MAP Growth College Readiness Benchmarks: An Addendum with Preliminary Results Keyed on the SAT

Yeow Meng Thum, Ph.D.

June 21, 2017

Suggested Citation:

Thum Y. M. (2017). *MAP Growth College Readiness Benchmarks: An Addendum with Preliminary Results keyed on the SAT*®. NWEA Research Report. Portland, OR: NWEA

Disclaimer: This report is the product of research conducted by the Northwest Evaluation Association. Neither the report nor its findings are sponsored or endorsed by ACT, Inc. or The College Board.

MAP Growth College Readiness Benchmarks: An Addendum with Preliminary Results Keyed on the SAT*

Yeow Meng Thum

NWEA

Portland, Oregon

June 21, 2017

As elementary and high-school educators are increasingly encouraged to prepare their students for college and life, early indicators of whether students are on-track for college success will no doubt play a stronger role. Early indicators of "college preparedness" were not readily available until 2015 when NWEA™ released college readiness benchmarks on NWEA MAP® Growth™ mathematics and reading scales for 5th through 9th grade students (Thum & Matta, 2015b). These benchmarks were based on the longitudinal MAP Growth test-score histories of all students in several districts, regardless of whether they had or had not taken the ACT® before they graduated from high school.

Results in Table 5 of Thum and Matta (2015b) indicated that elementary and high-school students who performed at the 60th percentile or above on either MAP Growth math or reading are generally on-track in their preparation for college success by virtue of the fact that they are likely to obtain an ACT score of 22 or higher in high-school. Using these benchmarks, a parent or teacher may now gauge a student's level of preparedness during the upper elementary and middle school grades, and thus have some time to course-correct appropriately. This addendum provides functionally equivalent backmapped benchmarks on MAP Growth mathematics and reading scales which are keyed on students' likely high school SAT® results instead.

Procedure

The ACT informs secondary school educator and college admissions officers about the preparedness of their students for post-secondary education (Clough & Montgomery, 2015). The SAT Suite of Assessments measure skills and knowledge that are essential ingredients for college and career readiness and success. Like the ACT, it is employed as a college-entrance test for college-bound high school students. For further details, see The College Board Research Report by Shaw, Marini, Beard, Shmueli, Young, and Ng (2016).

In this study, we report a set of college preparedness benchmark estimates for use with MAP reading and

*We thank Don Draper for leading an immense data collection effort and Branin Bowe for data preparation. Rebecca Moore and Dr. Michael Finger offered very helpful suggestions Please direct all correspondence to Y. M. Thum, at yeow.meng@nwea.org.

Table 1:
Descriptive statistics for the five districts in the study, MAP Growth Reading.

	Count				Took SAT	SCI	
District	Schools	Students	Tests	Cohorts	(%)	Mean	SD
1	7	1226	7416	4	51	67	5
2	58	15450	128531	4	79	50	15
3	71	8194	43091	4	20	75	13
4	6	1683	12683	4	24	20	10
5	8	2887	18934	4	58	40	17
Summary	150	29440	210655	20	46.4⋆	50.5⋆	12.0⋆

Note: ⋆ Average.

mathematics tests from grades 5 through 9, by exploiting the strong relationship between MAP Growth and the SAT in student longitudinal test score histories. A fairly diverse group of five small to medium-sized school districts from across the country participated in the study. In all, over 210,655 test events from 150 schools that serve a total of 29,440 students are analyzed.

Table 1 provides the counts of students, schools, MAP Growth reading test events, and cohorts per district. Districts ranged in size from large (District 2, $N = 15,450$) to small (District 1, $N = 1226$). A total of 29,440 students from 150 schools, in 20 cohorts were used to estimate the MAP Growth reading and mathematics college preparedness benchmarks. Additionally, the proportion of students taking the SAT varied across districts with 75% of students taking the SAT in District 3 and 17% of students taking the SAT in District 2. On average, across all 5 districts, 34% of the graduating class completed the SAT during high school. The counts for MAP Growth for mathematics are highly comparable.

As Table 1 shows, participating districts have average SCIs[1] from 20 to 75. These schools cluster around the national average of 50.5. The SCI standard deviation of 12.0 indicates that study schools represent the middle 44% of public schools in the US on this school poverty composite index. Because the study schools are less extreme in terms of the SCI spectrum than public schools in the US, results should be regarded cautiously as preliminary.

