Four Dyslexia Screening Myths That Cause More Harm than Good in Preventing Reading Failure and What You Can Do Instead

VanDerHeyden, Amanda M.; Burns, Matthew K.
Communique, v45 n7 p1, 26, 28 May 2017

Research-Based Practice

Four Dyslexia Screening Myths That Cause More Harm Than Good in Preventing Reading Failure and What You Can Do Instead

By Amanda M. VanDerHeyden & Matthew K. Burns

Thirty-nine states have recently passed legislation that focuses on identifying and remediating dyslexia (Eide, 2016). Most of the recent legislation requires schools to screen for dyslexia, which has resulted in new screeners being developed and purchased all across the country. At a time when schools are administering more screening to detect risk for reading failure than at any time in the history of education, it is interesting that legislative mandates are prescribing more reading screening in the name of better identification and treatment of dyslexia.

Given that most schools already conduct reading screening multiple times per year, often using multiple measures, it makes sense to revisit why all this screening may not be giving schools the desired return on investment that they are after. The purpose of this article is to equip school psychologists with an understanding of four common screening myths that cause more harm in screening than good, and to share specific tactics for smarter screening.

Myth 1: More Screening Can Only Improve Correct Identification of Students With Dyslexia

Most people are familiar with screening accuracy and likely participate in annual medical screening on the advice of their physicians. Two metrics are commonly used to quantify a screener's accuracy: sensitivity (test correctly identifies those who truly have a condition) and specificity (test correctly identifies those who truly do not have a condition). Identifying a student as having dyslexia when the student does not is a false positive error, and missing someone who truly has dyslexia is a false negative error. In education, it is not uncommon for error rates to range from 50%–60%, meaning if a school assesses 100 children for whom 20 are "true positives" (i.e., truly have dyslexia), then most of the 20 (approximately 16–18) will be identified, but 50 to 60 students will be identified as false positive errors in the process. Just think of what that means for a moment. To find the "real" 2 in 10 children with dyslexia, we will misidentify 4 children who do not have dyslexia and we will still miss 1 to 2 of every 10 true positives.

There is almost an insatiable appetite for screening in schools, and many school leaders believe that more screening can only return positive benefits for students. After all, what harm can it do? The answer is that screening when the probability of a condition actually being present is very, very low can do a great deal of harm. If we give pregnancy tests to all humans, then we will end up receiving positive pregnancy test results in males. Not only will that decision to screen result in unnecessary expenditures of resources having screened someone who could not ever have been pregnant; but because of false-positive error rates inherent to the test, there will also be some measurable percentage of incorrect/bad information (i.e., positive pregnancy results in males).

Myth 2: Failing to Learn to Read Means the Child Most Likely Has Dyslexia

When a child struggles to learn to read, that is often the first sign that the child may have dyslexia. However, the hallmark of dyslexia is not poor reading performance; rather, it is poor reading performance in the face of effective reading instruction. Most children who struggle to learn to read do not have dyslexia, which creates a terrible diagnostic conundrum. We suggest that poor reading performance should signal the need for screening. Screening then must combine controlled doses of instruction to rule out lack of instruction as a cause for poor reading performance.

Myth 3: Screening Accuracy for a Published Tool Will Be Similar Across Schools

Screening measures are inherently unstable across settings, no matter what the publishers say. All screening tools are systematically affected by the prevalence context in which they are used. What do we mean by prevalence context? If you visit your physician with upper respiratory illness symptoms, you have had a known exposure to flu, and flu is prevalent enough in your community, your physician will most likely tell you that there is no need to test you for the flu because you most likely have the flu no matter what the test says. In other words, the probability of your having the flu exceeds the probability that you do not have the flu given a negative flu test result. In education, there are schools and there are groups of children, among whom the risk of academic failure is so high, that even when children pass the screening test, they will have an intolerably high risk of academic failure. In other words, these children are likely to fail no matter what the test says. The reverse is also true. There are schools and groups of children among whom risk is so low, that a failed screening test may be more likely false than true. These are not novel concepts to psychological screening (Meehl & Rosen, 1955); yet, in education, clinical application of screening ignores prevalence in determining actual risk for students, which causes incorrect decisions (VanDerHeyden, 2013).

