September 2016

Applied Research Methods

Predicting math outcomes from a reading screening assessment in grades 3–8

Adrea J. Truckenmiller
Yaacov Petscher
Florida Center for Reading Research at Florida State University

Linda Gaughan
Ted Dwyer
Hillsborough County School District

Key findings

This study finds that a reading screening assessment used to identify students who may be at risk of low reading achievement can predict end-of-year math outcomes with a level of accuracy similar to that of math screening assessments. Therefore school districts could use an assessment of reading skills to screen for risk in both reading and math at the same time, potentially reducing costs and testing time. Furthermore, the analyses in this study produced decision trees that may offer practitioners a more transparent link between screening and outcomes than does logistic regression, another commonly used method for determining screening accuracy.

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences
U.S. Department of Education

REL SOUTHEAST
Regional Educational Laboratory
At Florida State University

U.S. Department of Education
John B. King, Jr., *Secretary*

Institute of Education Sciences
Ruth Neild, *Deputy Director for Policy and Research*
Delegated Duties of the Director

National Center for Education Evaluation and Regional Assistance
Joy Lesnick, *Acting Commissioner*
Amy Johnson, *Action Editor*
Sandra Garcia, *Project Officer*

REL 2016–180

The National Center for Education Evaluation and Regional Assistance (NCEE) conducts unbiased large-scale evaluations of education programs and practices supported by federal funds; provides research-based technical assistance to educators and policymakers; and supports the synthesis and the widespread dissemination of the results of research and evaluation throughout the United States.

September 2016

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-IES-12-C-0011 by Regional Educational Laboratory (REL) Southeast administered by Florida State University. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Truckenmiller, A. J., Petscher, Y., Gaughan, L., & Dwyer, T. (2016). *Predicting math outcomes from a reading screening assessment in grades 3–8* (REL 2016–180). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from http://ies.ed.gov/ncee/edlabs.

This report is available on the Regional Educational Laboratory website at http://ies.ed.gov/ncee/edlabs.

Summary

District and state education leaders frequently use screening assessments to identify students who are at risk of performing poorly on end-of-year achievement tests. This study examines the use of a universal screening assessment of reading skills for early identification of students at risk of low achievement on nationally normed tests of reading and math and provides support for the interpretation of screening scores to inform instruction.

Several members of the Regional Educational Laboratory Southeast Improving Literacy Alliance already use a reading screening assessment—the Florida Center for Reading Research Reading Assessment (FRA)—for all students in grades 3–8 to identify students who may be at risk of poor end-of-year reading outcomes. To gain more information to drive instruction without students having to spend more time taking tests, these alliance members wanted to know whether the FRA could also be used to identify students at risk of poor end-of-year math outcomes. Data on students in grades 3–8 from one large Florida school district were available to answer these questions.

The study found that the FRA identified students at risk of poor performance in mathematics on the Stanford Achievement Test, Tenth Edition, with a level of accuracy similar to that of screening assessments that measure math skills. The findings indicate that school districts could use an assessment of reading skills to screen for risk in both reading and math at the same time, potentially reducing costs and testing time.

This report provides decision trees to support implementation of screening practices and interpretation by teachers.

Contents

Tables

Why this study?

Every state tests students at the end of the school year in at least two academic subjects: reading (or English language arts) and math. Because state tests have high stakes—such as grade retention or promotion, or teacher performance evaluation—many educators try to identify which students are at risk of failing end-of-year tests as early as possible in order to change those students' trajectory. A variety of screening assessments (including AIMSweb, Measures of Academic Progress, and STAR assessments) are used to identify students at risk of failing (National Center on Response to Intervention, 2012). Many of these assessments also have diagnostic features that allow educators to identify specific skill weaknesses that can guide differentiated instruction. Educators use these data to set goals and improve reading and math instruction (Gersten et al., 2008; Jenkins, 2003).

Just as doctors use a thermometer to screen for a variety of illnesses, a reading screening assessment can be used to identify students at risk of failure in a variety of academic areas

As of 2010, 46 states recommended or required that schools conduct universal screening—that is, administer screening assessments to all students at the beginning of the school year to identify which students may need supplemental or differentiated instruction to meet end-of-year expectations (Zirkel & Thomas, 2010). Although some schools use screening assessments for several subjects and student behaviors, most schools that employ universal screening past grade 3 use only a reading screening assessment. The majority of research on universal screening has been conducted on reading screening and has shown positive results in predicting reading achievement (Wayman, Wallace, Wiley, Tichá, & Espin, 2007). But because the stakes have been raised in other subjects (for example, requiring passing scores on math tests in order to graduate from high school), educators may want to use screening assessments that accurately identify students for more intensive instruction in those subjects as well (Crawford, Tindal, & Steiber, 2001). At the same time, educators and education leaders must weigh the benefits of screening students for outcomes in addition to reading (Gersten et al., 2008) against the potential costs in money and instructional time of multiple screening assessments.

Because reading and math difficulties often occur together and may have similar underlying causes (Crawford et al., 2001; Fletcher, 2005; Helwig, Rozek-Tedesco, Heath, & Tindal, 1999; Thurber, Shinn, & Smolkowski, 2002), screening assessments in reading and math may identify many of the same students. So using an existing reading screening assessment to identify students at risk of poor outcomes in other subjects, such as math, could increase efficiency.

Two studies suggest that reading screening assessments can be used in addition to math screening assessments to improve prediction of math outcomes. They found that using a reading screening assessment in addition to a math screening assessment significantly improved the prediction accuracy of math outcomes and that the reading screening assessment was critical to identifying at-risk students in grades 3, 5, and 7 (Codding, Petscher, & Truckenmiller, 2014; Jiban & Deno, 2007).

Just as doctors use a thermometer to screen for a variety of illnesses, a reading screening assessment can be used to identify students at risk of failure in a variety of academic areas. Educators report being very confused about how performance on a screening measure directly relates to performance on an outcome measure, especially when students have similar scores in one area of reading but have different levels of risk on the whole screening assessment. Since the typical analysis of screening prediction (logistic regression) uses an unseen formula to predict outcomes, educators frequently report that they distrust the prediction. Previous research shows the theoretical link between reading screening assessments and math outcomes, but to inform educators' practices, models other than logistic regression are needed to delineate which scores on a specific reading screening assessment directly relate to scores on a math outcome (Koon & Petscher, 2015; Koon, Petscher, & Foorman, 2014).

This study was requested by a Florida-based member of the Regional Educational Laboratory (REL) Southeast's Improving Literacy Alliance. In 2002 Florida school districts began using universal reading screening assessments in kindergarten through grade 3 to predict important reading outcomes and in 2007 expanded the practice to all grades. Currently, the Florida Center for Reading Research Reading Assessment (FRA) is available in Florida for use as a free reading screening assessment for grades 3–10. The primary purpose of this study was to determine the screening accuracy of the FRA for an additional important academic outcome: the Stanford Achievement Test, Tenth Edition (SAT-10) Mathematics. A secondary goal of this study was to use an analysis that would produce a decision tree for interpreting FRA scores instead of the probability score that is currently produced for the FRA.

The primary purpose of this study was to determine the screening accuracy of the Florida Center for Reading Research Reading Assessment (FRA) for the SAT-10 Mathematics. A secondary goal was to use an analysis that would produce a decision tree for interpreting FRA scores instead of the probability score that is currently produced for the FRA

The FRA produces an overall probability score that indicates the likelihood of a student passing the end-of-year reading outcome test based on a complex formula of that student's performance on FRA tasks. The alliance member reports that some practitioners are hesitant to use FRA scores to inform instructional practices because probability scores do not demonstrate a direct connection between the component reading skills and important outcomes. In addition to answering the question about reading screening predicting math outcomes, the alliance member is also seeking a way to make the relationship between each FRA task score and outcomes in math and reading more transparent and less difficult to interpret. Given that the reading screening scores would be used to predict reading and math outcomes, this report compares the differences in identifying students at risk in math and students at risk in reading.

What the study examined

Three research questions guided the study:

- How are FRA task scores associated with SAT-10 Reading Comprehension and Mathematics scores?
- How well does a universal screening assessment of reading performance identify students at risk of not meeting expectations in math?
- Are the reading skills that predict math outcomes similar to the skills that predict reading outcomes?

The study was conducted using data on students in grades 3–8 in a large urban school district in Florida (see box 1 for a summary of the data and methods used in the study and appendix A for more details).

