Predicting public school students at risk for standardized testing failure

Jennifer Brite, MPH[1], Christina Pollari[1], Jasmine Abdelnabi[1], Ragheed Al-dulaimi[1], William Stumbo[3], Levi Waldron, PhD[1]

1. CUNY School of Public Health
2. CUNY Institute for Demographic Research
3. Xerox Company

Abstract

Although standardized testing is taking a more central role in school districts nationwide, models to predict performance are lacking, even as more data is gathered about students. This study will use data from Massachusetts Comprehensive Assessment System, a standardized test taken by all public school 10th graders in Massachusetts. We will use principal component analysis to group demographic, personality trait, academic courses, and career plans variables taken from a student questionnaire and other sources. We will use ROC analysis to identify the components with the highest predictive power, and will then build a model using 2013 data. Finally, we will evaluate our model by using it to predict scores in previous years.

Background

Nationwide, more public school districts are adopting standardized tests as a means of determining promotion and graduation (1). Predicting students at risk for standardized test failure is therefore increasingly important. Several studies have shown demographic characteristics, such as race and ethnicity, gender, and socioeconomic status, are associated with standardized test performance (1-4). Moreover, personality traits, such as self-efficacy, openness and conscientious, and academic motivation (5, 6), have been found to be predictors of general academic success. However, robust predictive models for standardized test score performance are sorely lacking at time when school districts are collecting more student data than ever before (7). Predictive models are needed that incorporate demographic and personality traits, as well as other variables such as coursework prior to the test, interest in particular academic subjects, and plans after high school.

In 1993, Massachusetts enacted its landmark Massachusetts Education Reform Act (MERA) (8). The Massachusetts Comprehensive Assessment System (MCAS) was designed to meet the requirements of MERA. This law specifies that MCAS must test all public school students in Massachusetts, measure performance based on the Massachusetts Curriculum Framework learning standards, and report on the performance of individual students, schools, and districts. All students educated with public funds in the tested grade are required to participate. More than 20% of students failed the 10th grade math test.(9)

MCAS is comprised of 3 subject areas: English language arts (ELA), mathematics, and science and technology/engineering. Brennan et al. assessed the equitability effects of the 1998's 8th grade MCAS across the three academic subjects and according to student characteristics (10). They found MCAS math results for African American and Latino/Latina students were significantly lower than their white counterparts. This finding, however, was not found among Asians. They also found that females' scores were significantly lower on MCAS math and science than their male counterparts; however, girls scored higher in English by similar margins.

The objective of this study is to build a model to predict a student's score on the MCAS that uses demographic, personality, and other available variables regarding a student's academic life and future plans.

Methods

Data

Our study sample includes all 10th grade students in Massachusetts who took the Massachusetts Comprehensive Assessment System (MCAS) from 2011 to 2013. Each year, approximately 70,000 students complete the exam. The data include: students' demographics (gender, race, and first language); responses and scores for each of three tests: English language arts (ELA), math, and science; and a set of questionnaire responses related to coursework completed, interest in math and science, and plans after high school.

Outcome variables

We will analyze test results in three ways. First, we will categorize performance levels based on the ranges defined by the Board of Elementary and Secondary Education: Advanced (260-280), Proficient (240-258), Needs Improvement (220-238) and Failing (200-218). Second, we will create a “pass/fail” binary outcome based on the categorical ranges; pass is any score over 220, while fail is 200. Finally, we will analyze the results as a continuous outcome using overall test results.

Statistical analysis

We will construct, via principal component analysis for ordinal and nominal variables, a set of components that compress the most variability of the answers to the student questionnaire from 2013. The analysis of correlation previous to the PCA analysis will be done using the polychoric correlation coefficients. The PCA analysis will be done with the aim of then condensing the variables in the analysis into a more manageable set variables—called principal components—that account for most of the variability of the original variables. The PCA analysis will be done using the CATPCA procedure in SPSS, which transforms categorical variables into quantities using optimal scaling, resulting in optimal principal components. The CATPCA procedure also imputes missing values with the mode, after internal correlation analysis is performed. The resulting principal components are each a linear combination of the original variables.