The approach pools the longitudinal test score histories from multiple time-adjacent age-cohorts of students within a district. Typically, a major cohort of students with MAP Growth scores and SAT scores provides the information needed to link MAP Growth and SAT results (age-cohort 7 in Table 2). MAP Growth data from several additional cohorts (age-cohorts 6, 8, and 9 in Table 2) support the description of the trends in MAP Growth for the district.

Correlations between MAP Growth and SAT mathematics and reading scores provide the necessary infor-

[1]The SCI, for *School Challenge Index*, is a school-level indicator of how public schools compare in terms of the challenges and opportunities they operate under as reflected by an array of factors they do not control (NWEA, 2015). This composite indicator is keyed on the proportion of students who are eligible for a free-and-reduced-priced lunch program in a school. Thus, it generally taps the collective economic circumstance of its students but it also offers a broader view of the economic circumstances they experience, as seen through a relevant set of socio-demographic, organizational, and educational policy programming factors. The SCI ranges from 1 to 99, with higher values for schools serving lesser-privileged student bodies. It has an average of 50 and a standard deviation of 21 among public schools in the US.

Table 2:
Illustrative cohort structure for a district.

		2010		2011		2012		2013		2014		2015		2016	
	Grade	Fa	Sp	Fa	Sp	Fa	Sp	Fa	Sp	Fa	Sp	Fa	Sp	Fa	Sp
SAT	12	1	1	2	2	3	3	4	4	5	5	6	6	7	7
	11	2	2	3	3	4	4	5	5	6	6	7	7	8	8
	10	3	3	4	4	5	5	6	6	7	7	8	8	9	9
MAP	9	4	4	5	5	6	6	7	7	8	8	9	9	10	10
	8	5	5	6	6	7	7	8	8	9	9	10	10	11	11
	7	6	6	7	7	8	8	9	9	10	10	11	11	12	12
	6	7	7	8	8	9	9	10	10	11	11	12	12	13	13
	5	8	8	9	9	10	10	11	11	12	12	13	13	14	14
	4	9	9	10	10	11	11	12	12	13	13	14	14	15	15

Table 3:
Sample size and correlation between SAT and MAP Growth scores for age-cohort 7 from District 2.

		5		6		7		8		9		
Subject		Fa	Sp	Fa	Sp	Fa	Sp	Fa	Sp	Fa	Sp	SAT
Mathematics	N	1956	1968	1990	2000	2022	2035	843	819	741	451	2209
	Corr.	.79	.80	.80	.80	.82	.83	.80	.74	.69	.69	
Reading	N	1949	1972	1991	1997	2021	2035	948	875	908	813	2218
	Corr.	.74	.74	.76	.75	.77	.76	.73	.70	.77	.78	

mation for linking one scale to another. Table 3 shows the sample correlations of seniors from one cohort in District 2 who have taken the SAT and MAP scores they received when they attended middle and high school. The sample correlations are moderately high (from 0.69 to 0.83) but they do not show a trend towards higher values in the higher grades as one expects.

Critical to the benchmarking effort, we recognize that participating districts vary widely, from 20% to 78%, in the proportion of high school students who take the SAT during their junior or senior years. It seems reasonable that whether or not a student takes the SAT is not a random outcome but that it reflects some degree of self-selection on the part of schools, parents, or students. It also suggests that growth modeling of longitudinal data is superior to bivariate or cross-sectional analyses. With longitudinal data, shared information across the grades and terms is maximized and estimates of the links between early MAP scores and SAT are improved.

To mitigate estimation bias from self-selection in taking the SAT, the modeling strategy of Thum and Matta (2015b), introduced in Thum (2011) and Thum and Matta (2015a), is employed. Missing data on the

Table 4:
Normative stringency and classification accuracy of pooled benchmarks for MAP Mathematics and Reading tests.