Box 1. Data and methods

Participants

Data were analyzed for all 7,803 students in grades 3–8 attending 15 elementary schools and 5 middle schools in a large urban district in Florida who took the Florida Center for Reading Research Reading Assessment (FRA) midyear and the Stanford Achievement Test, Tenth Edition (SAT-10), at the end of the school year. (In this school district, every student takes the FRA and the SAT-10 Reading Comprehension and Mathematics, unless the student has an exemption.) The participants were demographically diverse: 58 percent received free or reduced-price lunch, 10 percent were English learner students, 12 percent were in special education, 3 percent were Asian, 18 percent were Black, 39 percent were Hispanic, 1 percent were American Indian, and 37 percent were White. The students in grades 3–5 were high-achieving in both reading and math (see table B1 in appendix B); the students in grades 6–8 demonstrated average achievement in reading and math (see table B2 in appendix B).

Measures of student achievement

The study used two measures of student achievement.

Florida Center for Reading Research Reading Assessment (FRA). The FRA (Foorman, Petscher, & Schatschneider, 2015) comprises four computer adaptive tasks,[1] each producing a score, for a total of four scores per student:

- Word recognition: Measures a student's ability to decode words and requires the student to listen to a word pronounced by the computer and choose the correctly spelled word from three choices. Items include real words and nonwords.
- Vocabulary knowledge: Measures a student's vocabulary ability. The items consist of a sentence with one word missing. The missing word is replaced with a choice of three morphologically related words.
- Syntactic knowledge: Measures a student's comprehension of components in a sentence by having the student select the correct connective word, the correct pronoun reference, or the verb that creates appropriate subject–verb agreement. This task includes an audio assist for all students.
- Reading comprehension: Consists of a passage with seven to nine multiple-choice questions.

Stanford Achievement Test, Tenth Edition (SAT-10). The SAT-10 is a nationally normed test for grades 1–12 and adults and comprises eight subjects (Harcourt, 2003). The current study uses the Reading Comprehension score and the Mathematics score. The district participating in the current study found that the 50th percentile of SAT-10 Reading Comprehension and Mathematics scores most closely aligned with the passing score on its state-mandated accountability test in reading and math, so the scaled score associated with the 50th percentile for each grade level was used as the cutpoint for meeting grade-level expectations (passing) in each subject.

Analyses

This study uses classification and regression tree (CART) models to determine how well FRA task scores identify students at risk of failing the SAT-10 Reading Comprehension and Mathematics. CART helps identify students who are at risk based on multiple scores with cutpoints for each score and clear decision rules for combining the results. CART models yield decision trees based on these cutpoints to show which students are at risk and which students are not at risk. CART models can be adjusted to balance several types of screening accuracy (see box 2).

Note

1. Computer adaptive administration means that not all students are administered the same items. The difficulty of the items administered to a student depends on the student's previous responses. If a student is responding correctly to items, more difficult items are administered. If a student is responding incorrectly to items, easier items are administered.

Box 2. Understanding screening accuracy

The screening accuracy produced from the classification and regression tree model is represented by several statistics. The definition for each screening accuracy statistic is listed below. More detailed descriptions of these statistics are available in Schatschneider, Petscher, and Williams (2008).

Sensitivity. Percentage of students identified as at risk by the Florida Center for Reading Research Reading Assessment (FRA) out of all students who failed the SAT-10.

Specificity. Percentage of students identified as not at risk by the FRA out of all students who passed the SAT-10.

False positive rate. Percentage of students who were identified as at risk by the FRA but passed the SAT-10.

False negative rate. Percentage of students who were not identified as at risk by the FRA but failed the SAT-10.

Positive predictive power. Percentage of students correctly identified as at risk by the FRA out of all students identified as at risk.

Negative predictive power. Percentage of students correctly identified as not at risk by the FRA out of all students identified as not at risk.

Overall accuracy rate. Percentage of students correctly identified as at risk or not at risk by the FRA out of all students in the sample.

Educators who choose screening assessments may wish to prioritize some accuracy statistics over others. For example, the Regional Educational Laboratory Southeast's Improving Literacy Alliance members want to identify as many students as possible who may be at risk (that is, prioritizing negative predictive power) even though this approach may increase the number of false positives. Many commercial screening measures also prioritize negative predictive power (National Center on Response to Intervention, 2012). Models that determine the cutpoints on the screening assessment can be adjusted to maximize other screening accuracy statistics that may be needed for different purposes.

Because adjustments to improve one accuracy statistic necessarily affect other such statistics, educators must clearly specify the decisions they plan to make from the screening assessment data and set cutpoints appropriately. For example, if the identification of a student as at risk of not meeting expectations was used in high-stakes decisions (such as grade retention), educators would choose an assessment and cutpoint with more of a focus on reducing the rate of false positives. Conversely, if a school wanted to identify students who need further supplemental instruction and the school has enough resources to meet those demands, it may not be concerned about the resource cost associated with trying to increase negative predictive power. Cost is not the only consideration, nor the most important. Decisionmakers will want to carefully consider all implications of over- and underidentification at their school, including the student's placement in the core curriculum and other typical school activities as well as the process for determining when a student may successfully exit an intervention or intervention setting.

After deciding whether to prioritize underidentification or overidentification, practitioners can draw guidance for minimum accuracy thresholds from research or compare choices among other available screening assessments. Some researchers suggest that values above .75 for sensitivity, specificity, and predictive power provide evidence for good classification (Swets, 1992). Others recommend that sensitivity values be above .90 (Compton, Fuchs, Fuchs, & Bryant, 2006; Jenkins, 2003) and negative predictive power be above .80 (Petscher, Kim, & Foorman, 2011). The current study sought to keep negative predictive power above .80 while keeping specificity near .75.

What the study found

This section describes the results of each step of the analysis.

Reading skills are moderately to strongly correlated with math skills

Scores for each FRA task were significantly correlated with SAT-10 Mathematics scores at all grade levels (see table B3 in appendix B). Altogether, the FRA task score explained almost 50 percent of the variance in the SAT-10 Mathematics score (r^2 = 46).

- The FRA reading comprehension score had the strongest association (average correlation = .64) with the SAT-10 Mathematics score.
- The FRA syntactic knowledge score also had a strong association (average correlation = .51) with the SAT-10 Mathematics score in most grade levels, ranging from .43 in grade 3 to .56 in grade 8.
- The FRA vocabulary knowledge score (average correlation = .43) and word recognition score (average correlation = .42) had a moderate to strong correlation with the SAT-10 Mathematics score.

Scores for each FRA task were significantly correlated with SAT-10 Mathematics scores at all grade levels. Altogether, the FRA task score explained almost 50 percent of the variance in the SAT-10 Mathematics score

The correlations between FRA task scores and SAT-10 Reading Comprehension scores showed a similar pattern, though the correlations were generally larger. For example, the FRA reading comprehension score was most strongly correlated with the SAT-10 Reading Comprehension score (average correlation = .70), followed by syntactic knowledge score (average correlation = .51).

Reading screening accurately predicts math outcomes

After determining that the screening assessment was related to the outcome test, the study team analyzed how well the screening assessments predicted the outcomes by calculating screening accuracy statistics output from classification and regression tree (CART) models. CART models were used to find the minimum scores (cutpoints) that students needed on the combination of FRA tasks that most accurately predicted performance on the SAT-10 Mathematics. The development of CART models involves multiple iterations, and explaining each step and the rationale for each decision can be challenging. The steps for these models are discussed in appendix A. Although some screening accuracy statistics did not reach all the thresholds recommended by researchers and the screening accuracy statistics varied across grade levels, the FRA generally performed well. Sensitivity exceeded .80 in all grade levels, and negative predictive power exceeded .80 in all grade levels except grade 7 (table 1). Specificity, false positive rate, and negative predictive power were noticeably poorer for grade 7 than for other grade levels.

The results for the FRA are comparable to those for seven other commercial math screening assessments in sensitivity, specificity, negative predictive power, and overall accuracy rate (table 2; see also table C1 in appendix C). The accuracy of the FRA is similar to or slightly higher than that of three of these commercial math screening assessments in all screening accuracy categories.

As in other evaluations of the FRA's screening accuracy (Foorman et al., 2015), the statistics on how well FRA task scores identify students at risk of scoring below the 50th percentile on the SAT-10 Reading Comprehension indicated good screening accuracy (see

Table 1. Screening accuracy statistics on how well FRA task scores identify students at risk of scoring below the 50th percentile of Stanford Achievement Test, Tenth Edition, Mathematics scores

Grade	Sensitivity	Specificity	False positive rate	False negative rate	Positive predictive power	Negative predictive power	Overall accuracy rate
3	.84	.74	.26	.16	.66	.89	.78
4	.86	.74	.25	.14	.69	.89	.79
5	.84	.73	.25	.14	.69	.89	.79
6	.84	.73	.27	.16	.58	.91	.76
7	.84	.68	.32	.16	.77	.77	.77
8	.86	.80	.20	.14	.82	.84	.83

FRA is Florida Center for Reading Research Reading Assessment.

