

Problems with the New PSAT

A Report by Compass Education Group



COMPASS

COMPASSPREP.COM

Executive Summary

The need to roll out the PSAT while still constructing the SAT puts College Board test developers and executives in a Catch-22. On the one hand, PSAT reporting must, in part, rely on preliminary data as it represents a work-in-progress toward the goal of a redesigned SAT. On the other hand, the merits of the PSAT are the best evidence the public has for judging the credibility of the new SAT. Moreover, the PSAT itself has important purposes. Students make vital decisions between SAT and ACT based on PSAT reports, and the exam gives students feedback on their college readiness. PSAT data is now linked to Khan Academy, where students can work on areas that the PSAT identified for improvement. On the school side, College Board has encouraged weaving the PSAT into a variety of counseling and tracking roles. Encouraging those uses comes with the responsibility of educating the educators on proper interpretation and use of scores. It's not clear that this standard has yet been met for the 2015 PSAT/NMSQT. Counselors are struggling to interpret some of the changes that have been made and may not be aware of some of the shifts or inconsistencies that exist. There have been mechanical challenges in rolling out the new test and reporting, frustrating counselors and students. Those issues have been covered elsewhere. Instead, this three-part analysis focuses on issues of interpretation and validity. Does the PSAT live up to its promise, and what does it portend for the SAT?

Compass Education Group's analysis is based on examination of student and counselor PSAT reports; discussions with students, parents, and counselors; and the tables and publications provided by College Board. The most important source document is College Board's PSAT/NMSQT Understanding Scores 2015; additional links will be provided where applicable. Rather than attempt a definitive exploration of the 2015 PSAT, this report examines three problematic areas of the new exam.

This report was co-authored by Art Sawyer, Bruce Reed, and Adam Ingersoll, the founders of Compass Education Group. Comments or queries can be sent to guide@compassprep.com.



A student with a backpack looking over a mountain range.

Part 1: Percentile Inflation

A series of changes has greatly increased the percentile scores that students and educators are seeing on PSAT score reports. College Board has not been transparent about all of the changes and the ways in which they can distort score interpretation.

Part 2 : Score Discrepancies

An historically narrow gap between sophomore and junior performance does not seem credible and leads to questions about how scoring, scaling, and weighting were performed and reported.

Part 3: Lowered Benchmark

A dramatic lowering of the college and career readiness benchmark for the “verbal” portion of the PSAT and SAT calls for a deeper examination and reveals potential structural problems with the new exam.



Part 1: Percentile Inflation

A series of changes has greatly increased the percentile scores that students and educators are seeing on PSAT score reports. College Board has not been transparent about all of the changes and the ways in which they can distort score interpretation.

Moving the Goalposts: A New, Hypothetical Measuring Stick

A shift in percentile reference groups has, perhaps, caused the most immediate and pervasive confusion in interpreting score reports. Many PSAT report recipients assume that percentiles are calculated directly from the current pool of test-takers. Surely, a 55th percentile score for a section should mean that of current test-takers, 55 out of 100 scored at or below that section score. That's not the case and has almost never been the case.

College Board has long used the previous year's test-takers as the reference group for PSAT/NMSQT percentile calculations. Those students provided the measuring stick, so to speak. Since the pool of test-takers evolves slowly, the difference between comparing 2014 students to 2013 students and comparing 2014 students to each other would not have been pronounced. With the 2015 PSAT, College Board has introduced an entirely new measuring stick — a “nationally representative sample” [also referred to as the “National Representative sample”] — and made it the default norm for student score reports. Here is College Board's definition:

“Nationally representative percentiles are derived via a research study sample of U.S. students in the student's grade (10th or 11th), weighted to represent all U.S. students in that grade, regardless of whether they typically take the PSAT/NMSQT.”

In other words, test-takers are compared to students who didn't even take the test and may never take the test. These percentiles are displayed prominently alongside student scores in both online and printed reports. We find that parents and students are using these percentiles as their primary source of information. Unfortunately, these “nationally representative” percentiles have several problems:

- they provide a source of percentile inflation
- they do not accurately compare students to the pool of students likely to take the SAT or ACT
- they represent a break from past reporting and mean that these figures cannot be compared to any prior data
- they represent a “black box” — it is unclear exactly how the national sample is derived, how accurately it reflects the national pool of students, or when or if it will be modified in the future.

