THE STATE EDUCATION DEPARTMENT / THE UNIVERSITY OF THE STATE OF NEW YORK / ALBANY, NY 12234

Methodological Summary
The New York State (NYS) Testing Program
Grades 3-8 English Language Arts (ELA) and Mathematics
Common Core Tests

2013 External Benchmark Studies Summary

With the adoption of the NYS P-12 Common Core Learning Standards (CCLS) in ELA/Literacy and Mathematics, the Board of Regents signaled a shift in both instruction and assessment. In Spring 2013, NYS administered the first set of tests designed to assess student progress on the Common Core State Standards (CCSS) and whether students are on track to succeed in college and careers. The new tests require a standard setting process to establish the cut scores that determine levels of proficiency. During the standard setting process, it is common to use *external benchmark data* – that is, student performance on other assessments that measure or predict similar outcomes. To inform New York State Education Department (NYSED) leaders, policy makers, expert Standard Setting panelists,[1] and other stakeholders, NYSED approached The College Board to analyze NYS student performance on nationally-recognized assessments of college readiness: the SAT and the PSAT/NMSQT (Pre-SAT/National Merit Scholarship Qualifying Test).

Several national studies have examined the predictive relationships between The College Board's PSAT/NMSQT[2] and SAT[3], respectively, and first-year grade point average. For NYS, The College Board ultimately employed an approach similar to that of the national studies, but adapted to better reflect course-taking patterns of college students from New York.

The College Board first replicated existing research to identify the scores on the PSAT/NMSQT and SAT that predicted success in first-year of college for the sample of students entering four-year colleges and universities in Fall 2010. Scores on the PSAT/NMSQT and SAT math sections were studied to see how well they predicted performance in first year college math courses. Similarly, scores on the PSAT/NMSQT and SAT Critical Reading sections were studied to see how well they predicted performance in courses requiring significant reading, including English (excluding writing), business and communications, history, humanities, and social sciences. The PSAT/NMSQT and SAT writing sections were examined in terms of writing intensive courses in these same areas plus English composition.

---

[1] Information on Standard Setting, including details on standard setting panelists, please visit: http://www.regents.nysed.gov/meetings/2013Meetings/July2013/StandardSetting.pdf

[2] Proctor, T. P., Wyatt, J. N., & Wiley, A. (2010). PSAT/NMSQT Indicators of College Readiness. New York: The College Board. Retrieved from: http://research.collegeboard.org/rr2010-4

[3] Wyatt, J. N., Remigio, M. J., & Camara, W. J. (2012). SAT content area benchmarks: An analysis conducted by Research and Development for the National Assessment Governing Board (NAGB). New York: The College Board. Retrieved from: http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/statistical-relationships/sat-content-area-benchmarks.pdf.

The analyses used the following combinations of probability and course grade to determine college readiness. These criteria, established using college performance of a nationally representative sample of four-year college students, reflect current expectations for college readiness benchmarks for NYS students. ***The probabilities and expected grades differ in ELA and Math because, generally speaking, it is much easier to obtain a higher grade in first-year credit-bearing ELA courses than in first-year credit-bearing math courses.***

English Language Arts: What score on the PSAT/NMSQT and SAT (in Critical Reading and Writing) yields a 75% probability of attaining at least a grade of B- or its equivalent in those applicable courses?

Math: What score on the PSAT/NMSQT and SAT in Math yields a 60% probability of attaining at least a grade of C+ or its equivalent in those applicable courses?

Results
The results of The College Board studies provided meaningful information on how NYS students perform upon entering college. The results of the studies were shared with the panelists who set the performance standards for college readiness on the new Common Core grades 3-8 Math and ELA Tests administered in the 2012-13 school year.

Minimum cut scores
The table below summarizes the college readiness benchmarks on the PSAT/NMSQT and SAT established by the college readiness criteria above.

	75% probability of B- or higher	60% probability of C+ or higher
SAT – Critical Reading	**560**	380
SAT - Writing	**530**	360
SAT - Math	710	**540**
PSAT/NMSQT – Critical Reading	**49**	33
PSAT/NMSQT – Writing	**45**	29
PSAT/NMSQT - Math	65	**45**

Impact of college-readiness benchmarks
Once the above criteria were established, additional analyses identified the percent of NYS students who would be judged college ready if expected to meet the nationally-determined PSAT/NMSQT and SAT benchmarks. The criteria were applied to the SAT scores and 10th grade PSAT/NMSQT scores of a cohort of NYS students expected to have graduated from high school in the summer of 2010. Only 25% of NYS students scored high enough on SAT-Critical Reading to meet the criteria for college readiness in ELA. Only 36% of NYS students scored high enough on SAT-Math to meet the criteria for college readiness in Math. At the June 2013 Grades 3-8 ELA and Math Standard Setting, The College Board analysis helped to set the context for

considering college-ready cut scores on the NYS Common Core Grades 3-8 ELA and Math Tests by establishing a benchmark on another test with which the standard setting panelists were already familiar.