		Mathematics, SAT=530					Reading, SAT=480				
Grade	Term	Benchmark	SE	Pct	TPR	FPR	Benchmark	SE	Pct	TPR	FPR
5	Fall	216.67	0.07	65	0.60	0.28	202.76	0.06	43	0.69	0.30
5	Spring	224.65	0.07	59	0.62	0.27	208.72	0.06	42	0.69	0.28
6	Fall	224.75	0.07	68	0.62	0.26	208.83	0.06	45	0.71	0.26
6	Spring	231.52	0.07	66	0.63	0.25	213.77	0.05	45	0.73	0.26
7	Fall	231.27	0.07	69	0.63	0.24	213.68	0.05	49	0.73	0.24
7	Spring	236.81	0.07	68	0.63	0.24	217.60	0.05	50	0.75	0.24
8	Fall	236.23	0.07	71	0.63	0.23	217.29	0.05	49	0.76	0.25
8	Spring	240.55	0.08	70	0.63	0.23	220.20	0.05	50	0.74	0.23
9	Fall	239.62	0.08	70	0.64	0.25	219.68	0.05	49	0.74	0.23
9	Spring	242.72	0.08	69	0.64	0.25	221.57	0.05	50	0.74	0.24

Note: SE = Std. Error Pct = Percentile TPR = True Positive Rate FPR = False Positive Rate

SAT are treated as *missing not at random*, or MNAR (Rubin & Little, 2002), per the shared-parameter model in Thum and Matta (2015b, Appendix A). To set the MAP Growth college readiness benchmark for mathematics and reading at each grade and term, given an SAT benchmark score of 530 or 480[2], respectively, and a 0.5 probability[3] that a student will opt to take a college test, are examined. Finally, predicted district means and variance-covariances are then pooled over districts, weighted by the amounts of data employed in each districts, using a multiple sample covariance structure analysis. The pooled results are the basis for benchmark estimations.

Results

College preparedness benchmark estimates for each grade, term and subject are listed in Table 4, along with their standard errors, 2015 normative percentiles, true positive, and false positive rates. The standard errors indicate that the benchmarks are well-estimated. As expected, benchmarks increase with grade level and they also appear to be more stringent, as they approach high school. For example, the mathematics benchmark (corresponding to the SAT score of 530) for fall term of grade 7 is 231.27 which corresponds with the 69^{th} percentile based on the 2015 NWEA MAP national norms.

The quality of benchmark are gauged by considering two classification accuracy standards. The first is the *true positive rate* (TPR) and the second is the *false positive rate* (FPR)[4]. The true positive rate is the proportion of students who are considered college ready based on a given MAP Growth score for a grade and term and who are actually college ready (based on a score of 530 for math (or 480 for reading) or better on the SAT during high school), among all those students who scored a 530 (or 480) or better on the SAT.

[2]These new SAT college and career readiness benchmark scores represent a 75% likelihood of a student achieving at least a "C" grade in a first-semester, credit-bearing college course in a related subject.

[3]This probability value merely reflects the situation in which the reader admits ignorance about how likely a student will be taking the SAT in high school.

[4]In the wider research literature on decision quality, false positive rate is the Type I error rate and false negative rate is the Type II error rate.

Table 5:
Area under the curve (AUC) for MAP college readines benchmarks.

	5		6		7		8		9	
Subject	Fa	Sp	Fa	Sp	Fa	Sp	Fa	Sp	Fa	Sp
Mathematics	.795	.806	.813	.821	.823	.828	.827	.830	.824	.825
Reading	.818	.830	.843	.853	.862	.868	.871	.871	.869	.863

The false positive rate is the proportion of students who are considered college ready based on a given MAP Growth score but do not score a 530 (or 480) or higher on the SAT among all those students who did not score a 530 (480) or better on the SAT. Locating a MAP Growth score that balances high true positive rate with low false positive rate is key to determining a benchmark for each term and grade.

