Stated Briefly

Can scores on an interim high school reading assessment accurately predict low performance on college readiness exams?

Sharon Koon
Yaacov Petscher
Florida State University

This study examines whether scores from an interim reading assessment in grade 9, the Florida Assessments for Instruction in Reading—Florida Standards, can be used to identify students who may score below the college readiness benchmark on the Preliminary SAT/National Merit Scholarship Qualifying Test and ACT Plan in grade 10. Using scores on an existing interim reading assessment in an early warning system could enable districts to identify at-risk students without an additional testing burden, time away from instruction, or added cost.

Why this study?

District leaders in two Florida school districts partnered with Regional Educational Laboratory Southeast to study the extent to which scores on Florida's interim reading assessment in grade 9, the Florida Assessments for Instruction in Reading—Florida Standards (FAIR-FS), could identify students who may

This brief summarizes the findings of Koon, S., & Petscher, Y. (2016). *Can scores on an interim high school reading assessment accurately predict low performance on college readiness exams?* (REL 2016–124). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. That report is available at http://ies.ed.gov/ncee/edlabs/projects/project.asp?ProjectID=4464.

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

score below the college readiness benchmark on the Preliminary SAT/National Merit Scholarship Qualifying Test (PSAT/NMSQT) or the ACT Plan in grade 10. During the 2013/14 school year the participating school districts sought to develop an early warning system to identify students at risk of low performance on college readiness measures in grade 11 or 12 in order to support them with remedial coursework prior to high school graduation. If specific skills can be identified as those most likely to indicate college readiness, early identification of deficiencies in those skills could enable teachers and schools to provide targeted intervention.

This study used the FAIR-FS to examine the relationship of reading comprehension, decoding, and language to college-ready performance on the PSAT/NMSQT or ACT Plan. The FAIR-FS is a screening and diagnostic assessment system aligned with the Language Arts Florida Standards. Grade 10 students who achieve a PSAT/NMSQT score of 42 in critical reading, 44 in math, and 42 in writing meet college and career readiness benchmarks (College Board, 2013). Students who achieve an ACT Plan score of 15 in English, 18 in reading, 20 in science, and 19 in math are considered ready for college (ACT, 2013).

FAIR-FS data were obtained for five high schools in district 1 and two high schools in district 2. PSAT/NMSQT data were obtained from district 1, and ACT Plan data were obtained from district 2.

Two research questions guided the study:

- How do scores on the FAIR-FS in grade 9 predict performance on the PSAT/NMSQT in grade 10?
- How do scores on the FAIR-FS in grade 9 predict performance on the ACT Plan in grade 10?

Classification and regression tree (CART) analyses, which provide results in an easy-to-interpret "tree" format, were used to classify students as at risk or not at risk of low performance on the PSAT/NMSQT or ACT Plan college readiness measures based on a set of "if-then" statements related to FAIR-FS scores (Berk, 2008; Breiman, Friedman, Olshen, & Stone, 1984; Lewis, 2000). While multiple measures of classification accuracy are reported, the current study used sensitivity values to evaluate the CART results (see the appendix for details). In the current study sensitivity refers to the percentage of students identified as at risk on FAIR-FS among all students who did not meet the college readiness benchmark of interest.

What the study found

The findings indicate that FAIR-FS scores can predict performance on college readiness assessments with acceptable sensitivity (at least 80 percent correct; Piasta, Petscher, & Justice, 2012).