Note: See box 2 for definitions of screening accuracy statistics.

Source: Authors' analysis of school district data for 2012/13.

Table 2. How screening accuracy statistics of the FRA compare with those of selected commercial math screening assessments

Assessment	Sensitivity	Specificity	False positive	False negative	Positive predictive power	Negative predictive power	Overall accuracy rate
Acuity	↑	↓	↓	↑	↑	↓	↓
AIMSweb	↑	↔	↔	↔	↑	↓	↔
Discovery Education Predictive Assessment	↔	↔	↔	↔	↔	↔	↔
EasyCBM	↔	↔	↔	↔	↔	↔	↔
Measures of Academic Progress	↑	↓	↓	↑	↔	↓	↓
Iowa Test of Basic Skills	↑	↓	↓	↑	↔	↔	↔
STAR	↑	↔	↔	↑	↑	↔	↑

FRA is Florida Center for Reading Research Reading Assessment.

↑ indicates that the Florida Center for Reading Research Reading Assessment (FRA) predicts math outcomes slightly better than the commercial screening assessment.

↔ indicates that the FRA predicts math outcomes approximately as well as the commercial screening assessment.

↓ indicates that the commercial screening assessment predicts math outcomes slightly better than the FRA.

Note: The range of FRA statistics for grades 3–8 was compared to the range of statistics that each commercial assessment reported.

Source: Authors' analysis based on National Center on Response to Intervention (2012).

table C3 in appendix C). Almost all the sensitivity, specificity, and negative predictive power statistics for predicting SAT-10 Reading Comprehension were at or above .80. In this way the screening accuracy statistics for the FRA were considered better for predicting the SAT-10 Reading Comprehension than for predicting the SAT-10 Mathematics.

Interpreting screening task scores using a classification and regression tree

A member of the REL Southeast's Improving Literacy Alliance requested a more transparent explanation of how a screening measure relates to an outcome measure. One commonly used analysis for determining screening accuracy (logistic regression) produces a formula that differentially weights components of the screening assessment and produces a probability that a student will be at risk or not at risk. By contrast, classification and regression tree (CART) analyses identify the scores on a screening assessment that serve as the cutpoint between students who are considered at risk and not at risk.

The alliance member reports that the educators using the data frequently inquire about the direct relationship between scores on the screening assessment and risk status. Educators are confused by the weighting process that is used in and the probabilities that come from logistical regression models. Educators who use screening assessments to inform instruction may find CART analyses easier to interpret than the currently used probability-of-success score (though the interpretability of CART analyses can vary depending on the number of decision rules used; see appendix D). This is because CART analyses specify the cutpoints on the screening assessment that directly relate to the cutpoint on the outcome test.

Educators who use screening assessments to inform instruction may find classification and regression tree (CART) analyses easier to interpret than the currently used probability-of-success score because CART analyses specify the cutpoints on the screening assessment that directly relate to the cutpoint on the outcome test

The criteria for identifying students as at risk or not at risk can be displayed to practitioners in two ways: a list of criteria or a decision tree. For example, there are three criteria for identifying grade 4 students as at risk and two criteria for identifying them as not at risk (figure 1; see appendix E for decision rules for other grades).

A student in grade 4 is likely at risk if his or her:

- Reading comprehension score is between 380 and 437, syntactic knowledge score is 400 or higher, and word recognition score is less than 411 (31 percent of students).
- Reading comprehension score is between 380 and 437 and syntactic knowledge score is less 400 (12 percent of students).
- Reading comprehension score is less than 380 (3 percent of students).

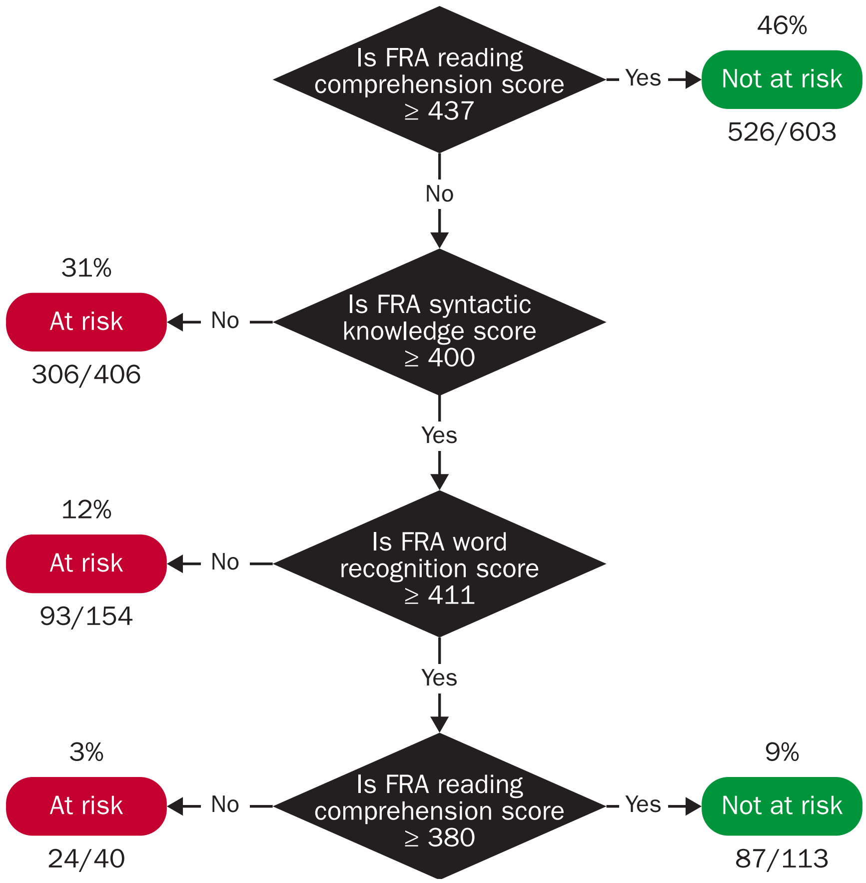
A student in grade 4 is likely not at risk if his or her:

- Reading comprehension score is 437 or higher (46 percent of students).
- Reading comprehension score is 380 or higher, syntactic knowledge score is 400 or higher, and word recognition score is 411 or higher (9 percent of students).

To use the decision tree in figure 1, an educator would compare a student's FRA scores to the cutpoints identified in the CART model for that student's grade level. Each diamond represents a decision cutpoint.

- The first cutpoint identifies a group of students who were classified as not at risk. Forty-six percent of the sample (603 students) were identified as not at risk by having an FRA reading comprehension score of 437 or higher. In this category 526 students (87 percent) were correctly identified as having an SAT-10 Mathematics score at or above the 50th percentile.
- The second cutpoint was for students with an FRA reading comprehension score below 437. If their FRA syntactic knowledge score was below 400, they were classified as at risk (31 percent of the sample). In this category 306 students (75 percent) were correctly identified as not having an SAT-10 Mathematics score at or above the 50th percentile.

Figure 1. Decision tree for identifying students at risk of scoring below the 50th percentile on the Stanford Achievement Test, Tenth Edition, Mathematics



FRA is Florida Center for Reading Research Reading Assessment.

Note: Diamonds represent decision rules, and ovals represent categories of students identified as at risk or not at risk. The denominator of the fraction is the number of students from the sample who were classified in that category, and the numerator is the number of students in that category who were correctly identified. The percentage identifies the proportion of the sample that fell into that category. Percentages do not sum to 100 because of rounding.

Source: Authors' analysis of school district data for 2012/13.

A student in grade 4 is likely at risk of scoring below the 50th percentile on the SAT-10 Mathematics if his or her reading comprehension score is between 380 and 437, syntactic knowledge score is 400 or higher, and word recognition score is less than 411; reading comprehension score is between 380 and 437 and syntactic knowledge score is less 400; or reading comprehension score is less than 380

- The third cutpoint was for students with an FRA syntactic knowledge score of 400 or higher. If their FRA word recognition score was below 411, they were classified as at risk (12 percent of the sample). In this category 93 students (60 percent) were correctly classified.
- The fourth cutpoint was for students with an FRA word recognition score above 411. If their FRA reading comprehension score was below 380, they were classified as at risk (3 percent of the sample). In this category 24 students (60 percent) were correctly classified.
- The remaining 9 percent of students were classified as not at risk. In this category 24 students (77 percent) were correctly classified.

