U.S. Department of Education
September 2013

Evaluating the screening accuracy of the Florida Assessments for Instruction in Reading (FAIR)

Barbara R. Foorman
Florida Center for Reading Research
at the Florida State University

Sarah Kershaw
Florida Center for Reading Research
at the Florida State University

Yaacov Petscher
Florida Center for Reading Research
at the Florida State University

Key findings

The reading comprehension screen of the Florida Assessments for Instruction in Reading (FAIR) was more accurate than the 2011 Florida Comprehensive Assessment Test (FCAT) 2.0 scores in correctly identifying students as not at risk for failing to meet grade-level standards on the 2012 FCAT 2.0. Using both the FAIR screen and the 2011 FCAT 2.0 lowered the underidentification rate of at-risk students by 12–20 percentage points compared with the results using the 2011 FCAT 2.0 score alone.

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

REL SOUTHEAST
Regional Educational Laboratory
At Florida State University

REL 2013–008

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 2013

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-IES-12-C-0011 by Regional Educational Laboratory Southeast administered by Florida Center for Reading Research, 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.

Foorman, B.R., Kershaw, S., & Petscher, Y. (2013). *Evaluating the screening accuracy of the Florida Assessments for Instruction in Reading (FAIR).* (REL 2013–008). 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

Florida requires that students who do not meet grade-level reading proficiency standards on the end-of-year state assessment (Florida Comprehensive Assessment Test, FCAT) receive intensive reading intervention. With the stakes so high, teachers and principals are interested in using screening or diagnostic assessments to identify students with a strong likelihood of failing to meet grade-level proficiency standards on the FCAT. Since 2009 Florida has administered a set of interim assessments (Florida Assessments for Instruction in Reading, FAIR) three times a year (fall, winter, and spring) to obtain information on students' probability of meeting grade-level standards on the end-of-year FCAT.

In 2010/11 the Florida Department of Education aligned the FCAT to new standards (Next Generation Sunshine State Standards) and renamed it the FCAT 2.0 but retained the 2009/10 cutscores. In 2011/12 it changed the FCAT 2.0 cutscores. The share of students meeting grade-level standards on the FCAT 2.0 fell to 53 percent in 2012 from 72 percent in 2011. This drop led the Florida Department of Education to partner with the Regional Educational Laboratory Southeast to analyze student performance on the FAIR reading comprehension screen and FCAT 2.0 to determine how well the FAIR and the 2011 FCAT 2.0 scores predict 2012 FCAT 2.0 performance.

The study addresses two research questions:

- What is the association between performance on the 2012 FCAT 2.0 and two scores from the FAIR reading comprehension screen across grades 4–10 and the three FAIR assessment periods (predictive validity)?
- How much does adding the FAIR reading comprehension screen affect identification errors beyond those identified through 2011 FCAT 2.0 scores (screening accuracy)?

A stratified random sample of student-level archival data for approximately 700,000 students in grades 4–10 was obtained from the state's Progress Monitoring and Reporting Network. Data included the spring 2011 and 2012 FCAT 2.0 reading standard scores and proficiency levels, the FAIR reading comprehension ability scores, and the FCAT success probability scores (which combines FAIR reading comprehension ability scores and 2011 FCAT 2.0 scores) for the fall, winter, and spring assessment periods in the 2011/12 school year.

Performance on the 2012 FCAT 2.0 was found to have a stronger correlation with FCAT success probability scores than with FAIR reading comprehension ability scores. In addition, using 2011 FCAT 2.0 scores alone to predict 2012 FCAT 2.0 scores underidentified 16–24 percent of students as at risk. Adding FAIR reading comprehension ability scores dropped the underidentification rate by 12–20 percentage points.

Contents

What motivated the study?

Florida requires that students who do not meet grade-level reading proficiency standards on the end-of-year state assessment (Florida Comprehensive Assessment Test, FCAT) receive intensive reading intervention (see box 1 for definitions of key terms). Placement decisions are informed by performance on the prior-year FCAT. As a result, teachers and principals are interested in using screening or diagnostic assessments to identify students with a high likelihood of failing to meet the grade-level proficiency standards on the FCAT. Since 2009 Florida has used the Florida Assessments for Instruction in Reading (FAIR)—a set of interim assessments (fall, winter, and spring)—to obtain information on students' likelihood of meeting grade-level standards on the end-of-year FCAT.

The FAIR is a statewide K–12 literacy screen and diagnostic interim assessment system developed by the Florida Center for Reading Research and owned, hosted, and maintained by the Florida Department of Education. The FAIR consists of a K–2 component and a grade 3–12 component, with multiple assessment periods (fall, winter, and spring). In the grade 3–12 component, students are administered three tasks: a computer-adaptive test of reading comprehension, a computer-adaptive test of spelling, and an assessment of reading efficiency. Because all three FAIR tasks can be given in the same 45-minute class period, teachers frequently give all three tasks to all students. The FAIR reading comprehension screen yields two scores: the FAIR reading comprehension ability (RCA) score and the FCAT success probability (FAIR FSP) score. The FAIR FSP score is calculated using the FAIR RCA score and the 2011 FCAT 2.0 score.

The priority for the FAIR screen is to minimize under-identification of at-risk students because students who are underidentified miss the opportunity for timely interventions

Districts and schools are required to use interim assessments in reading. Although they are not required to use the FAIR, a majority of districts and schools in Florida do. Students in schools that administer the FAIR first take the reading comprehension screen, which consists of up to four passages with questions and multiple-choice answers written to the FCAT 2.0 specifications. Generic estimates of reliability from item response theory range from .90 in grade 3 to .92 in grades 5–12 (Florida Department of Education, 2009a,b). The FAIR reading comprehension screen was designed to identify students who are not likely to meet grade-level proficiency standards on the FCAT 2.0.

So as not to miss students needing intensive reading intervention, the FAIR screen was designed to maximize the negative predictive power of its scores (Petscher, Kim, & Foorman, 2011) such that 85 percent of students identified as not at risk on the screen would meet grade-level proficiency standards on the FCAT 2.0. The 85 percent cutscore was selected to reduce the percentage of students who do not meet grade-level standards on the 2012 FCAT 2.0 but who were not identified as being at risk by the screen (false negative error).

While many screening assessments seek to maximize the percentage of students correctly identified as at risk of not meeting grade-level standards on an outcome assessment (the sensitivity of the screen), the priority for the FAIR screen is to minimize underidentification of at-risk students because students who are underidentified miss the opportunity for timely interventions. Trying to minimize false negative errors may raise the percentage of students identified as at risk by the screen but who actually meet grade-level standards on the outcome assessment (false positive error).

Box 1. Key terms

Diagnostic assessment. An assessment that is typically given after an initial screening assessment and that provides specific information to practitioners about a student's strengths and weaknesses.