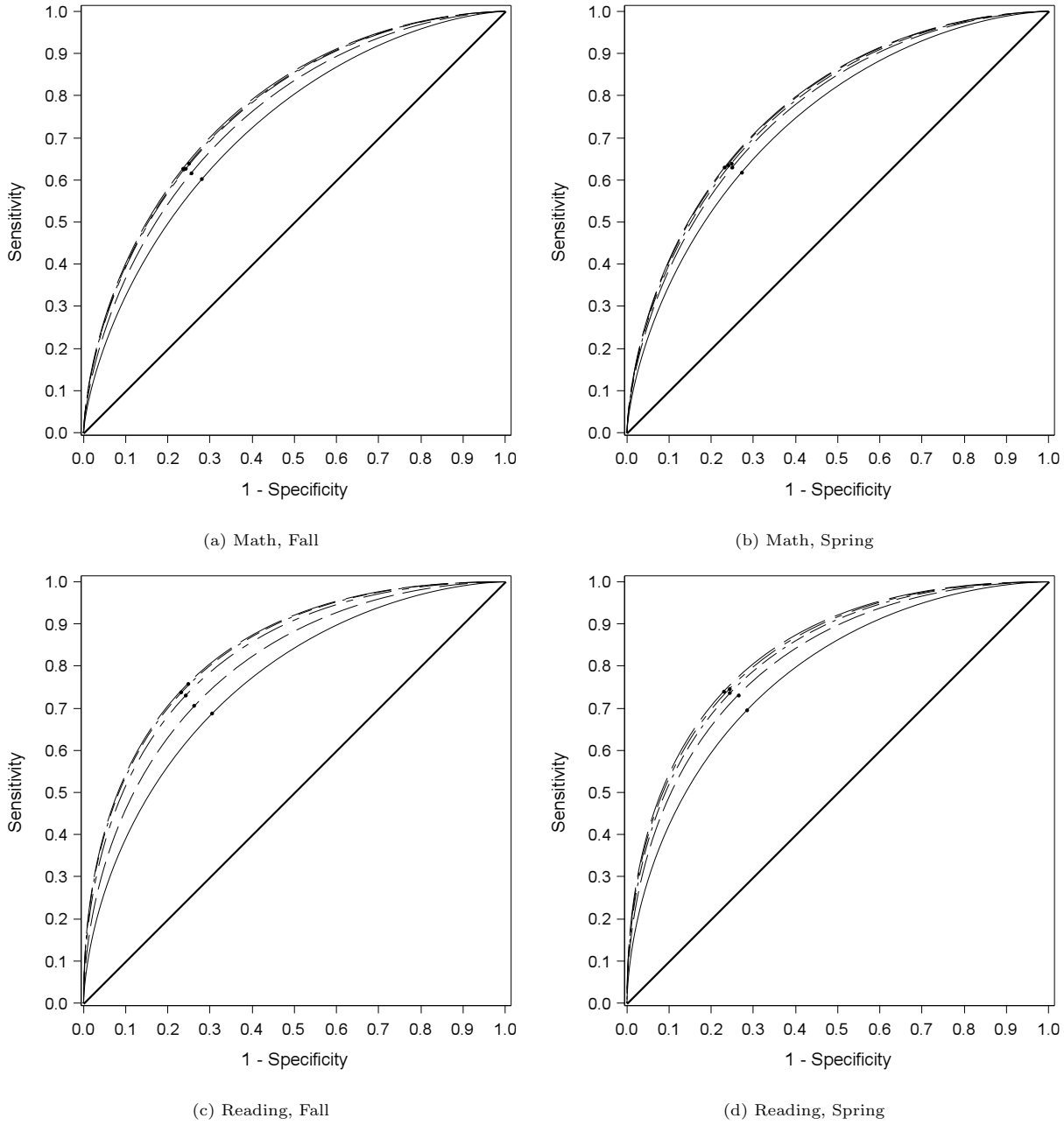
Using the benchmarks obtained correctly classify students who are college ready 63% of the time while falsely classifying students who are not college ready as college ready only 24% of the time. The true positive rate for mathematics ranges from 0.60 (fall, grade 5) to 0.64 (spring, grade 9; fall, grade 9). The false positive for mathematics benchmarks ranges from 0.23 (grade 5, fall and spring) to 0.28 (fall, grade 5). The true positive rates for reading benchmarks range from 0.69 to 0.76 and the false positive rates range from 0.24 to 0.30.

True positive rate are plotted against the false positive rate for all possible MAP benchmarks for a given grade and term generates an ROC curve[5]. The area under the ROC curve (AUC) is a well-known measure of predictive power, where a straight 45-degree line represents no predictive power (or 50-50 chance) and an AUC of 1 is perfect prediction[6]. As is shown in Table 5, AUC estimates for all benchmarks are relatively high, indicating predictions are well calibrated (or reliable).