The percentile ranks associated with these scores on the component skills of reading may help put these classification profiles into perspective. But the percentile ranks on the FRA tasks are not expected to line up perfectly with percentile ranks on the SAT-10 because of differences in norming samples and because of the nature of development of component reading skills. For grade 4 students, an FRA reading comprehension score of 437 is associated with the 50th percentile of the FRA normative group, an FRA syntactic knowledge score of 400 is associated with the 38th percentile, an FRA word recognition score of 411 is associated with the 50th percentile, and an FRA reading comprehension score of 380 is associated with the 20th percentile. Thus, the two groups of students not at risk have at least an average FRA reading comprehension score (above the 50th percentile) or an FRA reading comprehension score at least at the 20th percentile and no concurrent weaknesses in the FRA syntactic knowledge task (scoring above the 38th percentile) and the FRA word recognition task (scoring above the 50th percentile). These grade 4 students are likely to meet expectations for the SAT-10 Mathematics.

Profiles of risk in math differ slightly from profiles of risk in reading

Cutpoints on the FRA screening assessment differ depending on student grade level and whether the outcome is math or reading (see appendix E for the decision tree outputs from the CART models).

When identifying students at risk of scoring below the 50th percentile on the SAT-10 Mathematics, the score for just one FRA task (reading comprehension) serves as the best predictor in grades 5, 6, and 7. The decision trees show a simple one-step relationship between the FRA reading comprehension score and the SAT-10 Mathematics score (see figures E3–E5 in appendix E). In grade 8 the FRA syntactic knowledge score is most likely to differentiate students at risk of scoring below the 50th percentile on the SAT-10 Mathematics (see figure E6 in appendix E).

A variety of component reading and language skills are important for predicting math outcomes in the elementary years, and a relative strength in word recognition or syntax may serve a compensatory function for students with lower reading comprehension abilities

As with grade 4, identifying students at risk of scoring below the 50th percentile on the SAT-10 Mathematics in grade 3 involves scores for multiple reading skills. Three score profiles result in a classification of not at risk (see figure E1 in appendix E). Students are classified as not at risk if they have an FRA reading comprehension score above 400 or an FRA reading comprehension score from 330 to 400 and an FRA word recognition score above 370. Students are also classified as not at risk if their FRA reading comprehension score is between 360 and 400 and their FRA syntactic knowledge score is above 320. This shows that a variety of component reading and language skills are important for predicting math outcomes in the elementary years and that a relative strength in word recognition or syntax may serve a compensatory function for students with lower reading comprehension abilities.

The relationship between reading screening assessment and reading outcomes is more direct than the relationship between reading screening assessment and math outcomes. When identifying students at risk of scoring below the 50th percentile on the SAT-10 Reading Comprehension, one cutpoint—based on the FRA reading comprehension score—serves as the best predictor in grades 3–7. The models demonstrate a simple direct relationship between the FRA reading comprehension score and the SAT-10 Reading Comprehension score (see figures E7–E11 in appendix E). In grade 8 the profiles for students at risk and not at risk are more complex (see figure E12 in appendix E). Grade 8 students with an FRA reading comprehension score below 534 are at risk of receiving an SAT-10 Reading Comprehension score below the 50th percentile. Students with an FRA reading comprehension score above 534 and an FRA syntactic knowledge score above 541 are identified as not at risk of receiving an SAT-10 Reading Comprehension score below the 50th percentile. If their FRA reading comprehension score is above 534 but their FRA syntactic knowledge score is below 541, they need an FRA word recognition score above 507 to be categorized as not at risk.

Decision trees vary in both the identification of cutpoints and complexity. Figure 1 shows a relatively simple decision tree with four decision rules. Frequently, the most accurate CART model in this sample produced an easy-to-interpret decision tree with four or fewer decision rules. CART models are designed to balance the need to maintain reasonable values for the seven accuracy statistics and can sometimes produce a decision tree that has 10 or more decision rules (see tables D1 and D3 in appendix D), which reduces interpretability.

Although complex, educators may still implement models with multiple decision rules because most computer-delivered screening assessments can automate the algorithm to sort students into at risk and not at risk. Thus, as educators select their models for each grade and subject, they will have to consider the tradeoffs between simplicity and the balance of the accuracy statistics. For example, in grade 3 the model that optimized each accuracy statistic also had an uninterpretable number of decision rules (21; see table D1 in appendix D). When the CART model for grade 3 was restricted to produce five decision rules, the overall correct classification rate, negative and positive predictive power, and sensitivity statistics all declined, and false negatives increased (see table D2 in appendix D). These tradeoffs in classification accuracy for an improved number of decision rules were relatively small. For example, in grade 8 the statistics based on false positives (specificity, false positives, and positive predictive power) were approximately 10 percentage points worse, but negative predictive power (the statistic that was prioritized) changed little (from .84 to .82).

Members of the REL Southeast Improving Literacy Alliance who value the transparency of scores may want to accept slightly lower screening accuracy from models with fewer decision rules.

Implications of the study findings

Findings from this study suggest several implications for the relationship between reading screening and math outcomes:

- Reading screening assessments can be useful in identifying students who may be at risk of poor math outcomes.
- The accuracy of reading screening assessments in predicting math outcomes in this school district is similar to that of other commercial math screening assessments.
- The accuracy of the reading screening assessments in predicting math outcomes in this school district is similar to its accuracy in predicting reading outcomes.

The results suggest that students' ability in different reading component skills are important not only for reading and language arts teachers, but also for math teachers to understand how to best help students achieve important outcomes

In short, this school district, which is already screening its students for risk in reading can use the FRA to identify risk in math and thereby reduce the amount of time spent on universal screening.

The results suggest that students' ability in different reading component skills (such as text comprehension, sentence comprehension, vocabulary, and word recognition) are important not only for reading and language arts teachers, but also for math teachers to understand how to best help students achieve important outcomes. By providing actionable information to both reading and math teachers and by integrating literacy in math, schools may be able to improve overall instruction.

Integrating literacy instruction with solid content area instruction is recommended for secondary schools (Biancarosa & Snow, 2006). The same research suggests that integrating explicit literacy instruction when reading subject-specific texts results in greater improvements in students' content knowledge and comprehension. Content area teachers (such as math teachers) can also use information about students' literacy skills to identify situations where students may need more support.

For example, a school leader could meet with grade 6 language arts and math teachers to analyze the comprehension and language skills needed for success in reading comprehension (see figure E10 in appendix E) and math outcomes (see figure E4 in appendix E). The language arts teachers might already provide differentiated support to students who are at risk in reading comprehension (scoring below 487 on the FRA reading comprehension task). The language arts teachers can work with math teachers to understand the type of supplemental or differentiated support the students at risk of poor reading outcomes and the students at risk of poor math outcomes need when working with math text and in understanding verbal interactions (such as class lectures). The use of similar text and language strategies for all texts (not just texts encountered in language arts class) could improve students' academic skills, reinforce the reading skills required for both language arts and math classes, and ultimately improve students' ability to demonstrate on outcome tests their understanding of the content they need to master.

Schools might consider conducting a cost-benefit analysis of administering a math screening assessment in addition to the reading screening assessment. Previous research demonstrates that the best prediction of math outcomes is reached when a reading and math screening assessment are used together

Schools might also consider conducting a cost-benefit analysis of administering a math screening assessment in addition to the reading screening assessment. Previous research demonstrates that the best prediction of math outcomes is reached when a reading and math screening assessment are used together (Codding et al., 2014; Jiban & Deno, 2007). Regardless of which assessments are administered, teachers need more information specific to a student's math skills to guide math instruction once that student is identified as at risk. This study provides logistically appealing possibilities—one type of screening assessment that is already widely administered (a reading screening assessment) could also be used to identify students at risk in math. However, there are important limitations to the generalizability of these outcomes as well as considerations for the application of this conclusion.

Limitations of the study

Previous studies have found that scores on a reading screening assessment are more strongly related to math outcomes than scores on a math screening are (Codding et al., 2014; Jiban & Deno, 2007), but the current study cannot make the same claim. The previous studies used different types of screening assessments (curriculum-based measures) for both reading and math and administered both reading and math screening assessments at the same time. The current study used a different type of screening assessment and did not include a math screening assessment. Since the current study does not directly compare a math screening assessment with the FRA in the same sample, another math assessment may provide a better prediction of math outcomes and provide unique information regarding math achievement. A math screening assessment may also accurately identify students at risk in math who are not identified by the reading screening assessment. The lack of math-specific information in the current study also limits the utility of the findings for math instruction. The results do not suggest that math teachers should focus on reading skills to the detriment of the quality, depth, breadth, or amount of time spent in math instruction. Further diagnostic assessment or curriculum-embedded assessment of math is necessary to guide math instruction for students who are identified as at risk.