Florida Assessments for Instruction in Reading (FAIR). The K–12 screening and diagnostic assessment system used in Florida to identify students who are not likely to meet grade-level standards on the Florida Comprehensive Assessment Test (FCAT) at the end of the year.

FAIR FCAT success probability (FAIR FSP) score. A score derived from the prior-year FCAT score and from the current administration of the FAIR reading comprehension ability score that denotes the probability of meeting grade-level proficiency standards on the end-of-year FCAT.

Florida Comprehensive Assessment Test 2.0 (FCAT 2.0). The current annual standards-based, criterion-referenced outcome assessment (2012) used to measure student academic achievement on Florida's Next Generation Sunshine State Standards. FCAT 2.0 is administered in reading (grades 3–10), math (grades 3–8), writing (grades 4, 8, and 10), and science (grades 5 and 8; Florida Department of Education, 2011a,b). This study looks only at the reading component. Results on the FCAT 2.0 reading component are reported as a developmental scaled score (a standard score) and a proficiency level. The standard scores range from 140 to 302 for grades 3–10. The proficiency levels range from a low of 1 to a high of 5. Students are designated as meeting grade-level standards on the FCAT 2.0 if they achieve a proficiency level of 3 or higher.

Interim assessments. Quarterly or monthly assessments, generally administered district- or schoolwide, that evaluate a student's ability to meet grade-level standards on an outcome measure (Hamilton et al., 2009). Assessments often comprise a brief, universal screening assessment and a more in-depth diagnostic assessment. They provide reliable and valid scores and are used to predict student achievement on an outcome measure. Performance on the screen can be used to identify students needing further evaluation of skills as well as students who are expected to perform adequately or in an accelerated fashion on an outcome assessment. The FAIR is an interim assessment.

Negative predictive power. A measure of screening accuracy that reflects the proportion identified as not at risk on the screening assessment who pass the outcome assessment.

Positive predictive power. A measure of screening accuracy that reflects the proportion identified as at risk on the screening assessment who fail the outcome assessment.

Predictive validity. A term that describes the extent to which scores from a measure predict scores on an outcome measure administered at a later date, based on the correlation between scores on the measure and scores on the outcome, with a stronger correlation indicating strong predictive validity.

Screening accuracy. The ability of a measure to distinguish between students who are at risk and students who are not at risk for failing an outcome.

Screening assessment. Brief assessments designed to identify students at risk of failing an outcome. Performance on a screening assessment can be used to identify students who need

(continued)

Box 1. Key terms *(continued)*

further evaluation of skills as well as students who are expected to perform adequately or in an accelerated fashion on an outcome assessment.

Sensitivity. A measure of screening accuracy that reflects the proportion of true positives.

Specificity. A measure of screening accuracy that reflects the proportion of true negatives.

Previous studies found strong correlations between FAIR scores and end-of-year FCAT performance and strong predictive power for the FAIR for students in grades 4–10 (Foorman & Petscher, 2010a; Foorman & Petscher, 2010b; Petscher & Foorman, 2011). However, in the 2010/11 school year the FCAT was changed to align with a new set of standards (Next Generation Sunshine State Standards), and in 2011/12 the cutscores on the revised FCAT (FCAT 2.0) were changed. For example, following those changes, the share of grade 3 students meeting grade-level standards on the FCAT 2.0 fell from 72 percent for the 2011 version to 53 percent for the 2012 version. This drop made it important to study how well the findings of previous reports hold with the new changes made to cutscores on the 2012 FCAT 2.0.

Previous studies found strong correlations between FAIR scores and end-of-year FCAT performance, but in 2010/11 the FCAT was changed to align with a new set of standards, and in 2011/12 the cutscores were changed

How the study was conducted

This section lays out the research questions, describes the students included in the study sample, explains the assessments and scores used to assess student reading ability, and details the methods used to analyze the data.

The research questions guiding this study

The study sought to answer two questions:

- What is the association between performance on the 2012 FCAT 2.0 and two scores from the FAIR reading comprehension screen—the FAIR RCA score and the FAIR FSP score—across grades 4–10 and the three FAIR assessment periods (predictive validity)?
- How much does adding the FAIR reading comprehension screen affect identification errors beyond those identified through the 2011 FCAT 2.0 scores (screening accuracy)?

The study sample

Archival FAIR and FCAT data for 928,834 students in grades 4–10[1] in 2011/12 were obtained from the Progress Monitoring and Reporting Network, hosted and maintained by the Florida Department of Education. The study included students in grades 4–10 because these students had scores for both the 2011 and 2012 FCAT 2.0. The FCAT achievement distribution for all students who took the FAIR at each grade (top of table 1) did not precisely reflect the ability distribution of Florida students (bottom of table 1). Appropriately generalizing findings to students across the state required selecting a random subset of students within each grade to reflect the achievement distribution across the five FCAT levels (students achieving proficiency level 3 or higher are designated as having met grade-level standards). The proportion of students at each FCAT level for each grade was used to

Table 1. Proficiency-level distribution on the 2012 Florida Comprehensive Assessment Test 2.0 for three sets of students, by grade

2012 FCAT 2.0 proficiency level	Grade						
	4	5	6	7	8	9	10
Full sample							
1	15	17	23	22	21	22	23
2	26	26	27	28	30	33	34
3	26	26	26	27	24	23	21
4	23	20	16	16	15	15	15
5	9	11	8	8	9	7	7
Stratified sample							
1	13	15	19	18	17	19	20
2	25	24	24	25	27	29	30
3	27	27	28	29	26	24	22
4	25	22	19	18	18	18	19
5	10	12	10	11	12	9	9
State							
1	13	15	19	18	17	18	20
2	25	24	24	25	27	30	30
3	27	27	28	29	26	24	22
4	25	22	19	19	18	19	19
5	10	12	10	11	12	9	10

FCAT is Florida Comprehensive Assessment Test.

Note: This table displays the percentages of students in each grade scoring at one of five proficiency levels (scoring at level 3 or higher indicates that the student has met grade-level standards) on the 2012 Florida Comprehensive Assessment Test 2.0 for the full sample, a random subset of the full sample (stratified sample), and the population of the state. The stratified sample was used for all subsequent analyses so that findings would be generalizable to the state. Percentages may not sum to 100 percent because of rounding.

Source: Author's analysis based on 2012 data requested from the Progress Monitoring and Reporting Network of the Florida Department of Education.

The end-of-year FCAT 2.0 is a component of Florida's efforts to assess student achievement in reading, writing, math, and science

construct the stratified sample (middle of table 1) from state-aggregated data (http://fcat.fldoe.org/results/default.asp).

A stratified random sample of approximately 700,000 students (100,000 per grade) was drawn from the full sample of 928,834 students. The stratified random sample had the following demographic profile: 51 percent male and 45 percent White, 23 percent Black, 26 percent Hispanic, 2.5 percent Asian, 3 percent more than one race, and less than 1 percent other.[2] Approximately 7 percent of students were identified as having limited English proficiency and were designated as English language learner students, and 60 percent were eligible for free or reduced-price lunch. These characteristics were similar to those for Florida as a whole (table 2).