College Board often cites transparency as a goal for its programs and as a justification of the new PSAT and SAT. Nationally-representative percentiles seem far less transparent than traditional test-taker percentiles. The new percentiles are not based on college bound students. The new percentiles are not based on others taking the same exam. The new percentiles are based on numbers that can only be judged via a technical report — such a report has yet to be released.

National Users: Students Become “Users” and Study Samples Replace Actual Results

An alternate set of percentiles, “User: National,” is also provided, but each score in the second set is only found several clicks deep in the online version of the report. In fact, a full student report contains 25 separate percentile scores. The temptation is to view “User” as interchangeable with the traditional notion of “test-takers,” but that would be inaccurate. The PSAT previously presented percentiles based on “[students] who took the test last year,” but the new PSAT has no “last year” from which to draw. Rather than opting to use actual student data from 2015 test-takers, College Board created a new reference group:

“User group percentiles are derived via a research study sample of U.S. students in the student’s grade, weighted to represent students in that grade (10th or 11th) who typically take the PSAT/NMSQT.”



This procedure is not uncommon, but that does not ensure that it was done accurately this time. At minimum, it creates another black box for students and educators. It’s a remedy that did not need to exist. Consider, by contrast, that the SAT and ACT are taken on many different dates over sophomore, junior, and senior year of high school. If a student is to get an accurate sense of how she stacks up to other students in her class, data from her class’s testing history must be consolidated. Since percentiles cannot be calculated contemporaneously with score reports, the testing organizations use scores from a prior group of test-takers. College Board uses the previous class year for the SAT, whereas ACT traditionally uses the prior three years. The data consolidation rationale does not exist for PSAT/NMSQT percentiles. As of October 28, 2015, every student who would ever take the 2015 PSAT/NMSQT had done so. Full results could have been tabulated and used for percentile calculation and reporting. Instead, College Board elected to use a sampling method that has not been disclosed and that is subject to the error inherent in any sampling.

Nationally Representative Scores Result in Percentile Inflation

The table below shows how the Nationally Representative percentiles differ from those for User and increase expected percentile scores. ACT uses prior test-takers only, so the PSAT is unique in this source of inflation.

Percentile Inflation Due to the Nationally Representative Sample

Evidence-Based Reading and Writing				Math			
Converted Section Score	PERCENTILES			Converted Section Score	PERCENTILES		
	National Representative Sample	PSAT/NMSQT User: National	Difference		National Representative Sample	PSAT/NMSQT User: National	Difference
540	74	71	+3	540	73	70	+3
530	71	67	+4	530	69	66	+3
520	67	63	+4	520	65	61	+4
510	63	59	+4	510	62	59	+3
500	60	54	+6	500	58	55	+3
490	55	50	+5	490	54	50	+4
480	50	47	+3	480	49	45	+4
470	46	43	+3	470	43	39	+4

A Percentile by Any Other Name: College Board Changes a Definition

A more fundamental change underlies all of the percentile scores on the new PSAT report. Few people give much thought to the various ways percentiles are defined, because the measure seems so simple to understand. [In this report, the vernacular “percentiles” will be used with no attempt to distinguish among percentiles, percentile rank, or cumulative percentages.]

In standardized test reporting, the two most common ways of defining percentiles for test-takers vary slightly enough that the distinction often gets overlooked:

- Definition A: The percentage of students scoring below you.
- Definition B: The percentage of students scoring at or below your score.

Definition B produces higher values in almost all cases and never gives lower values. College Board shifted from Definition A to Definition B this year, introducing an additional source of percentile inflation. Understood in context, there is no negative implication to this inflation — the new definition is just as valid and, perhaps, easier for a layperson to understand. The context, though, is easily lost. There are no red asterisks alerting to the change, so students and educators are understandably — and incorrectly — comparing 2015 percentiles to those from previous years.

