



REFLECTIONS
ON
METHODOLOGY

MARCH 2019

Using Predictive Analytics to Combine Indicators of Third-Grade Reading Proficiency

By *Kristin E. Porter*

This post is one in a series highlighting MDRC's methodological work. Contributors discuss the refinement and practical use of research methods being employed across our organization.

Elementary schools often assess whether students are on track to read at grade level at future points much like high schools assess whether students are on track to graduate — relying on a set of indicators, or predictors, such as literacy screening tests. Each test produces a useful composite score and scores for subsections of the tests. Educators can then identify students with a low likelihood of future reading success and recommend interventions to help them improve. We wondered whether the *combined* scores and subscores from all the reading assessments administered over the years could provide more accurate information.¹ While it is difficult for educators to sift through this array of scores to understand the implications, predictive analytics, including machine learning, can provide substantial value here. Machine learning refers to a variety of algorithms that mine the data to help us better determine the relationships between the test scores and the outcome, such as third-grade reading proficiency.

OUR APPROACH IN A LARGE SCHOOL DISTRICT

MDRC studied a large school district that administers the DIBELS Next portfolio of screening tests to first-, second-, and third-grade students in the fall, winter, and spring. The DIBELS Next assessments are “universal” literacy screeners (administered to all children at a given grade level) that identify potential reading problems. The district also administers an end-of-grade English language arts (EOG ELA) test to third-graders, which is used to assess reading proficiency levels for state accountability reports. A baseline ELA test is given to third-graders at the beginning of the school year.

We used DIBELS scores and baseline ELA scores for students in third grade in 2014-2015 to build predictive models to estimate the likelihood of reading proficiency for the next year's third-graders. We then evaluated the predictions based on the true reading proficiency outcomes as measured by the students' actual scores on the 2015-2016 EOG ELA test. We estimated students' risks at two time points: (1) after the second-grade spring DIBELS screening test and (2) after the third-grade fall DIBELS screening test. Using both time points allows us to assess the value of combining both current and past DIBELS tests.

For each DIBELS screening test, [the publisher provides translations](#) from raw scores to grade-level reading categories. According to the DIBELS guidelines: (1) Students who score *above benchmark* on any given assessment have a very high likelihood of future reading success. We considered this a prediction that these students would score at a proficient level on the third-grade EOG ELA test. (2) Students who score *well below benchmark* have a very low likelihood of future reading success. We considered this a prediction that these students would *not* score at a proficient level on the third-grade EOG ELA test. (3) Students who score *below benchmark* but not *well below benchmark* have a medium likelihood of future reading success (40 percent to 60 percent). We considered this a prediction of not reaching proficiency on the third-grade EOG ELA test. (Repeating the analysis with the alternative assumption did not notably change the story when comparing the benchmark predictions with predictive analytics.)

NEW YORK
200 Vesey Street, 23rd Flr.
New York, NY 10281
Tel: 212 532 3200

OAKLAND
475 14th Street
Suite 750
Oakland, CA 94612
Tel: 510 663 6372

WASHINGTON, DC
1990 M Street, NW
Suite 340
Washington, DC 20036

LOS ANGELES
11965 Venice Boulevard
Suite 402
Los Angeles, CA 90066

www.mdrc.org



¹ See Biancarosa et al., “Constructing Subscores That Add Validity: A Case Study of Identifying Students at Risk,” *Education and Psychological Measurement* 79, no. 1 (2019), for research on subsection scores.

FINDINGS

We evaluated three sets of predictions: those based on DIBELS benchmarks; those produced by estimating the relationship between the district's DIBELS scores and EOG ELA performance for the 2014-2015 cohort; and our model derived from machine learning, using the same threshold to predict proficiency. (Using the district scores in the second comparison, rather than the national sample that DIBELS uses to produce its risk categories, allows us to disentangle the benefits of using local data from the benefits of more complex modeling.) In the table below, the true positive rate represents the percentage of at-risk students (those who did not score proficiently on the third-grade EOG ELA) correctly identified as at risk by a given method; the false positive rate represents the percentage of students incorrectly identified as at risk among those who did score proficiently.

District identification of at-risk students in the 2015-2016 cohort

MODEL	TRUE POSITIVE RATE	FALSE POSITIVE RATE
Using DIBELS guidelines		
At risk at end of second grade	50%	5%
At risk at start of third grade	54%	5%
Based on most recent composite score for 2014-2015 cohort		
At risk at end of second grade	55%	13%
At risk at start of third grade	57%	5%
Combining all available scores with machine learning for 2014-2015 cohort		
At risk at end of second grade	73%	18%
At risk at start of third grade	78%	14%

The machine learning model greatly increased the true positive rate. As the true positive rate rises, the false positive rate often rises as well; this is a typical trade-off. In the case of reading interventions, this is not a terrible thing. When there are false positives, it means that more students may receive intervention than truly need it, but they may benefit from it, and more students who need intervention will actually get it.

The more complex modeling using machine learning incorporates all available DIBELS composite and subtest scores (both levels and continuous scores) for the district's own students: 104 measures of student performance at the end of second grade and 137 measures in the fall of third grade (after the first round of third-grade assessments). This method also outperformed a simple regression model based on composite scores, which resulted in a 57 percent true positive rate and a 10 percent false positive rate.

CONCLUSION

Predictive analytics is able to do what teachers cannot — combine and weight many pieces of information in a way that is optimal for making predictions. When applied to multiple screening tests, the method allows educators to more accurately detect at-risk students who can benefit from intervention and reduce the substantial costs of providing an intervention to students who would succeed without it.

For the results presented here, we used just one cohort of students to develop our models and just one cohort to validate the models' performance. It is important to repeat the analysis with more cohorts for which there are consistent data over time. It would be valuable to include first-grade DIBELS scores, if the data become available; measures of students' attendance, tardiness, or other data could also improve the models' accuracy.