Universal screening is conceptualized in this study as a process by which schools identify students who may need differentiated instruction, supplemental instruction, or more individualized feedback within the core curriculum. The results of this study are targeted toward those types of instructional decisions and are not intended to support high-stakes decisions such as special education classification, retention, or any other removal from the general education curriculum. This study does not empirically evaluate the practical implications and potential misunderstanding of screening data by practitioners.

In addition, the models in this study may not generalize to other schools. The degree of screening accuracy and the cutpoints in CART models vary depending on the percentage of students scoring at or above the 50th percentile on the SAT-10 or the sample's demographics (Schatschneider et al., 2008). Other unmeasured factors may affect the variation in cutpoints. A school or school district that uses a reading screening assessment with decision rules (CART models) to identify students at risk in other subjects should tailor the models to the school or school district population and not use the cutpoints identified in this report. This study provides a precedent for exploring the predictive accuracy of other screening assessments predicting cutpoints on other important outcomes.

The results do not suggest that math teachers should focus on reading skills to the detriment of math instruction. Further diagnostic assessment or curriculum-embedded assessment of math is necessary to guide math instruction for students who are identified as at risk

If this study is used for educator training purposes, it is important for trainers to clearly define the purpose of the screening assessment and emphasize that using a reading screening assessment as a general academic gauge is not intended to reduce content-directed instruction or replace content-specific diagnostic measures. Students identified as at risk on the math outcome require diagnostic math screening assessments to help guide supplemental or intensive instruction and intervention.

Finally, the CART figures in this report provide another option for addressing a perceived lack of transparency in assessment results. Though CART analyses may be a more user-friendly approach, no empirical evaluation has been conducted that compares CART analyses with other methods (such as logistic regression) for interpretability by educators.

Appendix A. Study methodology

The study team used classification and regression tree (CART) analyses to evaluate the research questions because they produce a decision tree that explicitly illustrates the relationship between screening scores and outcomes. They also yield statistics on the accuracy of screening assessments that show the percentages of students correctly and incorrectly identified as at risk based on the particular specification of the CART model.

Before conducting the CART analyses, the dataset (see box 1 in the main report) was analyzed for missing data to reduce or eliminate potential biases. Significant results from Little's missing completely at random test (p = .000) indicated that missing data were not missing completely at random, but the study team determined that the missing data could be assumed to be missing at random because the reason for missing data (absences) was unrelated to performance on the measures used in the study. To address the missing data, SAS 9.4 software multiple imputation was used to create a dataset with complete cases for all variables. Currently, the research literature offers no recommended procedure for analyzing and summarizing classification trees generated from multiple imputed files. So a conservative 20,000 imputations were conducted, and the mean imputed value was used for each missing value.

CART analyses classify individuals into mutually exclusive subgroups of a population using a nonparametric approach that results in a decision tree. The subgroup splits in CART analyses are determined by a software program (R software, rpart package; Therneau & Atkinson, 2013) that is set to reduce the model relative error and simultaneously improve model fit. CART analyses use an exhaustive subgroup comparison to identify the predictors (tasks) and predictor levels (scores on the tasks) that best split the sample into the most homogeneous subgroups of students identified as at risk or not at risk based on observed scores.

For this study, CART analyses classified students as at risk or not at risk on the basis of individual performance of each Florida Center for Reading Research Reading Assessment (FRA) task at every possible cutpoint. To ensure a parsimonious model, several specifications were used to limit the number of splits. Guided by Compton et al. (2006), the analyses specified a stopping rule of a minimal parent node size of three students. In addition, the number of splits was limited by specifying a minimum reduction in the relative error (approximately equivalent to $1–R^2$), identified after running a base model with no minimum specified. Each grade-based model included tenfold cross validation to evaluate the quality of the decision tree and determine the appropriate minimum complexity parameter (Breiman, Freidman, Olshen, & Stone, 1984). The recommended minimum standard for complexity parameters is a cross-validation relative error less than one standard error above the minimum cross-validation relative error (Therneau, Atkinson, & Ripley, 2013.)

This study team intended to build and prune CART models to maximize the negative predictive power to as close to .90 as possible without allowing specificity to fall lower than most commercial math screening assessments (.68). To accomplish this, a loss matrix was specified in each model that added weights to specific classification categories. To increase the negative predictive power, weights were set to reduce false negatives. Each grade-based model included tenfold cross validations to evaluate the quality of the decision tree and determine the appropriate minimum complexity parameter (Breiman et al., 1984). A recommended minimum standard is the value of the complexity parameter that results in a cross-validation relative error of less than one standard error above the minimum cross-validation relative error (Therneau et al., 2013).

Appendix B. Descriptive statistics

This appendix provides the number and percentage of students in the sample meeting expectations on the Stanford Achievement Test, Tenth Edition (SAT-10), Reading Comprehension and Mathematics (table B1), the means and standard deviations for Florida Center for Reading Research Reading Assessment (FRA) scores and SAT-10 Reading Comprehension and Mathematics scores (table B2), and correlations between FRA task scores and SAT-10 Reading Comprehension and Mathematics scores (table B3).

Table B1. Students meeting expectations on the Stanford Achievement Test, Tenth Edition, Reading Comprehension and Mathematics

	Reading comprehension		Mathematics	
Grade	Number	Percent	Number	Percent
3	970	70	870	63
4	880	67	787	60
5	881	65	956	70
6	624	51	456	38
7	640	51	548	44
8	715	57	600	48
Total	4,710		4,217	

Source: Authors' analysis of school district data for 2012/13.

Table B2. Means and standard deviations for scores on FRA tasks and Stanford Achievement Test, Tenth Edition, Reading Comprehension and Mathematics scores

	FRA								Stanford Achievement Test, Tenth Edition			
	Reading comprehension score		Syntactic knowledge score		Vocabulary knowledge score		Word recognition score		Reading comprehension		Mathematics	
Grade	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
3	383	57	375	87	411	82	369	63	655	44	634	45
4	432	70	426	91	445	59	423	89	659	39	641	41
5	473	80	453	92	493	79	445	88	671	36	673	42
6	489	88	446	89	498	73	455	87	669	42	660	39
7	513	104	490	91	519	74	522	91	677	36	675	41
8	552	125	507	107	549	73	510	100	686	33	686	37

FRA is Florida Center for Reading Research Reading Assessment. SD is standard deviation.

Source: Authors' analysis of school district data for 2012/13.

Table B3. Correlations between FRA task scores and Stanford Achievement Test, Tenth Edition, Reading Comprehension and Mathematics scores, by grade level

Grade and score	SAT-10 Reading Comprehension score	SAT-10 Mathematics score
Grade 3		
SAT-10 Reading Comprehension	—	
SAT-10 Mathematics	.66*	—
FRA reading comprehension	.73*	.62*
FRA syntactic knowledge	.47*	.43*
FRA vocabulary knowledge	.55*	.47*
FRA word recognition	.40*	.39*
Grade 4		
SAT-10 Reading Comprehension	—	
SAT-10 Mathematics	.66*	—
FRA reading comprehension	.70*	.63*
FRA syntactic knowledge	.53*	.50*
FRA vocabulary knowledge	.41*	.37*
FRA word recognition	.41*	.43*
Grade 5		
SAT-10 Reading Comprehension	—	
SAT-10 Mathematics	.68*	—
FRA reading comprehension	.72*	.66*
FRA syntactic knowledge	.59*	.54*
FRA vocabulary knowledge	.58*	.49*
FRA word recognition	.45*	.39*
Grade 6		
SAT-10 Reading Comprehension	—	
SAT-10 Mathematics	.65*	—
FRA reading comprehension	.72*	.65*
FRA syntactic knowledge	.57*	.50*
FRA vocabulary knowledge	.52*	.41*
FRA word recognition	.47*	.43*
Grade 7		
SAT-10 Reading Comprehension	—	
SAT-10 Mathematics	.64*	—
FRA reading comprehension	.65*	.65*
FRA syntactic knowledge	.58*	.51*
FRA vocabulary knowledge	.45*	.41*
FRA word recognition	.43*	.39*
Grade 8		
SAT-10 Reading Comprehension	—	
SAT-10 Mathematics	.62*	—
FRA reading comprehension	.67*	.64*
FRA syntactic knowledge	.61*	.56*
FRA vocabulary knowledge	.47*	.42*
FRA word recognition	.48*	.47*

* Significant at $p < .01$.

SAT-10 is Stanford Achievement Test, Tenth Edition. FRA is Florida Center for Reading Research Reading Assessment.