Measures used

The end-of-year FCAT 2.0 is a component of Florida's efforts to assess student achievement in reading, writing, math, and science as represented in Florida's Next Generation Sunshine State Standards (Florida Department of Education, 2011a,b). The reading portion of the FCAT 2.0 is a group-administered, criterion-referenced test consisting of informational and literary passages followed by multiple-choice items (Florida Department of Education,

Table 2. Percentages of students in various demographic groups by grade for the stratified random sample are similar to those for Florida as a whole, 2011/12

	Grade						
Variable	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Stratified sample							
Male	51	51	51	51	51	51	51
Race/ethnicity[a]							
White	44	45	45	45	46	47	47
Black	22	22	23	23	23	23	22
Hispanic	27	26	26	25	25	24	25
Asian	2.5	2.5	2.5	2.5	2.5	2.5	2.5
More than one race	3	3	3	3	3	3	2.5
Other	0.5	0.5	0.5	0.5	0.5	0.5	0.5
English language learner	10	8	6	5	5	5	5
Eligible for free or reduced-price lunch	63	62	62	60	58	54	51
State							
Male	51	51	51	52	51	52	51
Race/ethnicity[a]							
White	42	42	42	43	43	44	45
Black	23	23	23	23	22	23	23
Hispanic	29	23	29	29	29	27	27
Asian	2.5	2.5	2.5	2.5	2.5	2.5	2.5
More than one race	3	3	3	3	3	3	2.5
Other	0.5	0.5	0.5	0.5	0.5	0.5	0.5
English language learner	9	7	5	5	5	5	5
Eligible for free or reduced-price lunch[b]	58	58	58	58	58	58	58

a. Percentages may not sum to 100 percent because of rounding.

b. The state-level value is used for all grades because the state does not report disaggregated information by grade.

Source: Author's analysis based on 2012 data requested from the Progress Monitoring and Reporting Network of the Florida Department of Education.

Two score types can be derived from the FAIR reading comprehension screen: the reading comprehension ability score and the FCAT success probability score, which conveys the likelihood that a student will meet grade-level standards on the 2012 FCAT 2.0.

2011a,b). The FCAT 2.0 yields a standard score and a proficiency level. The standard score ranges from 140 to 302 within each grade. The proficiency levels range from a low of 1 to a high of 5. Students are designated as meeting grade-level standards on the FCAT 2.0 if they achieve a proficiency level of 3 or higher. Reliability for the FCAT 2.0 reading assessment as estimated by Cronbach's alpha (α) ranges from .89 in grade 10 to .92 in grade 3 (Florida Department of Education, 2011a,b).

The grade 3–12 component of the FAIR includes two parts: a reading comprehension screen that is administered first and a follow-up diagnostic assessment that is given to students who meet specific criteria based on their performance on the reading comprehension screen. Two score types can be derived from the FAIR reading comprehension screen: the RCA score and the FSP score.

The FAIR RCA score is a developmental scaled score that can track changes (that is, growth) in reading comprehension over grades 3–10. Values range from 190 to 1000, with a mean of 500 and a standard deviation of 100. A score nearer 190 indicates that a student's ability is more closely aligned with that of a grade 3 student, whereas a score nearer

1000 indicates ability more closely aligned with that of a grade 10 student. Reported reliability for the FAIR RCA is at least .90 for 60 percent of students and .80 or greater for the remaining 40 percent of students.

The FAIR FSP score is a joint probability value comprising the 2011 FCAT 2.0 score and the FAIR RCA score. The FAIR FSP score conveys the likelihood (percentage chance) that a student will meet grade-level standards on the 2012 FCAT 2.0 (proficiency level 3 or higher). For example, a student with a FAIR FSP score of 20 percent has a 20 percent chance of meeting grade-level standards, and a student with a score of 83 percent has an 83 percent chance. Students with a FAIR FSP score of less than 85 percent are identified as at risk of not meeting grade-level standards on the 2012 FCAT 2.0.[3]

Study design and analysis

This study used correlations to investigate the predictive validity of the FAIR RCA and FSP scores from the fall, winter, and spring assessment periods of 2011/12

This study used the FAIR RCA and FSP scores from the fall, winter, and spring assessment periods of 2011/12. Correlations were used to investigate the predictive validity of these measures for students in grades 4–10. The goal was to replicate a series of historical reports investigating the relationship between FAIR scores and FCAT performance. To complement this investigation of the predictive validity of the FAIR screen, a series of correlation contrast tests (Meng, Rosenthal, & Rubin, 1992) were run to evaluate the extent to which one correlation was significantly different from another (Z test). Because the FAIR FSP score in grades 4–10 is a joint probability value comprising the FAIR RCA score and the 2011 FCAT 2.0 score, it was expected to correlate more strongly with FCAT 2.0 performance than was the FAIR RCA score alone.

To address the second research question, 2×2 contingency tables were created summarizing student performance in meeting a selected benchmark on the outcome assessment (2012 FCAT 2.0) and the FAIR FSP. Such tables are often used to determine the screening or diagnostic efficiency of assessments (Petscher, Kim, & Foorman, 2011). Students meeting grade-level standards (proficiency level 3 or higher) for the 2012 FCAT 2.0 were coded as "1," and students not meeting the grade-level standards (proficiency level 2 or lower) were coded as "0."

When 2011 FCAT 2.0 scores were used as the sole variable for predicting 2012 FCAT 2.0 performance in grades 4–10, the same cutscore was used to dichotomize performance as at risk (level 2 or lower) or not at risk (level 3 or higher). When FAIR FSP scores were used as the predictor, scores of 85 percent or greater were recoded as "0" to reflect that students were not at risk, and values below this threshold were recorded as "1" to reflect that students were at risk.

After taking the 2012 FCAT 2.0, students fell into one of four categories (table 3): at risk on the screen and not meeting the grade-level standards on the outcome assessment (cell A, true positive), at risk on the screen and meeting the grade-level standards on the outcome assessment (cell B, false positive), not at risk on the screen and not meeting the grade-level standards on the outcome assessment (cell C, false negative), and not at risk on the screen and meeting the grade-level standards on the outcome assessment (cell D, true negative).

Several indices of diagnostic efficiency can be calculated from these results:

- Sensitivity (proportion of true positives): A/(A+C).

Table 3. Sample 2×2 contingency table

	2012 FCAT 2.0	
FAIR FSP	Does not meet standards	Meets standards
At risk	A: True positive	B: False positive
Not at risk	C: False negative	D: True negative

FCAT is Florida Comprehensive Assessment Test. FAIR FSP is Florida Assessments for Instruction in Reading
FCAT success probability.

Source: Authors' illustration.