Traditionally, College Board used Definition A and ACT used Definition B. It seems fitting that as the SAT and ACT grow more similar in content that their respective organizations now agree on Definition B. It is unclear if College Board will be using this definition for all of its exams. Below is an excerpt of percentile tables for the new PSAT; columns have been added for Definition A to demonstrate how percentile inflation can be observed.

Percentile Inflation Due to Change in Percentile Definition

Converted Section Score	Evidence-Based Reading and Writing		
	PERCENTILES		
	As Reported (Definition B)	2014 Reporting (Definition A)	Difference
540	71	67	+4
530	67	63	+4
520	63	59	+4
510	59	54	+5
500	54	50	+4
490	50	47	+3
480	47	43	+4
470	43	39	+4

Converted Section Score	Math		
	PERCENTILES		
	As Reported (Definition B)	2014 Reporting (Definition A)	Difference
540	70	66	+4
530	66	61	+5
520	61	59	+2
510	59	55	+4
500	55	50	+5
490	50	45	+5
480	45	39	+6
470	39	36	+3

Neither of the percentile definitions provide full information, because percentiles do not convey how many people achieved the same score. Any air traveler has encountered a variant of this problem when “Group 1” is called and 200 passengers rise as one. Having a high score is not as good if too many people share your score.

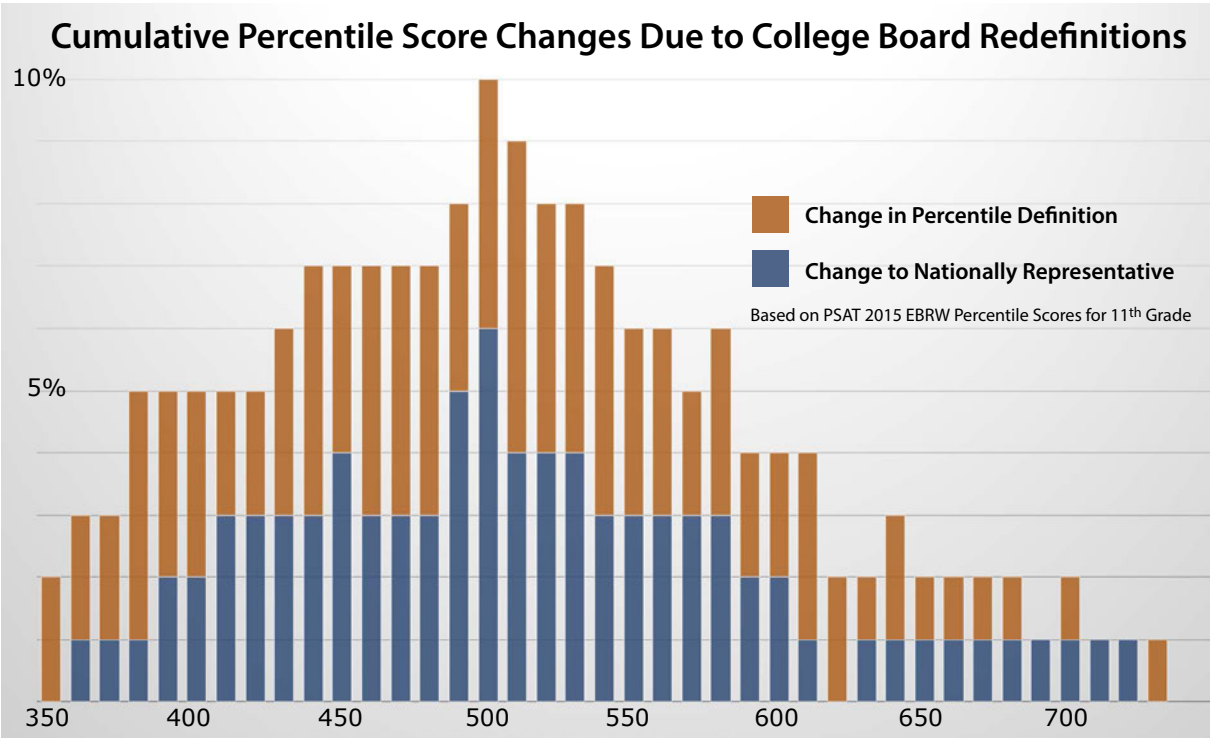


Definition A indicates what percentage of students achieved lower scores, but it cannot convey what percentage scored higher. For example, by looking at the table above at the cell for a score of 500 Math under Definition A, we can tell that 50 percent of students scored lower. Without referring to other cells, though, we do not know how many scored higher than 500 (it turns out to be 45 percent). In the case of Definition B, we would see that a 500 is the 55th percentile and know that 55 percent scored at 500 or below. This tells us that 45 percent scored higher ($100 - 55 = 45$). We would not know how many students scored below a 500.

The definitions do not change the underlying data, but students and educators are only provided a single value on their score reports, and it is the new, higher value. It can feel like having \$60 in your pocket rather than having \$50. It’s a nice feeling until you realize that prices have uniformly gone up by 20%. Across the middle, meatiest part of the score range, the change in definition “raises” percentiles by 2 – 6 points. It is plausible that College Board moved from Definition A to Definition B in part to give the feel-good impression of \$60 — especially since ACT was already handing out the extra bills.

Adding Up the Changes

The percentile inflation caused by the new definition and the new reference group are effectively additive. Under 2014’s percentile and reference group definitions, a 500 Math score would be presented as 50th percentile. On a 2015 score report, however, a 500 Math score is the 60th percentile. Percentile inflation is as high as 10 percent over part of the scale. At the edges of