Note: Correlations are based on data from all students in the grade level.

Source: Authors' analysis of school district data for 2012/13.

Appendix C. Comparing screening accuracy statistics

This appendix provides accuracy statistics to facilitate comparison of commercial math screening assessments (table C1) with the Florida Center for Reading Research Reading Assessment (FRA) for math outcomes (table C2) and of the FRA predicting math outcomes with the FRA predicting reading comprehension outcomes (table C3).

Table C1. Accuracy statistics of commercial math screening assessments

Assessment	Sensitivity	Specificity	False positive	False negative	Positive predictive power	Negative predictive power	Overall accuracy rate
Acuity	.57–.65	.86–.95	.05–.14	.35–.44	.10–.53	.93–.99	.83–.93
AIMSweb	.77–.83	.68–.82	.18–.32	.15–.25	.21–.56	.93–.97	.70–.82
Discovery Education Predictive Assessment	.81–.94	.57–.83	.16–.43	.06–.19	.49–.93	.75–.96	.73–.90
EasyCBM	.78–.93	.65–.85	.15–.35	.07–.22	.35–.76	.85–.98	.72–.85
Measures of Academic Progress	.63–.81	.89–.94	.06–.11	.22–.40	.65–.76	.88–.95	.84–.91
Iowa Test of Basic Skills	.39–.74	.86–.94	.03–.14	.26–.61	.41–.93	.64–.96	.70–.87
STAR	.75	.74–.79	.21–.26	.25–.26	.47–.59	.88–.91	.75–.77

Note: Includes math screening assessments that have screening accuracy statistics reviewed by the National Center on Response to Intervention.

Source: National Center on Response to Intervention, 2012.

Table C2. Screening accuracy statistics on how well FRA task scores identify students at risk of scoring below the 50th percentile of Stanford Achievement Test, Tenth Edition, Mathematics scores

Grade	Sensitivity	Specificity	False positive	False negative	Positive predictive power	Negative predictive power	Overall accuracy rate
3	.84	.74	.26	.16	.66	.89	.78
4	.86	.74	.25	.14	.69	.89	.79
5	.84	.73	.25	.14	.69	.89	.79
6	.84	.73	.27	.16	.58	.91	.76
7	.84	.68	.32	.16	.77	.77	.77
8	.86	.80	.20	.14	.82	.84	.83

FRA is Florida Center for Reading Research Reading Assessment.

Source: Authors' analysis of school district data for 2012/13.

Table C3. Screening accuracy statistics on how well FRA task scores identify students at risk of scoring below the 50th percentile of Stanford Achievement Test, Tenth Edition, Reading Comprehension scores

Grade	Sensitivity	Specificity	False positive	False negative	Positive predictive power	Negative predictive power	Overall accuracy rate
3	.72	.86	.14	.28	.69	.87	.82
4	.80	.83	.17	.20	.67	.91	.82
5	.80	.82	.18	.20	.71	.88	.81
6	.81	.88	.12	.19	.87	.82	.85
7	.87	.73	.27	.13	.75	.85	.79
8	.81	.85	.15	.19	.80	.86	.83

FRA is Florida Center for Reading Research Reading Assessment.

Source: Authors' analysis of school district data for 2012/13.

Appendix D. Increasing the transparency of screening accuracy decisions

The goal of a screening assessment is to accurately predict an end-of-year outcome. Sometimes, the highest accuracy also results in a classification and regression tree (CART) model with so many decision rules that it cannot be easily interpreted. The number of decision rules can be reduced to a more easily interpreted model, though this can result in poorer screening accuracy statistics. Tables D1 and D3 provide screening accuracy statistics for the models with the best screening accuracy statistics; tables D2 and D4 provide screening accuracy statistics for the models that reduce the number of decision rules to five or fewer. The screening accuracy statistics in table D1 for grades 3, 4, 7, and 8 are closer to desired values than the screening accuracy statistics in table D2; however, the number of decision rules is easier to interpret in the model for table D2 than in the model for table D1. The pattern is similar (with different grade levels) in tables D3 and D4.

Table D1. Screening accuracy statistics for the models with the best screening accuracy statistics when using FRA task scores to identify students at risk of scoring below the 50th percentile of Stanford Achievement Test, Tenth Edition, Mathematics scores

Grade	Sensitivity	Specificity	False positives	False negatives	Positive predictive power	Negative predictive power	Overall accuracy rate	Number of decision rules
3	.84	.74	.26	.16	.66	.89	.78	21
4	.86	.74	.25	.14	.69	.89	.79	8
5	.84	.73	.25	.14	.69	.89	.79	1
6	.84	.73	.27	.16	.58	.91	.76	1
7	.79	.77	.23	.21	.81	.74	.78	5
8	.86	.80	.20	.14	.82	.84	.83	9

FRA is Florida Center for Reading Research Reading Assessment.

Source: Authors' analysis of school district data for 2012/13.

Table D2. Screening accuracy statistics for the models that reduce the number of decision rules to five or fewer when using FRA task scores to identify students at risk of scoring below the 50th percentile of Stanford Achievement Test, Tenth Edition, Mathematics scores

Grade	Sensitivity	Specificity	False positives	False negatives	Positive predictive power	Negative predictive power	Overall accuracy rate	Number of decision rules
3	.77	.73	.27	.23	.63	.84	.75	5
4	.80	.78	.22	.20	.71	.86	.79	4
5	.84	.73	.25	.14	.69	.89	.79	1
6	.84	.73	.27	.16	.58	.91	.76	1
7	.85	.70	.31	.14	.76	.82	.78	1
8	.86	.70	.30	.14	.76	.82	.78	1

FRA is Florida Center for Reading Research Reading Assessment.

Source: Authors' analysis of school district data for 2012/13.

Table D3. Screening accuracy statistics for the models with the best screening accuracy statistics when using FRA task scores to identify students at risk of scoring below the 50th percentile of Stanford Achievement Test, Tenth Edition, Reading Comprehension scores

Grade	Sensitivity	Specificity	False positives	False negatives	Positive predictive power	Negative predictive power	Overall accuracy rate	Number of decision rules
3	.72	.86	.14	.28	.69	.87	.82	3
4	.80	.83	.17	.20	.67	.91	.82	3
5	.80	.82	.18	.20	.71	.88	.81	1
6	.81	.88	.12	.19	.87	.82	.85	11
7	.87	.73	.27	.13	.75	.85	.79	7
8	.81	.85	.15	.19	.80	.86	.83	11

FRA is Florida Center for Reading Research Reading Assessment.

Source: Authors' analysis of school district data for 2012/13.

Table D4. Screening accuracy statistics for the models that reduce the number of decision rules to five or fewer when using FRA task scores to identify students at risk of scoring below the 50th percentile of Stanford Achievement Test, Tenth Edition, Reading Comprehension scores

Grade	Sensitivity	Specificity	False positives	False negatives	Positive predictive power	Negative predictive power	Overall accuracy rate	Number of decision rules
3	.82	.77	.23	.18	.60	.91	.78	1
4	.80	.83	.17	.20	.67	.91	.82	1
5	.80	.82	.18	.20	.71	.88	.81	1
6	.80	.79	.21	.20	.80	.80	.80	1
7	.87	.66	.34	.13	.71	.84	.76	1
8	.87	.72	.28	.13	.70	.88	.79	3