Figure 1 provides four graphs, each one containing the ROC curves for grades 5 through 9 for different subjects (mathematics and reading) and terms (fall and spring) on assuming the SAT benchmark of 503/480. Graph (a) illustrates the ROC curves for fall term mathematics, Graph (b) illustrates spring term mathematics, Graph (c) illustrates fall term reading, and Graph (d) represents spring term reading. In each graph, the grade 5 scores are the shallowest (smaller AUC) and are represented by the solid curved line. As the grade level increases, the AUC becomes larger and the curve approaches the top-left corner of the plot. This indicates, as expected, that benchmarks at 8^{th} and 9^{th} grade are more predictive of college readiness than benchmarks in 5^{th} grade. On the whole, use of the benchmarks leads to highly accurate predictions. The point plotted on each curve is the MAP benchmark with the given true positive rate and false positive rate in Table 4.

[5]A "receiver operating characteristic" or ROC curve is a graphical device representing the trade-off between the hit and false alarm rates of a binary decision rule; here, the proposed benchmark for college readiness. See, *e.g.*, Swets, Dawes, and Monahan, (2000).

[6]The AUC is also called a "concordance" statistic.



(a) Math, Fall

(b) Math, Spring

(c) Reading, Fall

(d) Reading, Spring

Figure 1: ROC Plots for Mathematics and Reading Benchmarks, Grades 5 - 9

Conclusion

College readiness benchmarks for fall and spring terms of 5th through 9th graders on MAP Growth mathematics and reading, which are statistically anchored on SAT's college readiness cut-scores of 530 and 480 respectively, were obtained from examining more than 210,000 test events from 29,440 4th to 12th graders from 150 schools in 5 districts across the US. The study finds that middle and high school students in grades 5 through 9 are likely to be on-track in the preparation for college if they performed between the 60th to 70th percentiles, or above, in mathematics or between the 40th to 50th percentiles, or above, in reading. These MAP Growth benchmark estimates are anchored on the recommended SAT college and career readiness benchmark score of 530 for mathematics and 480 for Evidence-Based Reading and Writing. In terms of national percentile ranking, a SAT mathematics score of 530 registers at the 61st percentile and a SAT Evidence-Based Reading and Writing score ranks at the 41st percentile, according to The College Board (2016). It is important to note that, as with our previous benchmarks referencing the ACT, the results are generally applicable to all middle and high school students, rather than to those who are expected to take the ACT or the SAT before they leave high school.

Using these MAP Growth college preparedness benchmarks, about 60 to 75 students out of 100 who meet or exceed the benchmarks are correctly classified as on-track for college and only 25 to 30 students out of 100 of those students who are not on-track are misclassified (see Figure 1 above). The true positive classification rates are found to be sufficiently high, suggesting that, when such benchmarks are used, educators and parents can be confident that students are accurately identified as being college ready, or not. At the same time, the false positive classification error rates appear sufficiently low so that students requiring assistance to get back on track to being college ready are also accurately identified. These benchmarks are selected with the view that higher misclassification rates of off-track students are more costly than the misclassification of on-track student for all stakeholders. Mistaking a student to be on-track when he is not would mean missing an early opportunity to intervene and returning him to the path of college preparedness. Given the available information, these classification accuracy rates are robust for the SCI spectrum of schools represented in the study.

Selected References

Clough, S., & Montgomery, S. (2015). *How ACT Assessments Align with State College and Career Readiness Standards* (White Paper). Retrieved from http://www.discoveractaspire.org/pdf/ACT_Alignment-White-Paper.pdf.

Rubin, D. B. & Little, R. J. A. (2002). *Statistical analysis with missing data (2nd ed.)*. New York: Wiley.

Shaw, E., Marini, J. P., Beard, J., Shmueli, D., Young, Y., & Ng, H. (2016). *The Redesigned SAT Pilot Predictive Validity Study: A First Look* (Research report 2016-1). New York: The College Board.

The College Board (2016). *SAT Understanding Scores 2016*. Retrieved from http://collegereadiness.collegeboard.org/pdf/understanding-sat-scores-2016.pdf.