the scale, the absolute change is smaller, but the proportional impact is higher. For example, while 99th percentile is only 2 percentile higher than 97th percentile, the cumulative reporting changes from 2014 mean a doubling or tripling of students receiving the higher figure. With 3% of students boasting 99th percentile scores this year, there are important implications for how students, parents, and counselors forecast National Merit scores, for example. Counselors typically see hundreds of reports, so they have been observing this proliferation of high scores without necessarily knowing why.

Percentile Inflation is Distinct from Score Inflation

The changes described on the previous page all relate to how percentiles are higher than ones reported in the past. Although scoring inconsistencies also appear to exist this year as a separate concern, percentile inflation in and of itself does not provide evidence that scores have been miscalculated or mis-distributed.

Why Did College Board Make These Changes?

In the course of replacing the old PSAT with the new PSAT, College Board has drawn on samples rather than actuals, swapped in new measuring sticks, and redefined how the measuring gets done.

“Why?”

It is hard to make a case for how students benefit from these changes. Percentile inflation may shift test planning decisions in unwelcome ways, and the sampling methodology and expanded comparison pool do little to answer questions about how a student’s scores stack up against those of other college applicants.

College Board, though, has multiple motives for making these shifts. The organization is under intense competitive pressures from ACT and other testing companies in the fight over whose testing products will be chosen to assess students from middle school to high school graduation. College Board cannot tolerate a competitive disadvantage just to preserve an old definition. Rebranding its ReadStep product as PSAT 8/9, creating a vertical scale that tracks students across all of its PSAT and SAT instruments, and rebadging the PSAT as the PSAT 10 when taken by sophomores in the spring have all been decisions to expand what College Board now dubs “The SAT Suite of Assessments.” College Board has strived to close any real or perceived competitive deficit, and the shift to the national sample fits into the organization’s long-term plans. States and school districts are increasingly contracting with the organization to offer the PSAT or SAT to all of their students rather than just a self-selected group of college bound students. These bulk buyers prefer standards that compare their students to all grade-equivalent students. The PSAT

8/9 is taken by far fewer students than the PSAT/NMSQT. Test-taker or “user” percentiles are more susceptible to change from exam to exam. College Board would ultimately like to offer the PSAT to every student across the country. It is, in essence, setting a benchmark with the goal of growing into it.

There are statistical reasons, too, why preference was given to a research study sample. Test makers generally want reporting data such as percentiles calculated prior to the administration of a new form. In hindsight, this preference was a risky decision for the PSAT given scrutiny of the exam by both proponents and critics. Many people are left wondering, “Is there something to hide?”