- Specificity (proportion of true negatives): D/(B+D).
- Positive predictive power (proportion identified as at risk on the screen who fail on the outcome assessment: A/(A+B).
- Negative predictive power (proportion identified as not at risk on screen who pass the outcome assessment: D/(C+D).

While researchers have proposed different threshold values for sensitivity and specificity, many look for levels of at least .80, with some recommending at least .90 (Compton, Fuchs, Fuchs, & Bryant, 2006; Jenkins, 2003).

Sensitivity and specificity are examples of frequently reported population-based measures. Positive and negative predictive power are sample-based measures because they are influenced by student performance in the sample. Because schools and districts differ in demographic composition and student performance, these sample-based measures are likely to provide relevant and useful information for schools and districts that adopt a screen.

Correlations were strong within each grade between 2012 FCAT 2.0 performance and both the FAIR RCA and the FAIR FSP scores, indicating that both score types strongly predict 2012 FCAT 2.0 performance

The 2×2 contingency tables were generated first using 2011 FCAT 2.0 scores as the sole predictor of the 2012 FCAT 2.0 performance and then using FAIR FSP scores as the predictor. Sensitivity, specificity, positive predictive power, and negative predictive power were calculated for each condition and descriptively compared for each assessment period (fall, winter, and spring) across grades 4–10.

Study findings

During 2011/12 FAIR RCA scores ranged from 190 to 1000 across grades 4–10, and 2012 FCAT 2.0 standard scores ranged from 140 to 302 at each grade (table 4). Predictive validity between FAIR RCA and FSP scores and 2012 FCAT 2.0 performance ranged from .67 to .79 across grades 4–10 (table 5).

Correlation between Florida Comprehensive Assessment Test 2.0 performance and Florida Assessments for Instruction in Reading scores

The correlations were strong within each grade between 2012 FCAT 2.0 performance and both the FAIR RCA and the FAIR FSP scores (see table 5). These correlations indicate that both score types (FAIR RCA and FAIR FSP) strongly predict 2012 FCAT 2.0 performance for grade 4–10 students. As a follow-up to the predictive correlations, a series of correlation contrast tests (Meng, Rosenthal, & Rubin, 1992) were used to evaluate how significantly different from one another the correlations were (table A1 in appendix A). As expected, the Z test for the significance of the difference between the two correlations

Table 4. Means and standard deviations for 2011/12 Florida Assessments for Instruction in Reading reading comprehension ability scores and for 2011 and 2012 Florida Comprehensive Assessment Test 2.0 standard scores, by grade and assessment period

	2011/12 FAIR						FCAT 2.0			
	Fall		Winter		Spring		2011		2012	
Grade	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
4	404.01	68.35	400.23	78.01	443.13	94.37	202.62	20.28	212.95	20.34
5	455.95	91.12	458.82	79.93	487.02	103.50	211.98	20.99	220.79	21.64
6	477.18	93.35	489.56	99.48	492.37	96.21	218.38	21.04	224.57	21.42
7	499.97	101.94	520.47	90.98	523.29	102.15	224.97	21.05	231.22	21.56
8	533.49	108.91	549.98	104.64	560.96	105.96	230.40	21.21	237.24	22.46
9	552.08	103.17	566.74	104.17	575.08	100.65	235.15	21.91	239.69	21.81
10	599.64	104.26	640.45	94.51	595.41	101.88	239.91	21.43	243.76	20.60

FAIR is Florida Assessments for Instruction in Reading. FCAT is the Florida Comprehensive Assessment Test.

Note: During 2011/12 FAIR reading comprehension ability scores ranged from 190 to 1000 across grades 4–10, and 2012 FCAT 2.0 standard scores ranged from 140 to 302 at each grade.

Source: Author's analysis based on 2012 data requested from the Progress Monitoring and Reporting Network of the Florida Department of Education.

Predictive validity between FAIR RCA and FSP scores and 2012 FCAT 2.0 performance ranged from .67 to .79 across grades 4–10

Table 5. Correlations between 2012 Florida Comprehensive Assessment Test 2.0 performance and 2011/12 Florida Assessments for Instruction in Reading reading comprehension ability scores and FCAT success probability scores, by grade and assessment period

	FAIR					
	Fall		Winter		Spring	
Grade	Reading comprehension ability score	FCAT success probability score	Reading comprehension ability score	FCAT success probability score	Reading comprehension ability score	FCAT success probability score
4	.67	.77	.67	.77	.74	.77
5	.72	.77	.70	.78	.75	.77
6	.72	.79	.73	.79	.74	.79
7	.71	.78	.71	.79	.72	.78
8	.73	.78	.74	.79	.73	.79
9	.73	.78	.70	.78	.72	.78
10	.70	.77	.67	.77	.69	.77

FAIR is Florida Assessments for Instruction in Reading. FCAT is the Florida Comprehensive Assessment Test.

Note: The table displays the correlation between the reading comprehension ability score from the Florida Assessments for Instruction in Reading (FAIR) and Florida Comprehensive Assessment Test (FCAT) 2.0 performance (located in the reading comprehension ability score column of each assessment period) and the correlation between the FAIR FCAT success probability score and FCAT 2.0 performance (located in the FCAT success probability score column of each assessment period) by grade and assessment period using the stratified sample. During 2011/12 FAIR reading comprehension ability scores ranged from 190 to 1000 across grades 4–10, and 2012 FCAT 2.0 standard scores ranged from 140 to 302 at each grade.

Source: Author's analysis based on 2012 data requested from the Progress Monitoring and Reporting Network of the Florida Department of Education.

showed that the correlation with FCAT 2.0 performance was significantly stronger for FAIR FSP scores than for FAIR RCA scores across grades 4–10 and all assessment periods. This finding is not surprising in part because the FAIR FSP score includes the FAIR RCA and the 2011 FCAT 2.0 scores in its calculation and because of the study's large sample size.

Screening accuracy

In general, for predicting 2012 FCAT 2.0 performance, sensitivity and negative predictive power were better using FAIR FSP scores, and specificity and positive predictive power were better using 2011 FCAT 2.0 scores (table 6; a corresponding contingency table using a cutscore of 85 percent on the FAIR FSP for students in grades 4–10 is in table A2 in appendix A). For example, when only 2011 FCAT 2.0 scores were used to predict grade 4 student performance on the 2012 FCAT 2.0, 79 percent of students were correctly identified as not at risk for meeting grade-level standards (negative predictive power), meaning that 21 percent of students (100 percent minus 79 percent) were incorrectly identified as not being at risk when they in fact did not meet grade-level standards. Across all grades these values ranged from 76 percent (grade 8) to 84 percent (grade 10), meaning that 16–24 percent of students were mistakenly identified as performing on grade level when only 2011 FCAT 2.0 scores were used for screening. Such underidentification would have prevented these students from receiving appropriate interventions.