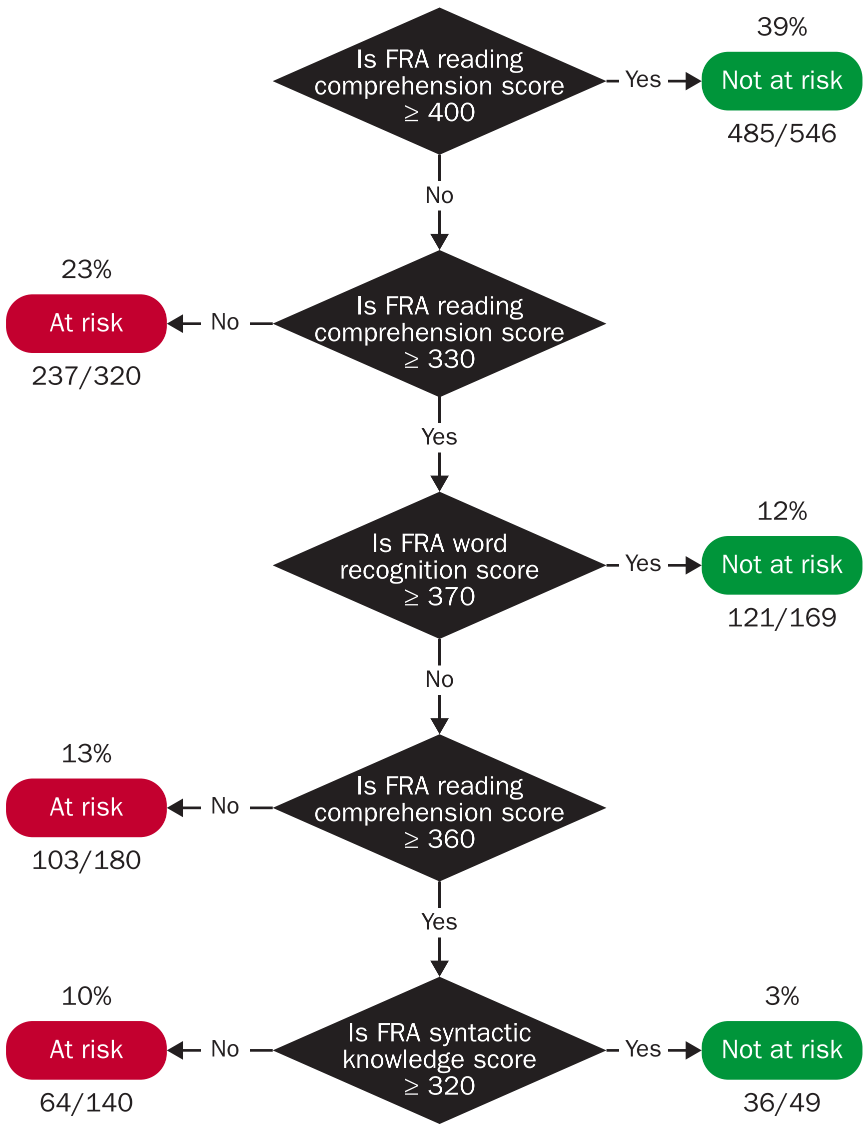
FRA is Florida Center for Reading Research Reading Assessment.

Source: Authors' analysis of school district data for 2012/13.

Appendix E. Decision trees for each grade level

This appendix provides decision trees for using Florida Center for Reading Research Reading Assessment (FRA) scores to identify students at risk of scoring below the 50th percentile of Stanford Achievement Test, Tenth Edition, Mathematics or Reading Comprehension scores. The models balance the need to maintain interpretability with the need to maintain acceptable levels of screening accuracy (see tables D2 and D4 in appendix D for screening accuracy statistics). In the figures diamonds represent decision rules, and ovals represent categories of students identified as at risk or not at risk. The denominator of the fraction is the number of students from the sample who were classified in that category, the numerator is the number of students in that category who were correctly identified, and the percentage identifies the proportion of the sample that fell into that category.

Figure E1. Decision tree for grade 3 math predictions

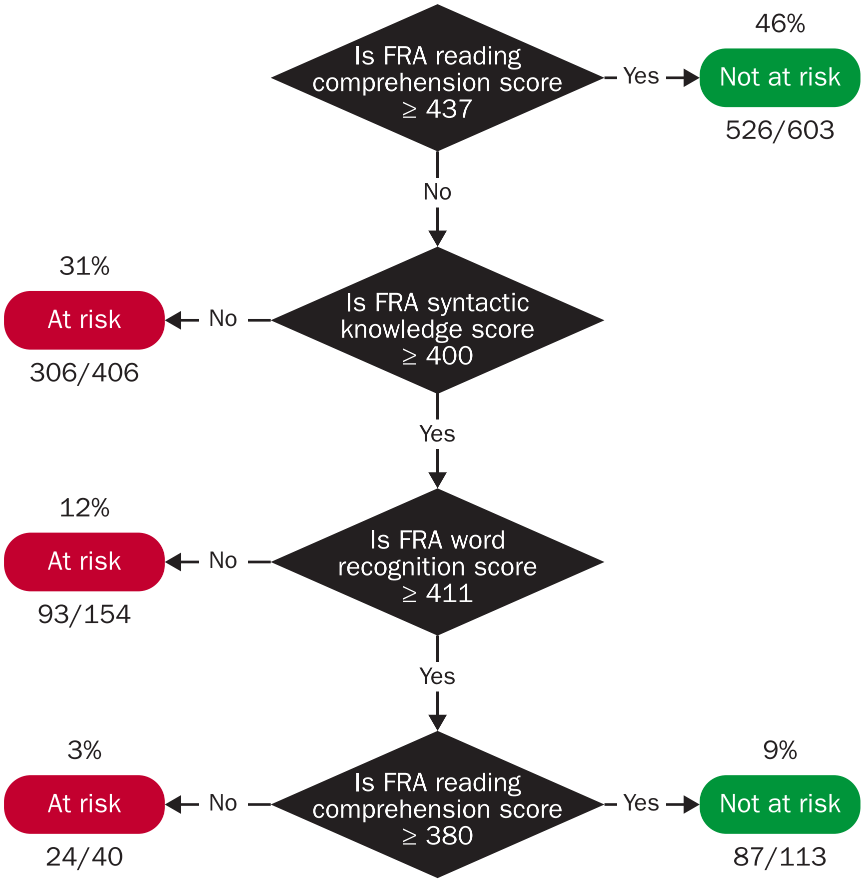


FRA is Florida Center for Reading Research Reading Assessment.

Note: Uses FRA task scores to identify grade 3 students at risk of scoring below the 50th percentile on the Stanford Achievement Test Mathematics, Tenth Edition.

Source: Authors' analysis of school district data for 2012/13.

Figure E2. Decision tree for grade 4 math predictions



FRA is Florida Center for Reading Research Reading Assessment.

Note: Uses FRA task scores to identify grade 4 students at risk of scoring below the 50th percentile on the Stanford Achievement Test Mathematics, Tenth Edition. Percentages do not sum to 100 because of rounding.

Source: Authors' analysis of school district data for 2012/13.

Figure E3. Decision tree for grade 5 math predictions

FRA is Florida Center for Reading Research Reading Assessment.

Note: Uses FRA task scores to identify grade 5 students at risk of scoring below the 50th percentile on the Stanford Achievement Test Mathematics, Tenth Edition.

Source: Authors' analysis of school district data for 2012/13.

Figure E4. Decision tree for grade 6 math predictions

FRA is Florida Center for Reading Research Reading Assessment.

Note: Uses FRA task scores to identify grade 6 students at risk of scoring below the 50th percentile on the Stanford Achievement Test Mathematics, Tenth Edition.

Source: Authors' analysis of school district data for 2012/13.

Figure E5. Decision tree for grade 7 math predictions

FRA is Florida Center for Reading Research Reading Assessment.

Note: Uses FRA task scores to identify grade 7 students at risk of scoring below the 50th percentile on the Stanford Achievement Test Mathematics, Tenth Edition.

Source: Authors' analysis of school district data for 2012/13.

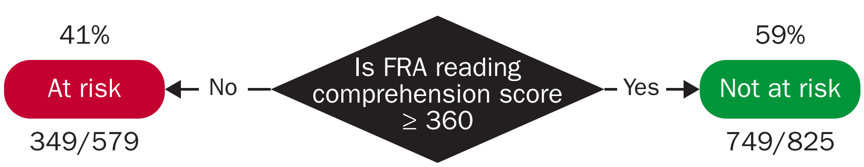
Figure E6. Decision tree for grade 8 math predictions

FRA is Florida Center for Reading Research Reading Assessment.

Note: Uses FRA task scores to identify grade 8 students at risk of scoring below the 50th percentile on the Stanford Achievement Test Mathematics, Tenth Edition.

Source: Authors' analysis of school district data for 2012/13.

Figure E7. Decision tree for grade 3 reading predictions



FRA is Florida Center for Reading Research Reading Assessment.

Note: Uses FRA task scores to identify grade 3 students at risk of scoring below the 50th percentile on the Stanford Achievement Test Reading Comprehension, Tenth Edition.

Source: Authors' analysis of school district data for 2012/13.

Figure E8. Decision tree for grade 4 reading predictions

FRA is Florida Center for Reading Research Reading Assessment.

Note: Uses FRA task scores to identify grade 4 students at risk of scoring below the 50th percentile on the Stanford Achievement Test Reading Comprehension, Tenth Edition.

Source: Authors' analysis of school district data for 2012/13.

Figure E9. Decision tree for grade 5 reading predictions

FRA is Florida Center for Reading Research Reading Assessment.

Note: Uses FRA task scores to identify grade 5 students at risk of scoring below the 50th percentile on the Stanford Achievement Test Reading Comprehension, Tenth Edition.

Source: Authors' analysis of school district data for 2012/13.