The reasons behind the decision to change the percentile definition and the default reference group may be valid, but the fact that the changes tend to amplify the percentiles and include an opaque leap from test-taker group to a Nationally Representative sample creates a dubious impression. A productive solution would be to release the actual numbers for test-takers and publish all research study results. The new SAT debuts on March 5, 2016, and many of its components are being built on the same research studies and with the same methods used for the PSAT. It would seem prudent to establish credibility with PSAT data now rather than play catch-up after final SAT numbers are released.





Part 2: Score Discrepancies

An historically narrow gap between sophomore and junior performance does not seem credible and leads to questions about how scoring, scaling, and weighting were performed and reported.

Sophomore Versus Junior Score Discrepancies Call Scoring Methodologies into Question

Percentile inflation caused by redefinition and re-norming creates unfortunate misinterpretations, but the sources of the change can be readily identified; previous percentile tables can be restated based on the new definition; the difference between Nationally Representative percentiles and User percentiles can be compared to gauge the difference added there. However, without further information from College Board it is impossible to know the accuracy of the 11th and 10th grade percentiles. Our analysis shows that there are significant problems in the way the numbers are being presented that mask the very thing the new test was meant to reveal — college readiness and academic progress. If score results between grades are suspect, it leads to questions about the pilot studies that were performed and how they inform the scoring for the PSAT and SAT.

Expected Versus Observed Score Differences Between Grades

Historically, juniors have outperformed sophomores on the PSAT/NMSQT by approximately 5 points per section [see table below]. Translated into SAT scores, the differences between 10th and 11th graders in 2014 were 48 points, 47 points, and 51 points in Critical Reading, Writing, and Math, respectively. On the new PSAT, however, the reported difference is only 12 points on Evidence-Based Reading and Writing (EBRW) and 19 points in Math. The average difference in 2014 is more than 3 times that seen in 2015. The 2014 grade differences were in line with those seen over the last decade, so they were not anomalous. The old and new PSAT are different tests, but student growth tends to show up similarly even on different college admission exams.



Grade Year Differences Fall From an Average of 49 Pts Per Section to 15 Pts

	2014 PSAT			2015 PSAT	
	Reading	Writing	Math	EBRW	Math
11 th Grade	474	459	486	480	489
10 th Grade	426	412	435	468	470
Difference	+48	+47	+51	+12	+19

*2014 PSAT scores presented on the old SAT scale.

Are Low Score Discrepancies Due to Differing Testing Populations?

Not all sophomores and juniors take the PSAT. Some take the PSAT as mandatory testing; some take the PSAT in order to qualify for National Merit; some take the ACT Aspire instead of the PSAT. If College Board's calculation of a nationally representative sample is correct, though, this year's grade differences should be immune from differences in test-taker demographics. Previous PSATs lacked a nationally representative sample, so sophomore to junior comparisons may be distorted by test-taker patterns. A way of removing potential distortion is to look at the results only for repeat testers — students who took the test in both school years. College Board has done research on the typical score change on the old PSAT by analyzing only students who took the test as sophomores and repeated the test as juniors [see table below]. The average increase, expressed in SAT points, was 33 points in Critical Reading, 33 points in Writing, and 40 points in Math. The figures are still twice what is being shown on PSAT reports as the 10th grade to 11th grade score differential.

Progress for Repeat Testers on Old PSAT

	Reading	Writing	Math
11 th Grade	487	481	507
10 th Grade	454	448	467
Change	+33	+33	+40

*Old PSAT scores presented on the old SAT scale.



Do Content Differences Between Old and New PSATs Provide an Explanation?

A remaining problem is that the old PSAT is not the new PSAT. Although the new and old tests cover roughly the same score range and do not have radically different means or standard deviations, we cannot be certain that year-over-year growth is identical. A third set of data is College Board’s own estimates of growth. Below are the College and Career Readiness Benchmarks.