For predicting 2012 FCAT 2.0 performance, sensitivity and negative predictive power were better using FAIR FSP scores, and specificity and positive predictive power were better using 2011 FCAT 2.0 scores

By contrast, using FAIR FSP scores (which combine the FAIR RCA score with the 2011 FCAT 2.0 score) reduced underidentification from 21 percent in grade 4 to 4–6 percent. The percentage point decrease in underidentification from adding the FAIR RCA to the 2011 FCAT 2.0 score is derived by subtracting the negative predictive power for the 2011 FCAT 2.0 from the negative predictive power for the FAIR FSP at each assessment period. Accordingly, the underidentification rate decreased from 12–14 percentage points in grade 10 to 19–20 percentage points in grade 8 (see table 6). The discrepancy between results using the two methods was similar for sensitivity: 2011 FCAT 2.0 scores correctly predicted at risk performance on the 2012 FCAT 2.0 for just 59–88 percent of students in grades 4–10, while FAIR FSP scores correctly predicted at risk performance for 93–99 percent.

For specificity and positive predictive power, misidentification was lower using 2011 FCAT 2.0 scores alone. For sensitivity, using 2011 FCAT 2.0 scores resulted in a misidentification rate that was 23–39 percentage points lower than when using FAIR FSP scores. A similar discrepancy was estimated for positive predictive power, with a misidentification rate of 12–26 percentage points lower across grades and assessment periods using 2011 FCAT 2.0 scores rather than FAIR FSP scores.

Although these findings support the goal of minimizing underidentification, there was a loss of diagnostic accuracy for positive predictive power and specificity when using a FAIR FSP cutscore of 85 percent to predict 2012 FCAT 2.0 performance. That means that more students will be identified as needing remediation than turn out to actually need it. Previous studies using similar screening and outcome measures have shown that lowering the threshold for risk from 85 percent to 70 percent maintains a similar level of negative predictive power while increasing the diagnostic accuracy of other measures (Foorman & Petscher, 2010a,b; Petscher & Foorman, 2011). Thus, lowering the cutscore on FAIR FSP to 70 percent was expected to keep negative predictive power high while lowering the false positive rate, thereby also improving positive predictive power.

Table 6. Measures of screening accuracy when using 2011 Florida Comprehensive Assessment Test (FCAT) 2.0 scores compared with using a Florida Assessments for Instruction in Reading FCAT success probability cutscore of 85 percent to predict 2012 FCAT 2.0 performance, by grade

Measure and grade	Sensitivity	Specificity	Positive predictive power	Negative predictive power
2011 FCAT 2.0				
4	59	94	86	79
5	64	92	84	80
6	70	90	85	78
7	74	89	85	81
8	70	90	86	76
9	84	78	82	81
10	88	73	80	84
2011/12 Fall FAIR FSP				
4	96	60	60	96
5	96	60	61	96
6	98	52	64	96
7	98	54	64	97
8	98	51	65	96
9	98	46	66	96
10	99	39	66	98
2011/12 Winter FAIR FSP				
4	96	59	61	96
5	96	58	61	96
6	97	56	65	96
7	97	56	65	96
8	97	54	66	96
9	98	48	67	96
10	99	44	68	96
2011/12 Spring FAIR FSP				
4	93	71	68	94
5	94	68	67	94
6	97	55	66	96
7	97	59	67	96
8	97	56	68	95
9	98	48	69	96
10	99	37	68	97

FCAT is Florida Comprehensive Assessment Test. FAIR FSP is Florida Assessments for Instruction in Reading FCAT success probability.

Note: The table displays the percentages of correct classification for sensitivity and specificity, and proportions of predictive accuracy for positive predictive power and negative predictive power (see box 1 for definitions) by grade when using performance on the reading portion of the 2011 FCAT 2.0 to predict performance on the reading portion of the 2012 FCAT 2.0 compared with using a cutscore of 85 percent on the FAIR FSP measured for three assessment periods (fall, winter, and spring) to predict performance on the reading portion of the 2012 FCAT 2.0. These percentages were calculated using a series of 2×2 contingency tables that can be found in table A2 in appendix A.

Source: Author's analysis based on 2012 data requested from the Progress Monitoring and Reporting Network of the Florida Department of Education.

Using FAIR FSP scores in place of the 2011 FCAT 2.0 score to predict student performance on the 2012 FCAT 2.0 reduced under-identification from 12–14 percentage points in grade 10 to 19–20 percentage points in grade 8

The pattern of results for the 70 percent cutscore was similar to that for the 85 percent cutscore in that sensitivity and negative predictive power were higher than specificity and positive predictive power (table 7; a corresponding contingency table using a cutscore of 70 percent on the FAIR FSP for students in grades 4–10 is in table A3 in appendix A). However, values for specificity and positive predictive power were higher for the 70 percent cutscore than for the 85 percent cutscore. For example, for the fall FAIR FSP in grade 4 specificity and positive predictive power were both 60 percent for the 85 percent cutscore compared with 76 percent for specificity and 71 percent for positive predictive power for the 70 percent cutscore. Across grades and assessment periods the 70 percent threshold resulted in greater diagnostic accuracy than the 85 percent threshold, with the slight loss in negative predictive power and sensitivity offset by a substantial increase in positive predictive power and specificity.

Implications of the findings

Results showed a strong correlation between FAIR FSP scores and 2012 FCAT 2.0 performance at all grades and assessment periods, in part because the FAIR FSP includes 2011 FCAT 2.0 scores

The study investigated the relationship between Florida's interim assessment (FAIR) and the state outcome test (FCAT 2.0) and examined the diagnostic accuracy of the FAIR and the 2011 FCAT 2.0 in predicting 2012 FCAT 2.0 performance for students in grades 4–10. These questions were addressed using correlations, a correlation contrast test (Meng, Rosenthal, & Rubin, 1992), and 2×2 contingency tables.

Results showed a strong correlation between FAIR FSP scores and 2012 FCAT 2.0 performance at all grades and assessment periods, ranging from .67 in grade 4 in the fall to .79 in several grades at all assessment periods (Cohen, 1988). The correlations were strong, in part because the FAIR FSP includes 2011 FCAT 2.0 scores. However, correlations were also strong, at .67–.75, between FAIR RCA scores and 2012 FCAT 2.0 performance, confirming that FAIR's reading comprehension screen is a valid predictor of 2012 FCAT 2.0 performance.

When investigating the diagnostic accuracy of the FAIR compared with the 2011 FCAT 2.0, the goal was to maximize negative predictive power so that students who are at risk for future reading problems are not falsely identified as not at risk. It is considered a greater error for a screen to identify a student as likely to meet grade-level standards on the criterion who ultimately does not than to identify a student as at risk who ultimately meets standards on the criterion. A student misidentified as not at risk might not receive additional services or support, potentially adversely affecting that student's future success in reading. By contrast, a student misidentified as at risk would be given two additional diagnostic assessments to better understand what type of further instruction might be needed, presenting opportunities to correctly classify the student.