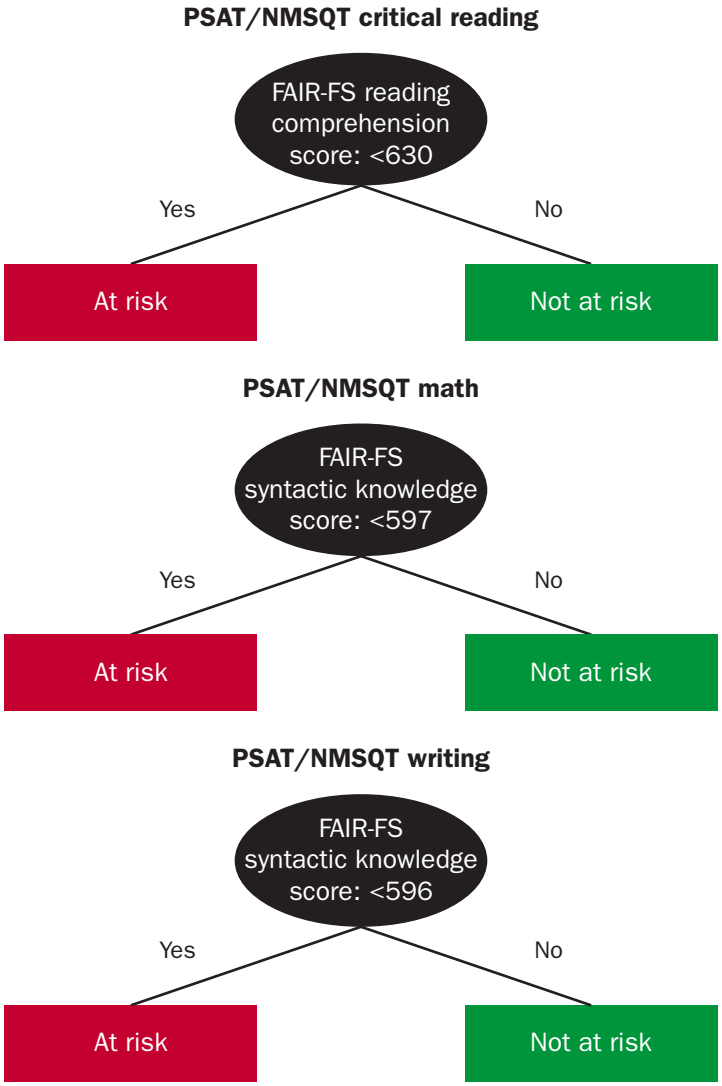
FAIR-FS reading comprehension scores predict PSAT/NMSQT critical reading performance with 89 percent sensitivity, and FAIR-FS syntactic knowledge scores predict PSAT/NMSQT math performance with 81 percent sensitivity and writing performance with 84 percent sensitivity

The CART analyses identified only one decision rule in each subject area for classifying students as at risk or not at risk of not reaching the college readiness benchmark on the PSAT/NMSQT (figure 1). Each rule is derived using the criterion specified followed by a "yes" or "no" answer.

The overall classification accuracy rate of the decision rules (the percentage of students who were correctly identified as either meeting or not meeting college readiness benchmarks on the PSAT/NMSQT) was 75–83 percent when tested with a validation sample of students (table 1). The sensitivity rate (the percentage of students correctly identified as at risk) was 81–89 percent, above the minimum standard of 80 percent. The specificity rate (the percentage of students correctly identified as not at risk) was much lower (66–80 percent), which affected the overall classification accuracy rate. Positive predictive power

rates generally exceeded negative predictive power rates. This study considered it more important to judge the prediction models based on sensitivity rates because the intention is to identify students at risk of low performance on the PSAT/NMSQT.

Figure 1. Classification and regression tree model decision rules for classifying Florida district 1 students as at risk or not at risk of not reaching the college readiness benchmark on the PSAT/NMSQT based on FAIR-FS score



PSAT/NMSQT is Preliminary SAT/National Merit Scholarship Qualifying Test. FAIR-FS is Florida Assessments for Instruction in Reading—Florida Standards.

Note: FAIR-FS scores range from 200 to 800, with a mean of 500 and a standard deviation of 100.

Source: Authors' analysis of data from the Florida Center for Reading Research and district 1.

Table 1. Preliminary SAT/National Merit Scholarship Qualifying Test classification and regression tree analysis results for Florida district 1 students (percent)

PSAT/NMSQT subject area	Sensitivity rate	Specificity rate	Positive predictive power rate	Negative predictive power rate	Overall classification accuracy rate
Critical reading	89	69	79	83	80
Math	81	66	79	70	75
Writing	84	80	89	72	83

PSAT/NMSQT is Preliminary SAT/National Merit Scholarship Qualifying Test.

Note: *n* = 210.

Source: Authors' analysis of data from the Florida Center for Reading Research and district 1.

FAIR-FS syntactic knowledge scores predict ACT Plan English performance with 81 percent sensitivity, reading performance with 84 percent sensitivity, and science performance with 96 percent sensitivity, and FAIR-FS reading comprehension scores predict ACT Plan math performance with 83 percent sensitivity

As with the PSAT/NMSQT results, the CART analyses identified only one rule in each subject area for classifying students as at risk or not at risk of not reaching the college readiness benchmark on the ACT Plan (figure 2).

The overall classification accuracy rate of the decision rules was 71–81 percent when tested with a validation sample of students (table 2). The sensitivity rate was 81–96 percent, above the minimum standard of 80 percent, while specificity rates were much lower (41–76 percent). Positive predictive power rates ranged from 59–89 percent, and negative predictive power rates ranged from 44–85 percent.