Thum, Y. M. (2011). *Measuring Student Growth and Achievement against College Readiness Benchmarks and the ACT.* Technical White Paper. Grand Rapids, MI: National Charter School Institute.

Thum Y. M., & Hauser, C. H. (2015). *NWEA 2015 MAP Norms for Student and School Achievement Status and Growth*. NWEA Research Report. Portland, OR: NWEA.

Thum, Y. M., & Matta, T. (2015a). *Predicting College Readiness from Interim Assessment Results: Selection Modeling for Longitudinal Data*. Paper presented at the Annual Meetings of the American Educational Research Association, April, Chicago.

Thum Y. M., & Matta, T. (2015b). *MAP College Readiness Benchmarks: A Research Brief*. NWEA Research Report. Portland, OR: NWEA

Author

Yeow Meng Thum, PhD., Senior Research Fellow at NWEA, is a research methodologist with substantive interests surrounding the use of testing results in formulating public educational policy. His methodological research focuses on multivariate, multilevel models for behavioral and educational data, with a strong emphasis on methods for describing growth and change.

Suggested Citation:

Thum, Y. M. (2017). *MAP College Readiness Benchmarks: An Addendum with Preliminary Results Keyed on the SAT*. NWEA Research Report. Portland, OR: NWEA

Table 6:
Predicted benchmark for MAP Mathematics by grade/term and percentile ranks for HS seniors expected to meet the Mathematics SAT=530 benchmark.

Grade		5		6		7		8		9	
Term		Fa	Sp	Fa	Sp	Fa	Sp	Fa	Sp	Fa	Sp
Benchmark		216.67	224.65	224.75	231.51	231.27	236.81	236.23	240.55	239.62	242.72
SD		10.52	10.44	10.24	10.40	10.38	10.80	10.91	11.59	11.87	12.84
Pct		65	59	68	66	69	68	71	70	70	69
Conditional Percentile Rank	10	203	211	212	218	218	223	222	226	224	226
	15	206	214	214	221	221	226	225	229	227	229
	20	208	216	216	223	223	228	227	231	230	232
	25	210	218	218	224	224	230	229	233	232	234
	30	211	219	219	226	226	231	231	234	233	236
	35	213	221	221	228	227	233	232	236	235	238
	40	214	222	222	229	229	234	233	238	237	239
	45	215	223	223	230	230	235	235	239	238	241
	50	217	225	225	232	231	237	236	241	240	243
	55	218	226	226	233	233	238	238	242	241	244
	60	219	227	227	234	234	240	239	243	243	246
	65	221	229	229	236	235	241	240	245	244	248
	70	222	230	230	237	237	242	242	247	246	249
	75	224	232	232	239	238	244	244	248	248	251
	80	226	233	233	240	240	246	245	250	250	254
	85	228	235	235	242	242	248	248	253	252	256
	90	230	238	238	245	245	251	250	255	255	259

Note: Pct = Percentile.

Appendix

Table 7:
Predicted benchmark for MAP Reading by grade/term and percentile ranks for HS seniors expected to meet the Evidence-Based Reading and Writing SAT=480 benchmark.

	Grade	5		6		7		8		9	
	Term	Fa	Sp	Fa	Sp	Fa	Sp	Fa	Sp	Fa	Sp
	Benchmark	202.76	208.72	208.83	213.77	213.68	217.60	217.29	220.20	219.68	221.57
	SD	11.06	10.18	9.68	9.00	8.74	8.36	8.40	8.40	8.78	9.23
	Pct	43	42	45	45	49	50	49	50	49	50
Conditional Percentile Rank	10	189	196	196	202	202	207	207	209	208	210
	15	191	198	199	204	205	209	209	211	211	212
	20	193	200	201	206	206	211	210	213	212	214
	25	195	202	202	208	208	212	212	215	214	215
	30	197	203	204	209	209	213	213	216	215	217
	35	199	205	205	210	210	214	214	217	216	218
	40	200	206	206	211	211	215	215	218	217	219
	45	201	207	208	213	213	217	216	219	219	220
	50	203	209	209	214	214	218	217	220	220	222
	55	204	210	210	215	215	219	218	221	221	223
	60	206	211	211	216	216	220	219	222	222	224
	65	207	213	213	217	217	221	221	223	223	225
	70	209	214	214	218	218	222	222	225	224	226
	75	210	216	215	220	220	223	223	226	226	228
	80	212	217	217	221	221	225	224	227	227	229
	85	214	219	219	223	223	226	226	229	229	231
	90	217	222	221	225	225	228	228	231	231	233

Note: Pct = Percentile.