Negative predictive power for meeting grade-level standards on the 2012 FCAT 2.0 was more accurate using FAIR FSP scores (94–98 percent across grades and assessment periods) than 2011 FCAT 2.0 scores alone (76–84 percent across grades). For example, in grade 4 the negative predictive power improved from 79 percent when using 2011 FCAT 2.0 scores to 96 percent when using FAIR FSP scores. When 2011 FCAT 2.0 scores alone were used to predict 2012 FCAT 2.0 performance, 63,527 students across grades 4–10 were identified as not at risk when in fact they did not meet grade-level standards on the 2012 FCAT 2.0 (from 4,429 in grade 10, an 18 percent error rate, to 13,192 in grade 4, a 21 percent error rate). When fall FAIR RCA scores were added to 2011 FCAT scores (the FAIR

Table 7. Measures of screening accuracy when using 2011 Florida Comprehensive Assessment Test (FCAT) 2.0 scores compared with using a Florida Assessments for Instruction in Reading FCAT success probability cutscore of 70 percent to predict 2012 FCAT 2.0 performance, by grade

Measure and grade	Sensitivity	Specificity	Positive predictive power	Negative predictive power
2011 FCAT 2.0				
4	59	94	86	79
5	64	92	84	80
6	70	90	85	78
7	74	89	85	81
8	70	90	86	76
9	84	78	82	81
10	88	73	80	84
2011/12 Fall FAIR FSP				
4	89	76	71	92
5	90	74	70	92
6	93	68	72	92
7	94	69	72	93
8	93	67	73	91
9	95	63	73	92
10	97	55	72	94
2011/12 Winter FAIR FSP				
4	89	76	70	91
5	89	75	71	91
6	92	71	73	95
7	92	72	74	92
8	92	70	74	91
9	94	64	75	91
10	95	61	75	91
2011/12 Spring FAIR FSP				
4	84	84	77	89
5	86	81	75	89
6	93	71	74	92
7	92	73	75	91
8	91	73	76	89
9	94	66	77	90
10	97	52	73	93

FCAT is Florida Comprehensive Assessment Test. FAIR FSP is Florida Assessments for Instruction in Reading FCAT success probability.

Note: The table displays the percentages of correct classification for sensitivity and specificity, and proportions of predictive accuracy for positive predictive power and negative predictive power (see box 1 for definitions) by grade when using performance on the reading portion of the 2011 FCAT 2.0 to predict performance on the reading portion of the 2012 FCAT 2.0 compared with using a cutscore of 70 percent on the FAIR FSP measured for three assessment periods (fall, winter, and spring) to predict performance on the reading portion of the 2012 FCAT 2.0. These percentages were calculated using a series of 2×2 contingency tables that can be found in table A3 in appendix A.

Source: Author's analysis based on 2012 data requested from the Progress Monitoring and Reporting Network of the Florida Department of Education.

Values for specificity and positive predictive power were higher for the 70 percent cutscore than for the 85 percent cutscore

FSP), roughly 90 percent fewer students were underidentified (4,097 fewer in grade 10, a 93 percent reduction, to 11,933 fewer in grade 4, a 90 percent reduction).

By contrast, positive predictive power was more accurate for 2012 FCAT 2.0 performance using 2011 FCAT 2.0 scores rather than FAIR FSP scores. Roughly 8,687–18,916 students across grades 4–10 and across assessment periods were overidentified using FAIR FSP scores, meaning that they were identified as at risk when they were not.

The impact of using a 70 percent cutscore on the FAIR FSP instead of the 85 percent threshold for predicting 2012 FCAT 2.0 performance was also investigated. The 70 percent alternative was investigated to replicate a series of historical reports that examined the use of a 70 percent cutscore on the FAIR FSP (Foorman & Petscher, 2010a; Petscher & Foorman, 2011). Using a 70 percent cutscore on the FAIR FSP to predict FCAT 2.0 performance slightly lowered the accuracy for negative predictive power and sensitivity but substantially increased the accuracy for positive predictive power and specificity, resulting in better balance across measures of diagnostic accuracy.

Although the alternative cutscore of 70 percent on the FAIR FSP resulted in better balance across measures of diagnostic accuracy, more than one alternative cutscore needs to be examined to determine an optimal balance

Limitations of the study

While the study findings provide valuable information for the Florida Department of Education on the predictive validity and diagnostic accuracy of the FAIR reading comprehension screen for predicting 2012 FCAT 2.0 performance, the work is limited by the methods used. Extending the research to examine other methods of predictive validity and diagnostic accuracy might improve understanding of the relationship between FAIR scores and 2012 FCAT 2.0 performance. Using more rigorous methods, such as structural equation modeling instead of correlations, to assess the predictive validity of the FAIR score types might lead to different findings.

Similarly, although the alternative cutscore of 70 percent on the FAIR FSP resulted in better balance across measures of diagnostic accuracy, more than one alternative cutscore needs to be examined to determine an optimal balance. Evaluation of a receiver operating characteristic curve could be used to identify an optimal cutscore on the FAIR FSP.

If such further investigation were of interest to the Florida Department of Education, future work could examine these various methods of predictive validity and diagnostic accuracy to better understand the relationship between FAIR RCA screen scores and 2012 FCAT 2.0 performance.

Appendix A. Additional statistics

Table A1. Correlation contrast tests comparing the correlations between 2011/12 Florida Assessments of Instruction in Reading reading comprehension ability scores and 2012 Florida Comprehensive Assessment Test (FCAT) 2.0 performance and between 2011/12 FCAT success probability scores and 2012 FCAT 2.0 performance, by assessment period and grade

Assessment period and grade	*Z*-score	95 percent confidence interval		*p*-value
		Lower bound	Upper bound	
Fall				
4	61.63	0.20	0.22	< .001
5	39.29	0.11	0.12	< .001
6	54.05	0.16	0.17	< .001
7	52.64	0.15	0.16	< .001
8	41.35	0.11	0.12	< .001
9	39.31	0.11	0.12	< .001
10	46.74	0.15	0.16	< .001
Winter				
4	63.63	0.20	0.22	< .001
5	56.19	0.17	0.18	< .001
6	47.00	0.14	0.15	< .001
7	57.36	0.18	0.19	< .001
8	41.43	0.12	0.13	< .001
9	56.92	0.17	0.18	< .001
10	65.37	0.20	0.22	< .001
Spring				
4	21.48	0.06	0.08	< .001
5	16.38	0.04	0.05	< .001
6	39.04	0.11	0.13	< .001
7	44.95	0.13	0.14	< .001
8	48.14	0.14	0.15	< .001
9	39.45	0.13	0.14	< .001
10	48.40	0.17	0.18	< .001

Note: A series of correlation contrast tests (Meng, Rosenthal, & Rubin 1992) were used to evaluate the extent to which one correlation was significantly different from another. This procedure yields a *Z* test for the significance of the difference between two correlation coefficients using a Fisher *z* transformation for the correlation between each predictor and outcome variable and the correlation between the two predictors.

