4-1-2008

How Should We Screen for Reading Problems

Evelyn S. Johnson
Boise State University

Michael Humphrey
Boise State University

Rebecca McLenna
Boise State University

How Should We Screen for Reading Problems

Evelyn S. Johnson, Boise State University, ID
Michael Humphrey, Boise State University, ID
Rebecca McLenna, Boise State University, ID

 Johnson, Ed.D. and Humphrey, Ed.C. are Assistant Professors of Special Education, and McLenna is Graduate Assistant in Special Education.

Abstract
Response to intervention (RTI) is increasingly used to organize reading instruction and assessment. One component of an RTI framework is the universal screening of students to determine who is at risk for developing reading problems. For screening to be effective, it must be efficient, accurate and have positive consequences for its use. This article discusses the current approaches to screening and their limitations, and provides recommendations for improvement.

Introduction
Screening students for reading problems is becoming more commonplace with the passage of Reading First and the Individuals with Disabilities Education Act (IDEA 2004). Both acts focus on the use of an RTI framework that includes the following components for organizing reading instruction: 1) high quality reading instruction, 2) universal screening, 3) interventions for at-risk students, 4) progress monitoring of students' response to the intervention, and 5) subsequent tiers of intervention for students who require more intense, targeted support. In this article, we focus on the importance of screening as the critical juncture in the RTI process by describing our current understandings of the construct of reading, providing a discussion on the critical requirements of an effective screen, analyzing the research on existing screening instruments, and concluding with recommendations to improve our current screening procedures.

Component Skills of Reading
Reading is a complex process comprised of many component skills. Several models of reading ability exist, but the most fundamental is the Simple View of Reading (Gough & Tunmer, 1986). The simple view states that reading (R) is a function of a student's decoding (D) and comprehension (C) ability, or D x C = R. Numerous studies provide validity evidence for the simple view, though in recent years, a number of studies have argued for its expansion in order to more clearly explicate the two subcomponents (D & C), as well as to include more comprehensive measures of literacy (R). In general, however, D comprises subskills such as phonological awareness, letter naming and letter sound knowledge. C comprises subskills such as listening comprehension and vocabulary. R is typically measured by standardized reading tests.

Although developmental models of reading seem to suggest that basic skills (D) must be acquired prior to the development of higher order skills (C), research suggests that both components (and their subcomponents) play important and continuous roles in the reading process, though the relative contribution of these components may vary depending upon the stage of reading development. For example, vocabulary knowledge is important for pre-reading skills (Wise et al. 2007), becomes less pronounced as children acquire decoding skills, and later reemerges as a significant predictor for later grades.

Models of early reading ability reinforce the understanding that reading ability does not necessarily progress through a lock step acquisition of basic skills. Wise et al. (2007) found that vocabulary knowledge is significantly related to pre-reading skills. Additionally, consistent with the Simple View, listening comprehension skills are found to significantly predict word identification abilities in early readers (Wise et al, 2007). This suggests the relationship between D and C is bidirectional (Adams, 1990). In other words, strong comprehension and vocabulary abilities facilitate word reading skill, and strong word identification ability promotes better comprehension.

One component of reading ability that 'cuts across' stages of reading acquisition is fluency. Fluency has emerged as an important component process of reading ability at all levels, including letter and word identification as well as the fluent reading of text. The theory underlying fluency's role in the reading process is that students who are more fluent in their skills have stronger representations of words and/or letters (Adams, 1990). When students can more fluently read a text, they are able to devote more processes to understanding the meaning of the text (LaBerge & Samuels, 1974). Much remains to be learned regarding the role of fluency in reading development however, and some argue that the current focus on developing fluency through rapid and repeated readings is not an effective teaching strategy (Samuels, 2007).

Overall, models of reading ability are generally consistent in the identification of phonological awareness, word reading, comprehension, vocabulary, and fluency as important component processes that contribute to a student's reading ability. These components do not operate in a lock-step fashion, but are recursive and vary in importance depending upon the phase of reading development. Understanding the component processes of reading has implications for developing effective screening measures.

What makes for an effective screening instrument?

Jenkins (2003) outlined three criteria to evaluate screening instruments: 1) accuracy, 2) efficiency, and 3) consequential validity. The primary purpose of a screening instrument is to accurately classify students into categories. Therefore, an effective screening instrument is one that correctly categorizes students as at-risk or not at risk for reading problems. Because screening is conducted universally (e.g. all students are screened), it is preferable to find the most efficient means of doing this. Which screening instrument takes the least amount of time to administer, is the most cost effective and easiest to implement? Finally, because screening leads to important instructional decisions for students, the overall consequences of a screening instrument's use must be positive.