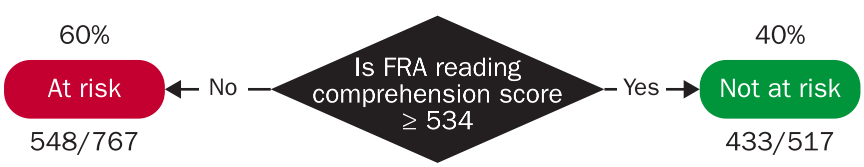
Figure E10. Decision tree for grade 6 reading predictions

FRA is Florida Center for Reading Research Reading Assessment.

Note: Uses FRA task scores to identify grade 6 students at risk of scoring below the 50th percentile on the Stanford Achievement Test Reading Comprehension, Tenth Edition.

Source: Authors' analysis of school district data for 2012/13.

Figure E11. Decision tree for grade 7 reading predictions

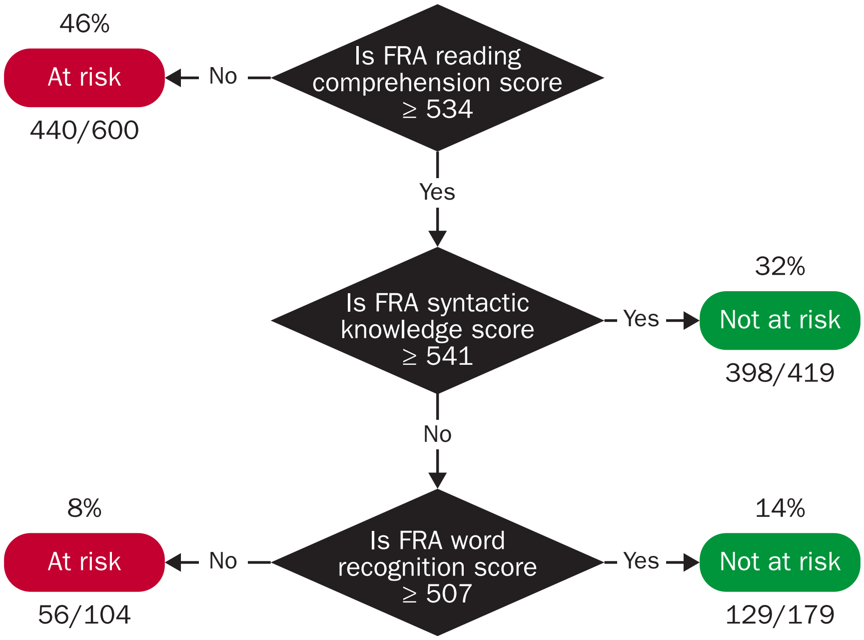


FRA is Florida Center for Reading Research Reading Assessment.

Note: Uses FRA task scores to identify grade 7 students at risk of scoring below the 50th percentile on the Stanford Achievement Test Reading Comprehension, Tenth Edition.

Source: Authors' analysis of school district data for 2012/13.

Figure E12. Decision tree for grade 8 reading predictions



FRA is Florida Center for Reading Research Reading Assessment.

Note: Uses FRA task scores to identify grade 8 students at risk of scoring below the 50th percentile on the Stanford Achievement Test Reading Comprehension, Tenth Edition.

Source: Authors' analysis of school district data for 2012/13.

References

Biancarosa, G., & Snow, C. E. (2006). *Reading next: A vision for action and research in middle and high school literacy: A report to Carnegie Corporation of New York* (2nd edition). Washington, DC: Alliance for Excellent Education.

Breiman, L., Friedman, J., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Boca Raton, FL: CRC Press.

Codding, R. S., Petscher, Y., & Truckenmiller, A. (2014). CBM reading, mathematics, and writing at the secondary level: Examining latent composite relations among indices and unique predictions with a state achievement test. *Journal of Educational Psychology, 107*(2), 437–450. http://eric.ed.gov/?id=EJ1061898

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(2), 394–409. http://eric.ed.gov/?id=EJ742190

Crawford, L., Tindal, G., & Steiber, S. (2001). Using oral reading rate to predict student performance on statewide achievement tests. *Educational Assessment, 7*(4), 303–323. http://eric.ed.gov/?id=EJ656682

Fletcher, J. M. (2005). Predicting math outcomes: Reading predictors and comorbidity. *Journal of Learning Disabilities*, *38*(4), 308–312. http://eric.ed.gov/?id=EJ695627

Foorman, B. R., Petscher, Y., & Schatschneider, C. (2015). *Florida Assessments for Instruction in Reading, aligned to the Language Arts Florida Standards, grades 3–12 technical manual*. Tallahassee, FL: Florida Center for Reading Research. Retrieved March 2, 2015, from http://www.fcrr.org/_/documents/FRA_3–12_Tech_Manual_v1.pdf.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., et al. (2008). *Assisting students struggling with reading: Response to Intervention and multi-tier intervention for reading in the primary grades. A practice guide* (NCEE No. 2009–4045). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. http://eric.ed.gov/?id=ED504264

Harcourt. (2003). *Stanford Achievement Test-10*. New York, NY: Author.

Helwig, R., Rozek-Tedesco, M. A., Heath B., & Tindal, G. (1999). Reading as an access to mathematics problem-solving on multiple choice tests for sixth grade students. *Journal of Educational Research, 93*(2), 113–125.

Jenkins, J. R. (2003, December). *Candidate measures for screening at-risk students*. Paper presented at the Conference on Response to Intervention as Learning Disabilities Identification, sponsored by the National Research Center on Learning Disabilities, Kansas City, MO.

Jiban, C. L., & Deno, S. L. (2007). Using math and reading curriculum-based measurements to predict state mathematics test performance: Are simple one-minute measures technically adequate? *Assessment for Effective Intervention*, *32*(2), 78–89. http://eric.ed.gov/?id=EJ793332

Koon, S., & Petscher, Y. (2015). *Comparing methodologies for developing an early warning system* (REL 2015–077). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. http://eric.ed.gov/?id=ED554441

Koon, S., Petscher, Y., & Foorman, B. R. (2014). *Using evidence-based decision trees instead of formulas to identify at risk readers* (REL 2014–036). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. http://eric.ed.gov/?id=ED545225

National Center on Response to Intervention. (2012). *Universal screening tools chart*. Washington, DC: Author. Retrieved March 2, 2015, from http://www.rti4success.org/resources/tools-charts/screening-tools-chart.

Petscher, Y., Kim, Y. S., & Foorman, B. R. (2011). The importance of predictive power in early screening assessments: Implications for placement in the response to intervention framework. *Assessment for Effective Intervention, 36*(3), 158–166. http://eric.ed.gov/?id=EJ925613

Schatschneider, C., Petscher, Y., & Williams, K. M. (2008). How to evaluate a screening process: The vocabulary of screening and what educators need to know. In Justice, L., & Vukelic, C. (Eds.), *Achieving excellence in preschool literacy instruction*, 304–316. New York, NY: Guilford.

Swets, J. A. (1992). The science of choosing the right decision threshold in high-stakes diagnostics. *American Psychologist, 47*(4), 522–532.

Therneau, T. M., & Atkinson, E. J. (2013). *An introduction to recursive partitioning using the RPART routines*. Technical report. Rochester, MN: Mayo Foundation. Retrieved December 16, 2013, from http://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf.

Therneau, T. M., Atkinson, B., & Ripley, B. (2013). *rpart: Recursive partitioning*. R package version 4.1–4. Project America. Retrieved December 16, 2013, from http://cran.r-project.org/web/packages/rpart/rpart.pdf.

Thurber, R. S., Shinn, M. R., & Smolkowski, K. (2002). What is measured in mathematics tests? Construct validity of curriculum-based mathematics measures. *School Psychology Review, 31*(4), 498–513. http://eric.ed.gov/?id=EJ667623

Wayman, M. M., Wallace, T., Wiley, H. I., Tichá, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education*, *41*(2), 85–120. http://eric.ed.gov/?id=EJ775114

Zirkel, P. A., & Thomas, L. B. (2010). State laws and guidelines for implementing RTI. *Teaching Exceptional Children*, *43*(1), 60–73. http://eric.ed.gov/?id=EJ898488

The Regional Educational Laboratory Program produces 7 types of reports

Making Connections
Studies of correlational relationships

Making an Impact
Studies of cause and effect

What's Happening
Descriptions of policies, programs, implementation status, or data trends

What's Known
Summaries of previous research

Stated Briefly
Summaries of research findings for specific audiences

Applied Research Methods
Research methods for educational settings

Tools
Help for planning, gathering, analyzing, or reporting data or research