Table 8:
Predicted probability of an observed MAP Mathematics score meeting or exceeding selected MAP benchmarks by grade/term for HS seniors expected to meet the Mathematics SAT=530 benchmark.

	Grade	5		6		7		8		9	
	Term	Fa	Sp	Fa	Sp	Fa	Sp	Fa	Sp	Fa	Sp
	Benchmark	216.67	224.65	224.75	231.51	231.27	236.81	236.23	240.55	239.62	242.72
	SD	10.52	10.44	10.24	10.40	10.38	10.80	10.91	11.59	11.87	12.84
	Pct	65	59	68	66	69	68	71	70	70	69
Observed Score	208	1	1	1	1	1	1	1	1	1	1
	210	2	1	1	1	1	1	1	1	1	1
	212	7	1	1	1	1	1	1	1	1	1
	214	20	1	1	1	1	1	1	1	1	1
	216	42	1	1	1	1	1	1	1	1	1
	218	66	2	2	1	1	1	1	1	1	1
	220	85	7	7	1	1	1	1	1	1	1
	222	95	20	20	1	1	1	1	1	1	1
	224	99	42	41	1	1	1	1	1	1	1
	226	99	66	65	4	5	1	1	1	1	1
	228	99	85	85	14	15	1	1	1	1	1
	230	99	95	95	32	35	2	3	1	1	1
	232	99	99	99	56	59	7	9	1	1	1
	234	99	99	99	78	80	19	24	2	4	1
	236	99	99	99	92	93	40	47	8	13	2
	238	99	99	99	98	98	64	71	21	31	7
	240	99	99	99	99	99	84	88	43	55	20
	242	99	99	99	99	99	95	96	68	77	41
	244	99	99	99	99	99	99	99	86	91	66
	246	99	99	99	99	99	99	99	96	98	85
	248	99	99	99	99	99	99	99	99	99	95
	250	99	99	99	99	99	99	99	99	99	99

Notes: SEM = 3.2; Pct = Percentile.

Table 9:
Predicted probability of an observed MAP Reading score meeting or exceeding selected MAP benchmarks by grade/term for HS seniors expected to meet the Evidence-Based Reading and Writing SAT=480 Benchmark.

Grade		5		6		7		8		9	
	Term	Fa	Sp	Fa	Sp	Fa	Sp	Fa	Sp	Fa	Sp
	Benchmark	202.76	208.72	208.83	213.77	213.68	217.60	217.29	220.20	219.68	221.57
	SD	11.06	10.18	9.68	9.00	8.74	8.36	8.40	8.40	8.78	9.23
	Pct	43	42	45	45	49	50	49	50	49	50
Observed Score	194	1	1	1	1	1	1	1	1	1	1
	196	2	1	1	1	1	1	1	1	1	1
	198	7	1	1	1	1	1	1	1	1	1
	200	19	1	1	1	1	1	1	1	1	1
	202	41	2	2	1	1	1	1	1	1	1
	204	65	7	7	1	1	1	1	1	1	1
	206	84	20	19	1	1	1	1	1	1	1
	208	95	41	40	4	4	1	1	1	1	1
	210	99	66	64	12	13	1	1	1	1	1
	212	99	85	84	29	30	4	5	1	1	1
	214	99	95	95	53	54	13	15	3	4	1
	216	99	99	99	76	77	31	34	9	13	4
	218	99	99	99	91	91	55	59	25	30	13
	220	99	99	99	97	98	77	80	48	54	31
	222	99	99	99	99	99	92	93	71	77	55
	224	99	99	99	99	99	98	98	88	91	78
	226	99	99	99	99	99	99	99	97	98	92
	228	99	99	99	99	99	99	99	99	99	98
	230	99	99	99	99	99	99	99	99	99	99

Notes: SEM = 3.4; Pct = Percentile.