A series of cross-validated (CV) receiver operating characteristic (ROC) curves will be used to determine which groups of variables in different models could accurately identify students at risk of being below the various cutoff points for all test scores. A ten-fold CV method with 10 random selections of the folds will used to draw the ROC plots, a method similar to that implemented in Robin et al (11). Models will be evaluated by mean area under the curve (AUC), with a greater AUC representing a better fit with the data. Model 1 will include demographic characteristics such as race, gender, and use of English as the first language. Subsequent models will include each of the principal components identified via PCA analysis with a lambda greater than 1. The final model will consist of demographic characteristics as well as all principal components with an AUC of 0.70 or greater, determined by ROC analysis.

Finally, we will evaluate our final model by using it to predict student MCAS scores for 2011 and 2012.

The analyses will be performed using SPSS 21 and SAS 9.3.

Preliminary Results

In a sub-sample of the 2013 MCAS made up of 7,677 students, principal component analysis identified seven components with a lambda value above 1. Those components accounted for 58.8% of the total variability found in the student’s questionnaire. The first component was made up of questions related to the frequency the students spent in their own investigations, discussing findings, doing computer simulations, using specialized software and digital tools to explain the results of their science enquiries. As a result the first component was labeled *frequency of time spent in science labs*. The questions that added most of the variability in the second component were the ones related to plans after graduating from high school, going right into college, and the willingness to follow a career in sciences. The second component was labeled *studying science in College, work in science*. The third component relates to the source of any job training, if their plan after high school was not going right in to college, and given the label *job training after graduating from high school*. The fourth component is weighted most heavily for questions related to the use of digital resources by both students and teachers, inside and outside class settings; therefore that component is called u*se of digital resources (class and at home)*. The fifth component captures whether the student is currently taking science courses. The sixth and seventh components both measured whether a computer was used for test taking and classroom assignments respectively during the school year. None of these principal components have not yet been validated via ROC curves.

Preliminary results suggest science skills seemed to help students across the board. Spending time in science labs, wanting to study science in college or work in science, and currently taking a science course were significantly protective against being at risk for failure not only in math and science but also in ELA. Use of technology seemed to take a more conflicting role. Taking a test on a computer and completing assignments on a computer was protective against being at risk for math and science. However, use of other digital resources both inside and outside the classroom was not significantly associated with being at risk in ELA, math, or science. Those planning to go into job training after high school rather than college had significantly higher odds of failing in math and science, but not ELA.

References

1. Brennan RT, Kim J, Wenz-Gross M, Siperstein GN. The relative equitability of high-stakes testing versus teacher-assigned grades: An analysis of the Massachusetts Comprehensive Assessment System (MCAS). Harvard educational review 2001;71(2):173-217.

2. Duckworth AL, Seligman ME. Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. Journal of educational psychology 2006;98(1):198.

3. Jencks C, Phillips M. The black-white test score gap: Brookings Institution Press; 2011.

4. Hunter RC, Bartee R. The achievement gap issues of competition, class, and race. Education and Urban Society 2003;35(2):151-160.

5. Caprara GV, Vecchione M, Alessandri G, Gerbino M, Barbaranelli C. The contribution of personality traits and self-efficacy beliefs to academic achievement: A longitudinal study. British Journal of Educational Psychology 2011;81(1):78-96.

6. Zimmerman BJ, Bandura A, Martinez-Pons M. Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. American educational research journal 1992;29(3):663-676.

7. Tamhane A, Ikbal S, Sengupta B, Duggirala M, Appleton J. Predicting student risks through longitudinal analysis. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining; 2014: ACM; 2014. p. 1544-1552.

8. MIE. Education Reform in Massachusetts - 20 year anniversary report. Boston, Massachusetts; 2013.

9. Feng M, Heffernan NT. Informing Teachers Live about about Student Learning: Reporting in the Assistment System. Tech., Inst., Cognition and Learning 2005;3:00-00.

10. Brennan RT, Kim J, Wenz-Gross M, Siperstein GN. The relative equitability of high-stakes testing versus teacher-assigned grades: An analysis of the Massachusetts Comprehensive Assessment System (MCAS). Harvard Educational Review 2001;71(2):173-216.

11. Robin X, Turck N, Hainard A, Lisacek F, Sanchez J-C, Müller M. Bioinformatics for protein biomarker panel classification: what is needed to bring biomarker panels into in vitro diagnostics? Expert Review of Proteomics 2009;6(6):675-689.