Myth 4: Screening Improves Reading Performance

Readers may be surprised to learn that there is not a direct positive relationship between screening assessments and improved reading outcomes. In medicine, mammograms are encouraged because women who have them at certain ages are less likely to die of breast cancer than similar-age women who do not have mammograms (Pace & Keating, 2014). This level of evidence for screening is a step beyond the basic decision accuracies of sensitivity and specificity that we report in education. Effective screening should predict future academic outcomes (Jenkins, Hudson, & Johnson, 2007) that are aligned with the school's curriculum and instruction (Ikeda, Neeson, & Witt, 2008). Most dyslexia screeners do not provide instructionally relevant data, which results in an expenditure of considerable resources with little opportunity to improve student outcomes. Screening alone does not improve outcomes. The screening must lead to effective remediation or the screening is not useful. Returning to the flu test for a moment, in deciding to give a flu test to a patient, the physician will also consider, "What difference does it make in my treatment?" For example, if the window within which a medication might be administered to reduce the duration of flu has passed, then giving the flu test has no treatment utility. The concept of treatment utility arose in psychology (Hayes, Nelson, & Jarrett, 1987). School psychologists must give voice to the idea of treatment utility in assessment, asking, "How will this information benefit this child if we collect it?"

The screening scale solution. Because 39 states now require schools to screen for dyslexia, there have been a number of newly developed screeners. The Shaywitz (2016) Dyslexia Screen is being used with increasing frequency and provides one example of the potential for errors in screening. The author is a leader in the field of dyslexia or reading disabilities, and

using a screening like the Shaywitz Dyslexia Screen may feel like a tidy solution to a legislative mandate for dyslexia screening. However, such a solution is not tidy.

The estimates of sensitivity and specificity reported by the publisher for the Shaywtiz scale were .73 and .71 respectively for kindergarten students, and .70 and .88 respectively for first grade (see http://www.pearsonclinical.com/education/products/100001918/shaywit-dyslexiascreen.html#tab-faq), which would be considered somewhat low according to screening standards in education (National Response to Intervention Center, n.d.). Moreover, if we assume that the percentage of students with dyslexia ranges from 5% (Cortiella & Horowitz, 2014) to 17% (Shaywitz, 1998), the probability of a correct decision is very low. The data in Table 1 suggest that if 100 students at each grade are identified as at-risk for dyslexia, we will likely misidentify (false positive) between 66 and 88 of kindergarteners and 46 and 77 of the first-graders, and will miss (false negative) 2 to 7 children who were actually dyslexic at each grade. Available, published measures that are already in use perform very comparably in terms of decision accuracy (National Response to Intervention Center, n.d).

The second problem with adopting a single-point-in-time measure of risk like the Shaywitz rating scale screener is that it does not inform or prompt a change in instruction that can better meet the needs of at-risk students. Some may argue that direct measures of reading proficiency for grade-level skills like Star Early Reading (Renaissance Learning, 2003), Measures of Academic Progress (Northwest Evaluation Association, 2009), or curriculum-based measures provide information to teachers about whether or not instruction is meeting the needs of students and in what specific skill areas teachers may need to provide reteaching. Yet, the extent to which teachers use the data to deliver more effective instruction is highly unstable, with some research studies reporting formative adjustments and associated achievement gains (Fuchs, Fuchs, Hamlett, & Stecker, 1991) and some studies reporting no formative adjustments and no associated achievement gains (Cordray, Pion, Brandt, Molefe, & Toby, 2012).

Selecting the "right" screener is not really the issue; using the screening effectively is. All of the focus on selecting a new screener for dyslexia is a red herring that distracts us from the real work of making sure every child has stable access to effective early reading instruction and more intensive instruction when they are struggling to learn to read.

Effective Screening Practices

So, what can be done to harness the power of academic screening and enhance the quality of life for the children that we serve, especially with regard to giving all children access to the best prevention and intervention efforts to assure reading proficiency with all the economic and social benefit that reading proficiency entails? Below we outline four recommendations for school personnel to consider to improve screening practices in their schools.

Be more selective about who is screened. One of the ways to improve screening accuracy is to screen only those students who cannot be ruled out based on other information. Use what is known about the risk of students to filter students into the "screening" and "no screening" groups. Somehow, decision makers must begin to understand the real harm that arises from screening children who have no signs of having dyslexia or a learning disability in reading. Giving a child a screening that the child does not need either confirms what we already knew (i.e., child is not at risk) or gives us bad information (i.e., as in the case of a false-positive error). Children who have shown no risk for reading failure should not be screened. Children who carry external risk factors (e.g., recently moving into a district, receiving special education services under any label, failing the preceding year's year-end test) should be screened. Furthermore, if a

child's risk of reading failure remains high (even if they pass the screening), the child should be provided with intervention.

