading disabilities: Current status and new directions. *Assessment for Effective Intervention*, 29(4), 13–24.

Schatschneider, C. (2006). *Reading difficulties: Classification and issues of prediction*. Paper presented at the Pacific Coast Research Conference, San Diego, CA.

Slavin, M. R., & Madden, N. A. (2006). *4Sight Benchmark Assessments*. Baltimore: Success for All Foundation.

Scholastic Reading Inventory. (n.d.). Retrieved December 31, 2007, from <http://teacher.scholastic.com/products/sri>.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Chen, R., Pratt, A., & Denckla, M. B. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology*, 88, 601–638.

Vellutino, F. R., Scanlon, D. M., Zhang, H., & Schatschneider, C. (2007, October 30). Using response to kindergarten and first grade intervention to identify children at-risk for long-term reading difficulties. *Reading and Writing*. doi:10.1007/s11145-007-9098-2

Visit www.rtinetwork.org for more information on this topic.

Copyright © 1999-2018 National Center for Learning Disabilities, Inc. All Rights Reserved.