Accuracy. The accuracy of a screening measure is most often reported through its classification rate, or sensitivity and specificity. Sensitivity refers to the percent of students who are correctly identified by the screen as at-risk for developing later reading problems. Specificity refers to the percent of students who are not at risk who are correctly identified. No screening measure is completely accurate, and rates of identification are easily manipulated by changing cut scores. For example, a school is likely more willing to overidentify students as being at-risk for reading problems, so raising the score needed to 'pass' a screen is an easy way to ensure that all students who are at-risk are identified. However, overidentifying students as at-risk may waste intervention resources; including too many students in interventions means that they may be less effective for those students who most need them.

Efficiency. Because screening is conducted schoolwide, a screening instrument must be efficient. If it is too cumbersome, teachers may be reluctant to implement the screening procedures (Jenkins, 2003). If a screening procedure is too costly, schools may not be able to

afford implementation. The results of a screen must also be easily understood. It is safe to argue that schools place efficiency as paramount in their consideration of which methods to adopt.

Consequential validity. Messick (1989) argued that the essential aspect of a test's validity was the overall consequences of the uses and interpretations of the test. Screening measures are meant to quickly determine who is at risk for developing reading problems so that interventions may be provided. Are the right students identified? Does the implementation of screening result in improved service delivery and outcomes for all students? What are some of the unintended consequences of implementation of the screening measure?

Many screening instruments to detect reading problems are available and in use, especially in schools that follow an RTI model. How do these measures rate in terms of efficiency, accuracy and consequential validity?

Current screening measures

Because the Reading First act targets K-3, and research supports early intervention of reading problems, much of the research on screening targets these early grade levels. Across these grade levels, a variety of instruments have been examined. A review by Jenkins, Hudson & Johnson (2007) found that in Kindergarten, screening instruments assessed some aspect of phonological awareness, letter knowledge, and fluency of these skills. Their review also found that when a combination of skills, such as blending and segmenting, were targeted, the screens were more accurate. In first grade, the content of screens changed depending upon the time of year (Jenkins et al.). For example, in fall of first grade, screens were similar to those used in Kindergarten. In spring of first grade and fall of second grade, measures of word reading and passage reading were used, though word reading was a better predictor (Jenkins et al.). By the end of second grade and beyond, oral reading fluency (ORF) tended to be the only screening instrument reported in the screening research with varying degrees of accuracy (Jenkins et al.).

Several important points are highlighted in the Jenkins et al. review. First, the reported accuracy of the screens vary, but many fall below rates at which schools should feel comfortable making important decisions about student performance. For example, in a study examining ORF as a screen, 74% sensitivity was achieved, but the screen was administered only two weeks prior to the outcome measure, and incorrectly classified 28% of the students who did not pass the outcome measure (McGlinchey & Hixson, 2004). Second, nearly all of the studies focused on only one component process of reading skill. Even when studies used a combination of measures, the skills assessed typically fell into one or two of the component processes (e.g. several tasks measured various aspects of phonological awareness). Finally, none of the studies used the same criterion measure. This is significant because the classification rate of similar screens varied depending upon the outcome measure used (Jenkins et al, 2007).

The problems with current screening measures

Recall that the primary goal of screening is to accurately identify students most at-risk for developing later reading problems. One of the most significant limitations with the current options for screening tools is that they fail on the accuracy front. For example, for students at the end of grade 2 and above, the primary screening measure in use for identifying reading problems is ORF, yet in the published screening studies, ORF has varying levels of sensitivity, none of which reach more than 75%. In one estimate, Pressley, Hilden, and Shankland (2005) suggest that the cost of collecting ORF measures three times a year is about $30 per student or the equivalent of paying a half time reading specialist in a small district for one school year. If a screen fails to accurately classify students, the authors question whether the money might not be put to better use by providing better reading instruction.

The classification accuracy in kindergarten and first grade is higher (Jenkins et al, 2007). However, in a critique of current approaches to screening for reading problems, Paris (2005) notes that the use of constrained skills (e.g. those skills with a finite set of performances such as naming letters or sounds; those used in the K-1 grade screens) can account for nearly all of the variance in reading outcomes because of the statistical properties of the measures and the characteristics of the sample. In other words, because these skills are assessed at a time when they are developing, there is a significant amount of variance in performance, which in turn provides greater capacity to correlate with the outcome measures (Pearson, 2006). While limiting screening instruments to a small number of skills may satisfy the efficiency criterion of effective screens, the problem with this approach is that it excludes other factors that comprise the reading construct, to include those with demonstrated causal relationships, such as vocabulary knowledge (Wise et al, 2007).