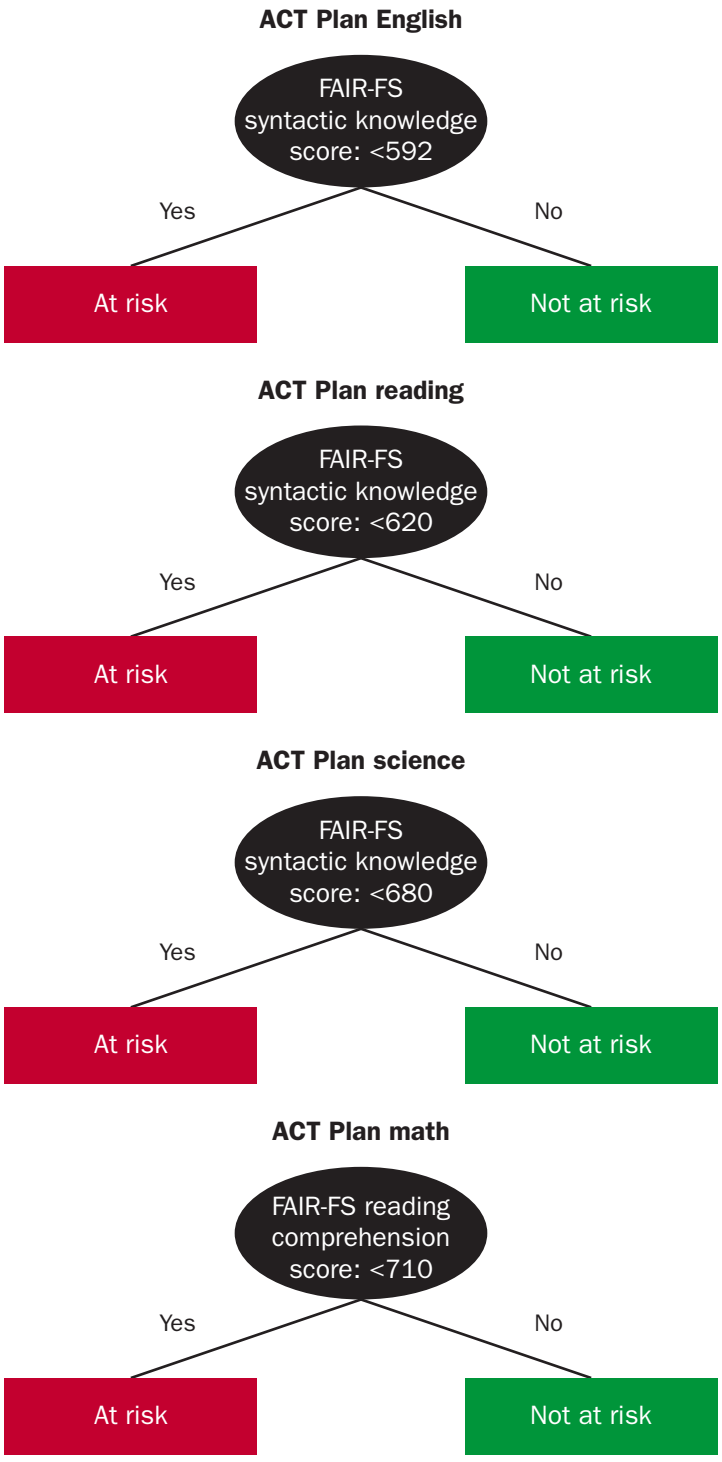
Table 2. ACT Plan classification and regression tree analysis results for Florida district 2 students (percent)

ACT Plan subject area	Sensitivity rate	Specificity rate	Positive predictive power rate	Negative predictive power rate	Overall classification accuracy rate
English	81	65	59	85	71
Reading	84	76	89	67	81
Science	96	48	79	85	80
Math	83	41	81	44	73

Note: *n* = 70.

Source: Authors' analysis of data from the Florida Center for Reading Research and district 2.

Figure 2. Classification and regression tree model decision rules for classifying Florida district 2 students as at risk or not at risk of not reaching the college readiness benchmark on the ACT Plan based on FAIR-FS score



FAIR-FS is Florida Assessments for Instruction in Reading–Florida Standards.

Note: FAIR-FS scores range from 200 to 800, with a mean of 500 and a standard deviation of 100.

Source: Authors' analysis of data from the Florida Center for Reading Research and district 2.

Implications of the study findings

This study provides preliminary evidence that FAIR-FS scores in grade 9 could be used to create an early warning system to identify students whose PSAT/NMSQT or ACT Plan performance in grade 10 may be below the college readiness benchmark. Using FAIR-FS scores in an early warning system could enable districts to identify these students without additional testing burden, time away from instruction, or added cost.

Limitations of the study

The study has two important limitations. First, the analyses are based on a single year's results, which does not allow for confirming the stability of the results or performing a longitudinal analysis. This limitation is especially relevant to FAIR-FS scores, which were drawn from the assessment's field-test sample. Second, the findings for each college readiness assessment are based on a single school district: district 1 for the PSAT/NMSQT and district 2 for the ACT Plan. The analyses could be replicated statewide to verify the stability of the models and the generalizability of the results to the larger Florida student population.