Brian F. Patterson & Wayne J. Camara, The College Board

INTRODUCTION

New York State Education Department (NYSED) approached The College Board to conduct a series of analyses to establish the relationship, for New York State (NYS) students, between their PSAT/NMSQT and SAT scores and their performance in the first year of college. The ultimate goal was to provide meaningful contextual information to the standard setting panels that were tasked with setting cut scores for college readiness on exams undergoing revision.

ESTIMATING COLLEGE BOARD TEST CUT SCORES

1) Wyatt, Remigio, and Camara (2012) estimated section-specific readiness indicators (i.e., cut scores) on the SAT critical reading (SAT-CR), mathematics (SAT-M), and writing (SAT-W) sections and this work mirrored their basic procedure, with the exception that this work used the PSAT/NMSQT (P/N) and SAT for the 2010 National SAT Validity Study cohort, while Wyatt, et al. (2013) used SAT data for 2009.

   a. The PSAT/NMSQT national samples consisted of: 113,122 students (from 113 institutions) for critical reading; 119,101 students (from 116 institutions) for writing; and 86,419 students (from 113 institutions) for mathematics. The SAT samples consisted of: 173,020 students (from 116 institutions) for critical reading; 182,254 students (from 118 institutions) for writing; and 131,299 students (from 117 institutions) for mathematics. For further information on the 2010 cohort of the National SAT Validity Study, of which a subset of up to 118 institutions (of a possible 160) were used in these analyses, see Patterson and Mattern (2013). Those 42 institutions for which course work had not been previously categorized for ready analysis were excluded.

2) In order to estimate external benchmarks for PSAT/NMSQT and SAT for three sections each, the first-year college GPA in relevant disciplines were used. In particular:

   a. SAT-M and PSAT-M were linked to all math courses;

   b. SAT-CR and PSAT-CR were linked to all "reading intensive" courses; i.e., business and communications, English (excluding writing), history, humanities, and social science; and

   c. SAT-W and PSAT-W were linked to all "writing intensive" courses; i.e., business and communications, English, history, humanities, and social science.

3) The sample was restricted to students expecting to graduate high school in 2010 with complete data on the predictor (e.g., 10th Grade PSAT math) and outcome of interest (e.g., earned a valid grade in at least one math course in the first year of college).

   a. Certain colleges may have been dropped despite having students with complete data in two cases: (1) all students earned mean course grades either above or below the selected GPA level for the discipline; or (b) there was quasi-complete or complete separation of the predictor for the dichotomized outcome. In either case, the maximum likelihood parameter estimates do not exist and hence cannot be used in the creation of external benchmarks.

4) Separate binary logistic regression models were estimated and the resulting parameter estimates were used via inverse prediction to identify the institution-specific cut score on the predictor of interest.

   a. Binary logistic regressions were estimated separately for each college with the dependent variable being the dichotomous indicator of the student having met or exceeded the selected GPA level.

   b. If the institution-specific cut score was outside the operational scale, that institution was dropped.

   c. The institution-specific cut scores are weighted by sample size and averaged to land on what we call the meta-analyzed logistic-regression-based cut score.

IDENTIFYING NEW YORK STATE PERCENTILES FOR STANDARD SETTING

Given the cut scores derived from the nationally representative set of four-year colleges and universities, the proportions of the national and New York State samples meeting or exceeding those cuts were estimated.

5) Based on a representativeness tolerance of +/- 5% for the gender and racial / ethnic identity groups, it became apparent that the PSAT/NMSQT sample under-represented White students (34.8% before re-sampling, see table below), so the PSAT/NMSQT group was re-sampled to better represent the Spring 2012 8th Grade NYSED assessment population.

Subgroup	NY State Grade 8 2012 Population (%)	All NY SAT (%)	PSAT/NMSQT	
			All NY (%)	Resampled (%)
		(n = 160,518)	(n = 105,564)	(n = 76,000)
Female	49.9	53.2	52.2	50.8
Male	50.1	46.8	47.8	49.2
African American	16.9	14.7	21.1	18.5
Hispanic	19.2	15.5	24.2	21.0
White	52.6	54.4	34.8	48.1
Other Ethnic Groups	11.3	13.6	16.2	12.4
Missing	0.0	1.8	3.7	0.0

6) For the New York State group of SAT and 10th grade PSAT/NMSQT examinees expecting to graduate in the Spring of 2010, the percentile p associated with the estimated cut score c is given by:

$$p = \sum_{k < c} (\% \text{ of students earning scores of } k) + \frac{1}{2} \cdot (\% \text{ of students earning scores of } c)$$

7) Based on a percentile of p, 100% - p of students met or exceeded the pth percentile rank (i.e., test cut score of c) and it was these percentages of students at or above the cut score that were the focus of the external benchmarks materials provided to standard setting panelists.

REFERENCES

Wyatt, J. N., Remigio, M. J., & Camara, W. J. (2013). SAT Subject Area Readiness Indicators: Reading, Writing, and STEM. New York: The College Board. Retrieved from: http://research.collegeboard.org/rn2013-1.pdf.

Patterson, B. F. & Mattern, K. D. (2013). Validity of the SAT for Predicting First-Year Grades: 2010 SAT Validity Sample. New York: The College Board. Retrieved from: http://research.collegeboard.org/sr2013-2.