Source: Author's analysis based on 2012 data requested from the Progress Monitoring and Reporting Network of the Florida Department of Education.

Table A2. 2×2 contingency table of sample size for 2011/12 Florida Comprehensive Assessment Test (FCAT) success probability with 85 percent cutscore and 2011 FCAT, by 2012 FCAT 2.0 performance and grade

2011/12 FCAT success probability	2012 FCAT 2.0			2011 FCAT 2.0	2012 FCAT 2.0		
	Does not meet standards	Meets standards	Total		Does not meet standards	Meets standards	Total
Grade 4							
Fall				At risk	18,672	2,933	21,605
At risk	32,991	21,559	54,550	Not at risk	13,192	48,254	61,446
Not at risk	1,259	32,379	33,638	Total	31,864	51,187	83,051
Total	34,250	53,938	88,188				
Winter							
At risk	33,506	21,849	55,355				
Not at risk	1,281	32,014	33,295				
Total	34,787	53,863	88,650				
Spring							
At risk	32,635	15,605	48,240				
Not at risk	2,400	38,132	40,532				
Total	35,035	53,737	88,772				
Grade 5							
Fall				At risk	20,644	3,900	24,544
At risk	33,615	21,070	54,685	Not at risk	11,538	45,986	57,524
Not at risk	1,402	31,350	32,752	Total	32,182	49,886	82,068
Total	35,017	52,420	87,437				
Winter							
At risk	34,064	21,586	55,650				
Not at risk	1,351	30,357	31,708				
Total	35,415	51,943	87,358				
Spring							
At risk	33,855	16,941	50,796				
Not at risk	2,212	36,341	38,553				
Total	36,067	53,282	89,349				
Grade 6							
Fall				At risk	22,493	3,882	26,375
At risk	36,259	20,742	57,001	Not at risk	9,765	34,795	44,560
Not at risk	851	22,634	23,485	Total	32,258	38,677	70,935
Total	37,110	43,376	80,486				
Winter							
At risk	36,306	19,475	55,781				
Not at risk	1,052	24,373	25,425				
Total	37,358	43,848	81,206				
Spring							
At risk	35,898	18,809	54,707				
Not at risk	934	23,374	24,308				
Total	36,832	42,183	79,015				

(continued)

Table A2. 2×2 contingency table of sample size for 2011/12 Florida Comprehensive Assessment Test (FCAT) success probability with 85 percent cutscore and 2011 FCAT, by 2012 FCAT 2.0 performance and grade *(continued)*

2011/12 FCAT success probability	2012 FCAT 2.0		
	Does not meet standards	Meets standards	Total
Grade 7			
Fall			
At risk	35,410	19,919	55,329
Not at risk	848	23,532	24,380
Total	36,258	43,451	79,709
Winter			
At risk	35,734	19,301	55,035
Not at risk	986	24,665	25,651
Total	36,720	43,966	80,686
Spring			
At risk	34,713	17,307	52,020
Not at risk	1,096	24,533	25,629
Total	35,809	41,840	77,649
Grade 8			
Fall			
At risk	37,105	20,030	57,135
Not at risk	880	21,038	21,918
Total	37,985	41,068	79,053
Winter			
At risk	37,468	19,446	56,914
Not at risk	1,010	22,490	23,500
Total	38,478	41,936	80,414
Spring			
At risk	36,279	17,236	53,515
Not at risk	1,160	22,223	23,383
Total	37,439	39,459	76,898
Grade 9			
Fall			
At risk	38,792	19,692	58,484
Not at risk	624	16,557	17,181
Total	39,416	36,249	75,665
Winter			
At risk	39,308	18,949	58,257
Not at risk	749	17,162	17,911
Total	40,057	36,111	76,168
Spring			
At risk	35,206	15,453	50,659
Not at risk	657	14,162	14,819
Total	35,863	29,615	65,478

2011 FCAT 2.0	2012 FCAT 2.0		
	Does not meet standards	Meets standards	Total
At risk	24,396	4,381	28,777
Not at risk	8,521	35,685	44,206
Total	32,917	40,066	72,983
At risk	24,212	3,807	28,019
Not at risk	10,610	33,984	44,594
Total	34,822	37,791	72,613
At risk	29,682	6,616	36,298
Not at risk	5,472	23,461	28,933
Total	35,154	30,077	65,231

(continued)

Table A2. 2×2 contingency table of sample size for 2011/12 Florida Comprehensive Assessment Test (FCAT) success probability with 85 percent cutscore and 2011 FCAT, by 2012 FCAT 2.0 performance and grade *(continued)*

2011/12 FCAT success probability	2012 FCAT 2.0		
	Does not meet standards	Meets standards	Total
Grade 10			
Fall			
At risk	40,536	20,874	61,410
Not at risk	332	13,105	13,437
Total	40,868	33,979	74,847
Winter			
At risk	40,987	18,945	59,932
Not at risk	584	15,031	15,615
Total	41,571	33,976	75,547
Spring			
At risk	35,424	17,014	52,438
Not at risk	283	9,843	10,126
Total	35,707	26,857	62,564

2011 FCAT 2.0	2012 FCAT 2.0		
	Does not meet standards	Meets standards	Total
At risk	33,762	8,327	42,089
Not at risk	4,429	22,954	27,383
Total	38,191	31,281	69,472

FCAT is Florida Comprehensive Assessment Test.

Source: Author's analysis based on 2012 data requested from the Progress Monitoring and Reporting Network of the Florida Department of Education.

Table A3. 2×2 contingency table of sample size for 2011/12 Florida Comprehensive Assessment Test (FCAT) success probability with 70 percent cutscore and 2011 FCAT, by 2012 FCAT performance and grade

2011/12 FCAT success probability	2012 FCAT 2.0		
	Does not meet standards	Meets standards	Total
Grade 4			
Fall			
At risk	30,625	12,687	43,312
Not at risk	3,625	41,251	44,876
Total	34,250	53,938	88,188
Winter			
At risk	30,986	13,082	44,068
Not at risk	3,801	40,781	44,582
Total	34,787	53,863	88,650
Spring			
At risk	29,432	8,615	38,047
Not at risk	5,603	45,122	50,725
Total	35,035	53,737	88,772
Grade 5			
Fall			
At risk	31,443	13,466	44,909
Not at risk	3,574	38,954	42,528
Total	35,017	52,420	87,437
Winter			
At risk	31,691	13,108	44,799
Not at risk	3,724	38,835	42,559
Total	35,415	51,943	87,358
Spring			
At risk	30,958	10,051	41,009
Not at risk	5,109	43,231	48,340
Total	36,067	53,282	89,349
Grade 6			
Fall			
At risk	34,596	13,712	48,308
Not at risk	2,514	29,664	32,178
Total	37,110	43,376	80,486
Winter			
At risk	34,475	12,581	47,056
Not at risk	2,883	31,267	34,150
Total	37,358	43,848	81,206
Spring			
At risk	34,078	12,056	46,134
Not at risk	2,754	30,127	32,881
Total	36,832	42,183	79,015