Expected Benchmark Progress

	Readiness Benchmark	
	EBRW	Math
11 th Grade	390	500
10 th Grade	360	470
Change	+30	+30

College Board assumes that students improve at roughly 30 points from sophomore year PSAT to junior year PSAT and another 20 points from junior year PSAT to SAT. The PSAT figures — which themselves seem conservative — are still twice that shown in the 2015 student data.

Percentile Data for Sophomores and Juniors May Prove the Existence of Errors in Presentation, Computation, or Norming

The low observed score differences between 10th and 11th graders do not fit into a historical pattern, match studies of repeat testers, or align with assumed College Board benchmark progress. As improbable as the small point discrepancy is, though, it seems impossible to go one step further and state that sophomores outperform juniors. But this is exactly what the published percentile tables show [below].

Significant Discrepancies Emerge When Comparing Grades

Converted EBRW Score	Nationally Representative Percentiles	
	10th Grade	11th Grade
550	79	77
540	76	74
530	73	71
520	70	67
510	66	63
500	63	60

Over much of the score range, percentile scores show juniors performing approximately 10 scaled points higher than sophomores. For example, 510 for juniors and 500 for sophomores are both 63rd percentile. This difference is improbably low, but not impossible.

Converted EBRW Score	Nationally Representative Percentiles	
	10th Grade	11th Grade
720	99+	99+
710	99	99+
700	99	99+
690	99	99
680	98	99
670	98	98

At the high end of the scale, relative performance inverts and sophomores show higher scores than juniors. For example, a 700 gets a junior to a percentile of 99+, but it takes a 720 for a sophomore to reach that level. This inversion seems impossible to explain without calling College Board figures into question.

As you move up the scale, the difference between 10th and 11th graders disappears and then turns in favor of the younger students. Read literally, the score tables say that more sophomores than juniors achieved top scores on the PSAT/NMSQT. There have always been talented sophomores who score highly on the PSAT, but as a group, these students should not do better on the PSAT in 10th grade than they do in the 11th.

These figures are for the Nationally Representative groups, so cannot be explained away by saying that the test-taking populations are different. There is no logical statistical or content explanation as to how sophomores could actually perform better than juniors. In fact, we should be seeing scores 30-50 points higher per section for juniors. The most likely explanation is that the surveying and weighting methods used for the PSAT did not properly measure the class year compositions. If we assume this to be the case, though, can we be assured that the studies did any better in measuring the intra-class composition? Will the SAT be immune from the same problems?

Can Anything Explain the Low Sophomore/Junior Score Differences and the Score Inversion?

A suspect in the mix is the PSAT 10. Although the content of the PSAT 10 is identical to that of the PSAT/NMSQT, it is positioned as a way for schools to measure how students perform near the end of the sophomore year rather than toward the outset of the year. The PSAT 10 will first be offered between February 22 and March 4, 2016. It is a safe assumption that spring sophomores, adjusted for differences in the testing pool, will score higher than fall sophomores. If College Board statistically accounted for PSAT 10 takers in their figures, the scores for sophomores would be inflated.

It seems academically inappropriate to lump PSAT/NMSQT and PSAT 10 scores into the same bucket. The tests are taken at different phases of a student's high school progress. In fact, one reason a PSAT 10 exists is because spring performance differs from fall performance. The only clue that College Board may have made such a combination is reproduced from its Understanding Scores 2015. Highlighting has been added.

Percentiles for total scores

Converted Total Score	10 TH GRADE		11 TH GRADE	
	PERCENTILES		PERCENTILES	
	National Representative Sample	PSAT/NMSQT and PSAT 10 User: National	National Representative Sample	PSAT/NMSQT and PSAT 10 User: National
1520	99+	99+	99+	99+
1510	99+	99+	99+	99+
1500	99+	99+	99+	99+
1490	99+	99+	99+	99+
1480	99+	99+	99+	99+

It's likely that this reference is simply the result of a production error. The document never makes this reference again in its 32 pages. In short, all figures likely measure October performance for sophomores and juniors. This final attempt to explain the anomalous supremacy of sophomores comes up short. Even had a PSAT 10 explanation proved successful, it would have raised more questions than it answered.