Both the College Board and ACT have implemented new assessments, replacing the versions of the PSAT/NMSQT and ACT Plan assessments used in this study (ACT, 2015; College Board, 2015). New analyses would be required to determine whether the study results remain valid with the new assessments.

PSAT/NMSQT and ACT Plan performance serves as a proxy measure of college readiness due to the complexity of identifying all the factors that determine whether a student is truly ready for success in college. District decisionmakers should carefully consider other issues, such as the implications of over- and under-identification, student access to the core curriculum and other typical activities at the school, and the school process for determining when a student may successfully exit an intervention.

Appendix. Measures of classification accuracy

Traditional measures of classification accuracy can be derived from a 2x2 classification table that provides counts of individuals in four categories (Schatschneider, Petscher, & Williams, 2008). In the current study students are categorized based on their score on an interim assessment, the Florida Assessments for Instruction in Reading—Florida Standards (FAIR-FS), and an outcome assessment, the Preliminary SAT/National Merit Scholarship Qualifying Test or the ACT Plan (table A1).

In addition to the overall classification accuracy rate (proportion of students correctly identified as meeting or not meeting the college readiness benchmark on the outcome assessment), four standard measures of classification accuracy were applied to determine how accurately the analysis identifies students at risk. The first measure, sensitivity, is the proportion of students identified as at risk on the interim assessment among all students who fail the outcome assessment—the number of true positives—divided by the sum of true positives and false negatives (A/[A+C]). The second measure, specificity, is the proportion of students identified as not at risk among all students who pass the outcome assessment—or the number of true negatives divided by the sum of true negatives and false positives (D/[D+B]). The third measure, positive predictive power, is the proportion of students who fail the outcome assessment among all students who are identified as at risk on the interim assessment—or the number of true positives divided by the sum of true positives and false positives (A/[A+B]). The fourth measure, negative predictive power, is the proportion of students who pass the outcome assessment among all students who are identified as not at risk on the interim assessment—or the number of true negatives divided by the sum of false negatives and true negatives (D/[C+D]).

Researchers have proposed different threshold values for sensitivity (the measure used in this study to judge the performance of the prediction models); many look for levels of at least .80 (Piasta et al., 2012), with some recommending at least .90 (Compton, Fuchs, Fuchs, & Bryant, 2006; Jenkins, Hudson, & Johnson, 2007). Based on these guidelines, a sensitivity value of .80 or higher (80 percent or higher when expressed in percentages) was used in the current study.

Table A1. Sample 2×2 classification table

	Outcome assessment (Preliminary SAT/National Merit Scholarship Qualifying Test or ACT Plan)	
Interim assessment (Florida Assessments for Instruction in Reading–Florida Standards)	Fail	Pass
At risk	A: True positive	B: False positive
Not at risk	C: False negative	D: True negative

Source: Authors' illustration.

References

ACT. (2013). *PLAN technical manual*. Iowa City, IA: Author. Retrieved January 14, 2014, from http://www.act.org/plan/pdf/PlanTechnicalManual.pdf.

ACT. (2015). *ACT Plan overview*. Retrieved December 1, 2015, from http://www.act.org/products/k-12-act-plan/.

Berk, R. A. (2008). *Statistical learning from a regression perspective*. New York, NY: Springer.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Belmont, CA: Wadsworth International Group.

College Board. (2013). *Understanding 2013 PSAT/NMSQT scores*. New York, NY: Author. Retrieved January 14, 2014, from http://professionals.collegeboard.com/profdownload/understanding-psat-nmsqt-scores.pdf.

College Board. (2015). *Redesigned SAT*. New York, NY: Author. Retrieved April 9, 2015, from https://www.collegeboard.org/delivering-opportunity/sat/redesign.

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

Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for service delivery in an RTI framework: Candidate measures. *School Psychology Review*, *36*(4), 582–601.

Lewis, R. J. (2000). An introduction to classification and regression tree (CART) analysis. In *Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California* (pp. 1–14). Retrieved July 9, 2015, from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.95.4103&rep=rep1&type=pdf.

Piasta, S. B., Petscher, Y., & Justice, L. M. (2012). How many letters should preschoolers in public programs know? The diagnostic efficiency of various preschool letter-naming benchmarks for predicting first-grade literacy achievement. *Journal of Educational Psychology*, *104*(4), 945–958.

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 L. Justice & C. Vukelic (Eds.), *Every moment counts: Achieving excellence in preschool language and literacy instruction* (pp. 304–317). New York, NY: Guilford Press.

REL 2016–134

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.

May 2016

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 the 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.

This REL report is in the public domain. While permission to reprint this publication is not necessary, it should be cited as:

Koon, S., & Petscher, Y. (2016). *Stated Briefly: Can scores on an interim high school reading assessment accurately predict low performance on college readiness exams?* (REL 2016–134). 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.

The Regional Educational Laboratory Program produces 7 types of reports