2011 FCAT 2.0	2012 FCAT 2.0		
	Does not meet standards	Meets standards	Total
At risk	18,672	2,933	21,605
Not at risk	13,192	48,254	61,446
Total	31,864	51,187	83,051
At risk	20,644	3,900	24,544
Not at risk	11,538	45,986	57,524
Total	32,182	49,886	82,068
At risk	22,493	3,882	26,375
Not at risk	9,765	34,795	44,560
Total	32,258	38,677	70,935

(continued)

Table A3. 2×2 contingency table of sample size for 2011/12 Florida Comprehensive Assessment Test (FCAT) success probability with 70 percent cutscore and 2011 FCAT, by 2012 FCAT performance and grade *(continued)*

2011/12 FCAT success probability	2012 FCAT 2.0		
	Does not meet standards	Meets standards	Total
Grade 7			
Fall			
At risk	33,982	13,545	47,527
Not at risk	2,276	29,906	32,182
Total	36,258	43,451	79,709
Winter			
At risk	33,808	12,182	45,990
Not at risk	2,912	31,784	34,696
Total	36,720	43,966	80,686
Spring			
At risk	32,815	11,211	44,026
Not at risk	2,994	30,629	33,623
Total	35,809	41,840	77,649
Grade 8			
Fall			
At risk	35,370	13,349	48,719
Not at risk	2,615	27,719	30,334
Total	37,985	41,068	79,053
Winter			
At risk	35,496	12,593	48,089
Not at risk	2,982	29,343	32,325
Total	38,478	41,936	80,414
Spring			
At risk	34,029	10,561	44,590
Not at risk	3,410	28,898	32,308
Total	37,439	39,459	76,898
Grade 9			
Fall			
At risk	37,411	13,489	50,900
Not at risk	2,005	22,760	24,765
Total	39,416	36,249	75,665
Winter			
At risk	37,705	12,825	50,530
Not at risk	2,352	23,286	25,638
Total	40,057	36,111	76,168
Spring			
At risk	33,739	10,179	43,918
Not at risk	2,124	19,436	21,560
Total	35,863	29,615	65,478

2011 FCAT 2.0	2012 FCAT 2.0		
	Does not meet standards	Meets standards	Total
At risk	24,396	4,381	28,777
Not at risk	8,521	35,685	44,206
Total	32,917	40,066	72,983
At risk	24,212	3,807	28,019
Not at risk	10,610	33,984	44,594
Total	34,822	37,791	72,613
At risk	29,682	6,616	36,298
Not at risk	5,472	23,461	28,933
Total	35,154	30,077	65,231

(continued)

Table A3. 2×2 contingency table of sample size for 2011/12 Florida Comprehensive Assessment Test (FCAT) success probability with 70 percent cutscore and 2011 FCAT, by 2012 FCAT performance and grade *(continued)*

2011/12 FCAT success probability	2012 FCAT 2.0		
	Does not meet standards	Meets standards	Total
Grade 10			
Fall			
At risk	39,638	15,357	54,995
Not at risk	1,230	18,622	19,852
Total	40,868	33,979	74,847
Winter			
At risk	39,582	13,277	52,859
Not at risk	1,989	20,699	22,688
Total	41,571	33,976	75,547
Spring			
At risk	34,701	12,840	47,541
Not at risk	1,006	14,017	15,023
Total	35,707	26,857	62,564

2011 FCAT 2.0	2012 FCAT 2.0		
	Does not meet standards	Meets standards	Total
At risk	33,762	8,327	42,089
Not at risk	4,429	22,954	27,383
Total	38,191	31,281	69,472

FCAT is Florida Comprehensive Assessment Test.

Source: Author's analysis based on 2012 data requested from the Progress Monitoring and Reporting Network of the Florida Department of Education.

Notes

1. This study uses data for students in grades 4–10. Grade 3 students are not included in the analyses because they have no 2011 FCAT 2.0 score. Students in grades 11 and 12 are not included in the analyses because they take the FCAT only if they do not meet grade-level standards in grade 10.
2. Hispanic includes Latino and Black includes African American.
3. For a detailed description of FAIR and its psychometrics, see the technical manual (Florida Department of Education 2009a).

References

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

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. Retrieved August 1, 2012, from www.eric.ed.gov/ERICWebPortal/detail?accno=EJ742190

Florida Department of Education. (2009a). *FAIR 3–12 Technical Manual*. Tallahassee, FL: Author. Retrieved August 1, 2012, from www.fcrr.org/FAIR/Technical manual - 3–12-FINAL_2012.pdf

Florida Department of Education. (2009b). *Florida Assessments for Instruction in Reading (FAIR)*. Tallahassee, FL: Author.

Florida Department of Education. (2011a). *Florida Comprehensive Assessment Test 2.0: Florida statewide assessments 2011 yearbook*. Tallahassee, FL: Author.

Florida Department of Education. (2011b). *Understanding FCAT 2.0 reports*. Tallahassee, FL: Author.

Foorman, B. R., & Petscher, Y. (2010a). Summary of the predictive relationship between the FAIR and the FCAT in grades 3–10. Tallahassee, FL: Florida Center for Reading Research.

Foorman, B. R., & Petscher, Y. (2010b). The unique role of the FAIR Broad Screen in predicting FCAT Reading Comprehension. Tallahassee, FL: Florida Center for Reading Research.

Hamilton, L., Halverson, R., Jackson, S., Mandinach, E., Supovitz, J., & Wayman, J. (2009). *Using student achievement data to support instructional decision making* (NCEE 2009–4067). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Educational Evaluation and Regional Assistance. Retrieved August 1, 2012, from www.eric.ed.gov/ERICWebPortal/detail?accno=ED506645

Jenkins, J. R. (2003, December). *Candidate measures for screening at-risk students*. Paper presented at the National Research Center on Learning Disabilities' Responsiveness-to-Intervention Symposium, Kansas City, MO. Retrieved August 1, 2012, from www.nrcld.org/symposium2003/jenkins/index.html

Meng, X., Rosenthal, R., & Rubin, D. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin, 111*(1), 172–175.

Perie, M., Marion, S., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice, 28*(3), 5–13. Retrieved August 1, 2012, from www.eric.ed.gov/ERICWebPortal/detail?accno=EJ853799

Petscher, Y., & Foorman, B. R. (2011). *Summary of the predictive relationship between the FAIR and the FCAT in grades 3–10: 2010–2011*. Tallahassee, FL: Florida Center for Reading Research.

Petscher, Y., Kim, Y., & Foorman, B. (2011). The importance of predictive power in early screening assessments: Implications for placement in the RTI framework. *Assessment for Effective Instruction, 36*(3), 158–166.