School personnel could also use existing data (e.g., year-end tests from the preceding spring) to sort children into more intensive instructional groups for the subsequent fall before conducting screening. This approach makes use of assessment data that are already in hand, removes the delay to start intervention in the fall, and performs as well as most actual direct screenings at forecasting academic risk. Any system can easily check the associated sensitivity and specificity of the previous year's spring screening in predicting failure on the year-end state test the next year. If sensitivity exceeds the conventional standard of .80 without too high of a false-positive error rate, it might work as a fall screening for the school (Gersten et al., 2009).

Implement class-wide interventions to decrease systemic risk (and improve screening accuracy). School personnel should not ignore systemic risk. When large numbers of students are at risk for reading failure, giving a reading screening is like giving flu tests during an epidemic. No screening measure can function accurately in the case of widespread risk. When many children are at risk, systemic intervention is necessary to improve the accuracy of applied screening tools. When risk is high, treat first and measure later. Research has consistently demonstrated the benefit of class-wide interventions on student learning for reading (Mathes, Howard, Allen, & Fuchs, 1998).

Include instructional trials in screening process. It would do more harm than good to simply prescribe a single-point-in-time universal screening to identify students who may have dyslexia. It is common for single-point-in-time universal screeners to return false positive error rates of 50% or more. Including additional screening measures with highly correlated scores, administered multiple times during the year, does not improve the accuracy of these screening measures. The use of serial assessments interspersed with well-controlled doses of instruction in between are the ingredients that improve accuracy. When this process is used, identification accuracy is enhanced and the tool becomes useful for identifying dyslexia.

School personnel should also do something about the quality and intensity of the small percentage of students who are not responsive to the most intensive instructional tactics we can deliver in schools. A mandate for more identification or screening should be accompanied by the opportunity for more effective prevention and remediation. Coupling assessment with intervention effect is necessary to meet contemporary standards of assessment validity and cultural and social justice (American Educational Research Association, American Psychological Association, & National Council for Measurement in Education, 2014).

Use filtered screening for students who struggle to learn to read. Filtered screening means that a screening is administered and only children who remain in the risk range participate in more intensive instruction and the next screening to determine continued risk. Instructional trials are necessary to provide the specificity needed to rule-in students as having an instruction-resistant reading trajectory, or one that merits consideration of eligibility or diagnosis. Contemporary evidence in screening and reading failure prevention offer a converging picture that the best signal that a child may have a reading disability is the failure to learn to read in the presence of effective reading instruction. It is not possible to correctly measure the risk that signals a reading disability without measuring instruction.

Use assessment data to drive instruction. If we screen smarter, we free up assessment time and opportunity for more meaningful assessment (i.e., assessment that makes a difference and not just a prediction; Reynolds, 1975). Skill-specific assessment that is integrated with instruction and probes a child's mastery of taught skills (and if needed, prerequisite skills),

retention of learned skills, and application of learned skills to new content and understandings can be used to differentiate interventions for individual students (e.g., Burns et al., 2016), but we lose this possibility with most dyslexia screeners because they do not provide that information. We need to screen with reading, but then train teachers how to use those data to drive instruction, and we have done so in states all over the country (Burns et al., 2016).

Conclusion

Despite historic investments in preventing reading failure, large numbers of children fail to learn to read proficiently by third grade. Illiteracy is a cancerous condition for children, growing into other areas of academic and social development and greatly affecting the quality of life that a child would otherwise experience. Luckily we know a great deal about how to deliver effective reading intervention to prevent illiteracy. Generally, the barrier to preventing reading failure is not associated with lack of screening; rather, the barrier to preventing reading failure is the consistency with which we provide effective, often intensive intervention to correct and close early learning gaps. Access to effective intervention is largely controlled by the efficacy of the teaching and leadership practices in the school that the child happens to attend. These risk factors are quantifiable and are not random. For example, schools that have a higher proportion of students receiving free or reduced lunch often have lower proficiency scores in reading. Children who have been made eligible for special education in any disability category often have much lower proficiency scores and trajectories in reading compared to same-school noneligible students.

Early identification and intervention for dyslexia is an important first step, but well-intended screening actions may result in unintended negative consequences. Screening children with dyslexia screeners will likely result in inaccurate decisions in which children will still be missed, with a large number of false positive errors. Instead, schools should implement reading screeners with instructionally relevant data in combination with class-wide and individual interventions as part of the screening process. Avoiding overscreening and screening error is not about cost-savings at the expense of child benefit. Avoiding overscreening and screening error is about increasing benefit to students. We have heard it said, "Weighing a cow doesn't make it fatter." Assessment is a critical driver of student achievement, but there is a point of diminishing returns, and we have reached that point in preventing reading failure. Adding yet more reading screeners is not going to improve reading outcomes for vulnerable children.