Narrowing our screens to focus on only one component of reading may be efficient, and in some cases, somewhat accurate, but it tends to have negative consequences on instruction. Pearson (2005) outlines principles of good assessment policy, one of which is that, "other things being equal, people will teach to tests, even if in their heart of hearts they know they should not" (p. xvii). The consequences of using screening measures that encompass only one basic reading skill are that it tends to lead to poor instructional decisions, such as trying to get students to read faster; spending time teaching students to decode nonsense words; and placing an inordinate amount of time and resources on developing basic reading skills at the expense of other important reading components such as vocabulary and language development. The logic to teaching to the test is reasonable enough; students who are more fluent readers tend to perform better on outcome measures; therefore, if I can get a student to read faster that will improve overall reading ability. However, the failure in this logic is also clear. Models of reading suggest that the relationships among many of the components are multi-directional (Adams, 1990). Reading faster won't necessarily improve comprehension. Increasing decoding skills without also increasing vocabulary knowledge won't necessarily improve comprehension. Teaching to the test misses the point of the purpose of the screening instrument.

Some possible solutions

There are basically two main approaches to address the problems with screens. They are not mutually exclusive, but for ease of discussion are elaborated separately here. The first approach is to help educators better understand the purpose of screening. For example, consider a vision screen, a process familiar to most of us. The examinee reads the lines on a chart a given distance away. Depending upon performance, one is either considered to have normal vision or is referred for an eye exam, during which vision problems are more carefully scrutinized. The follow-up exam will include many more tests of vision to determine the exact difficulties (i.e. are we near-sighted or far sighted?) and to identify the appropriate solution (i.e. what kind of prescription do we need?). In this scenario, few would advocate having someone memorize a screening chart, nor would they train a person to read a chart in preparation for the screen, because these things would not improve the person's vision. Corrective lenses tailored to improve the individual's specific vision problems will.

In the vision example, the purpose of the screen is clear, the corrective actions work immediately, and it is easy to see why teaching to the test is not helpful. Helping educators understand the role of screening is necessary, but the translation from vision to reading is complex. This is in part because at different developmental stages, different skills may predict reading ability, so the same screen at all stages is not possible. Also, even though there are strong predictors of reading ability at different stages, these predictors do not by themselves comprise overall reading ability. A poor screening outcome should indicate the need for more in-depth diagnosis, but our current system for screening does not necessarily promote this. If we mirrored reading screens to be like vision screens, we would, upon receiving a poor screening result for a

student, examine other areas more in-depth, to include miscue analysis and measures of language and comprehension so that we might provide a more tailored intervention to support student needs (Paris, 2005).

The second approach to improving screening for reading problems is to change the nature of the screens. We may, for example, include a more comprehensive assessment of reading as a screening measure across all grade levels by including tasks that represent all of the components that comprise reading ability. Even if the prediction rate is no better than when using a single skill, there may be instructional utility in including a more comprehensive measure of reading. If a more comprehensive measure to screen for reading problems were used, that could prompt teachers to develop more balanced and comprehensive instructional approaches to reading for all students. Focusing teachers on all that is important about reading may help increase their knowledge of the construct and enable them to develop and implement more appropriate, individualized interventions.

Conclusions

The focus on screening for early reading problems is a welcome development given what we know about the importance of early intervention and the importance of reading for overall school achievement. In our attempts to improve our reading instruction and determine who is in need of intervention, however, it is critical that we are mindful of the complexity of the reading construct; the complexity of the screening process; and the importance of getting this right.

References

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

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

Individuals with Disabilities Education Improvement Act of 2004 (Public Law 108-446).

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

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

LaBerge, D, & Samuels, S. J. (1974). Toward a theory of automatic information Processing in Reading. Cognitive Psychology, 6, 293-323.

McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments of reading. School Psychology Review, 33, 193-203.

Messick, S. (1989). Validity. In R. L. Linn (Ed), Educational Measurement, 3rd ed., New York: Macmillan.

Paris, S. G. (2005). Reinterpreting the development of reading skills. Reading Research Quarterly, 40 (2), 184-202.

Pearson, P. D. (2006). Foreword. In K. S. Goodman (Ed). The Truth about DIBELS: What it is, what it does. Portsmouth, NH: Heinemann

Pressley, M., Hilden, K., & Shankland, R., (2005). An evaluation of end-grade-3 Dynamic Indicators of Basic Early Literacy Skills (DIBELS): Speed Reading Without Comprehension, Predicting Little (Technical Report). East Lansing, MI: Literacy Achievement Research Center.

Samuels, S. J. (2007). The DIBELS tests: Is speed of barking at print what we man by reading fluency? Reading Research Quarterly, 42, 563-6.

Wise, J. C., Sevcik, R. A., Morris, R. D., Lovett, M. W. & Wolf, M. (2007). The relationship among receptive and expressive vocabulary, listening comprehension, pre-reading skills, word identification skills, and reading comprehension by children with reading disabilities. Journal of Speech, Language, and Hearing Research, 50, 1093 – 1109.