The dyslexia grass-roots movement presents a timely opportunity for our schools and the children that we serve, but it is an opportunity that cannot be squandered by selecting the option that is easy but wrong for children. Identifying and remediating dyslexia is yet another example in which the best option is the one that requires the most work, but we owe it to the children for whom reading is a labor instead of a joy.

References

American Educational Research Association, American Psychological Association, & National Council for Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Burns, M. K., Maki, E. E., Karich, A. C., Hall, M., McComas, J. J., & Helman, L. (2016). Problem-analysis at tier 2: Using data to find the category of the problem. In S. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (2nd ed., pp. 293–308). New York, NY: Springer.

Cordray, D., Pion, G., Brandt, C., Molefe, A, & Toby, M. (2012). *The impact of the Measures of Academic Progress (MAP) program on student reading achievement.* (NCEE 2013–4000). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Cortiella, C., & Horowitz, S. H. (2014). *The state of learning disabilities: Facts, trends and emerging issues.* New York, NY: National Center for Learning Disabilities.

Eide, F. (2016, March 29). Progress! Passed dyslexia laws in the United States – 2016. *Dyslexic Advantage.* Available at http://www.dyslexicadvantage.org/progress-passed-dyslexia-laws-in-the-united-states-2016/.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1991). Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Educational Research Journal, 28*, 617–641.

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., & Witzel, B. (2009). *Assisting students struggling with mathematics: Response to Intervention (RtI) for elementary and middle schools* (NCEE 2009-4060). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Hayes, S. C., Nelson, R. O., & Jarrett, R. B. (1987). The treatment utility of assessment: A functional approach to evaluating assessment quality. *American Psychologist, 42*, 963–974.

Ikeda, M. J., Neesen, E., & Witt, J. C. (2008). Best practices in universal screening. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 103–115). Bethesda, MD: National Association of School Psychologists.

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.

Mathes, P. G., Howard, J. K., Allen, S. H., & Fuchs, D. (1998). Peer assisted learning strategies for first grade readers: Responding to the needs of diverse learners. *Reading Research Quarterly, 62*, 62–94.

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin, 52*, 194–215. doi:10.1037/h0048070

National Response to Intervention Center. (n.d.). *Screening tools chart.* Washington, DC: American Institutes for Research. Retrieved from http://www.rti4success.org/resources/tools-charts/screening-tools-chart

Northwest Evaluation Association. (2009). *Technical manual for Measures of Academic Progress™ and Measures of Academic Progress for primary grades™*. Lake Oswego, OR: Author.

Pace, L. E., & Keating, N. L. (2014). A systematic assessment of benefits and risks to guide breast cancer screening decisions. *JAMA*, *311*, 1327–1335.

Renaissance Learning. (2003). *STAR early literacy*. Wisconsin Rapids, WI: Author.

Reynolds, M. C. (1975). Trends in special education: Implications for measurement. In W. Hively & M. C. Reynolds (Eds.), *Domain-referenced testing in special education*. Minneapolis, MN: University of Minnesota Leadership Training Institute/Special Education.

Shaywitz, S. E. (1998). Dyslexia. *New England Journal of Medicine*, *338*, 307–312.

Shaywitz, S. E. (2016). *Shaywitz dyslexia screen.* Bloomington, MN: Pearson Education.

VanDerHeyden, A. M. (2013). Universal screening may not be for everyone: Using a threshold model as a smarter way to determine risk. *School Psychology Review*, *42*, 402–414.

Amanda M. VanDerHeyden, Ph.D., is a researcher and consultant in Fairhope, Alabama. She is the founder of SpringMath.

Education Research and Consulting

Matthew K. Burns, Ph.D. is the Associate Dean for Research and Professor of Educational, School, and Counseling Psychology at the University of Missouri.

Table 1. ***Probability of an Accurate Test Finding for the Shaywitz Dyslexia Screen Given Estimates of Sensitivity, Specificity, and Prevalence***

Grade	% Population with Dyslexia	Probability Positive Test (Dyslexia) is Correct	Probability Negative Test (No Dyslexia) is Correct
Kindergarten	5%	.12	.02
	17%	.34	.07
First Grade	5%	.23	.02
	17%	.54	.07