Tables surrounding the PSAT are all marked as "Preliminary." College Board has made clear that final scaling for the redesigned SAT (and the PSAT is on the same scale) will not be completed until May 2016. Final concordance tables between old and new tests will replace any preliminary work. If the explanation of the statistical anomalies is that the paint is not yet dry, it begs the question as to what 3 million students and their educators are to do with the scores they have been presented. The new PSAT reports are the most detailed that have ever existed. They have total scores, section scores, test scores, cross-test scores, sub-scores, Nationally Representative percentiles, User percentiles, SAT score projections, sophomore and junior year benchmarks, and more. Which parts of the reports are reliable and which parts remain under construction? Should educators simply push these reports aside and wait until next year? Should students make test-taking and college choice decisions based on these scores?





Part 3: Lowered Benchmark

The dramatic lowering of the college and career readiness benchmark for the “verbal” portion of the PSAT and SAT calls for a deeper examination and reveals potential structural problems with the new exam.

College Board Sets an English Language Arts (ELA) Bar on the PSAT and SAT Far Below Comparable Standards

Despite its presumptive importance in college admission, the SAT has had a harder time establishing a wider role in K-12 education. Although revisions in 1994 and 2005 were meant, in part, to bring the SAT into closer alignment with academic standards, the exam still carried the residual suspicion that it only tested “test-taking skills,” was susceptible to short-term preparation, and was out-of-sync with education trends. Early on, ACT staked out a strong position in defining college readiness and in aligning test content with academic standards. The ACT surpassed the SAT in student volume by appealing to states and school districts to widen college access by offering testing to all students. As College Board plays catch-up — it only introduced its “SAT College Readiness Benchmark” in 2011 — ACT still brags about its historical advantage: “In ACT Aspire and the ACT, ACT has the only longitudinal college and career readiness assessments with scores linked to actual student performance in college courses and to descriptions of what students need to know and be able to do in college. No other organization can make this claim.” The redesigned SAT will need several years before it can make the same longitudinal and linking claims.

By the time David Coleman was hired in 2012 to lead College Board, it was clear that the organization was willing to prepare for the future even if it meant discarding much of what had come before. Coleman had been one of the architects of the Common Core standards for English Language Arts and was passionate about the SAT’s ability to be at the forefront of ensuring college and career readiness for all students.

An analysis of the new math and ELA college and career readiness benchmarks shows that while SAT Math standards were left largely untouched, the ELA benchmarks were lowered significantly. Below we explore five points that are at turns surprising, interesting, and troubling.



- College and career readiness ELA benchmarks on the PSAT and SAT have been lowered.
- The definition of “meets expectations” has changed.
- The “meets expectations” rate for the new ELA benchmarks is far above those of comparable exams.
- “Score pollution” at the low end of the EBRW scale impacts the validity of the 10th grade benchmark.
- No research has been published explaining how and why the ELA benchmarks were lowered or if any watering down occurred.

College Board had long left benchmarking to ACT. As the showdown between the companies moved more and more to state legislatures and departments of education, though, College Board decided to introduce a college and career readiness benchmark. A 1550 SAT was set in 2011 as the key benchmark, and section scores of 500 were set as benchmarks for Critical Reading, Writing, and Math [the composite score was higher than the sum of the individual sections]. PSAT score benchmarks were also developed to reflect progress toward those goals.

Revisions to the benchmarks were published along with the results of the new PSAT/NMSQT. Evidenced-Based Reading and Writing scores on the redesigned exam are nominally higher than the Critical Reading and Writing scores from the old PSAT, so it is surprising to see that the benchmarks were set lower on the new exam. A lot lower. Before comparing the new and old benchmarks, the old scores should first be concorded to new scores. When using a concordance, we cannot say that the new and old scores are completely interchangeable, but we can say that they represent comparable levels of performance — precisely what we want in the case of benchmarks.

It appears that the readiness benchmark for math has made it to the new PSAT and SAT with little change. The old SAT and new SAT are entirely different exams, so scores must be compared through a concordance that matches comparable levels of performance. When old and new Math scores are concorded, the benchmark differences are minimal.

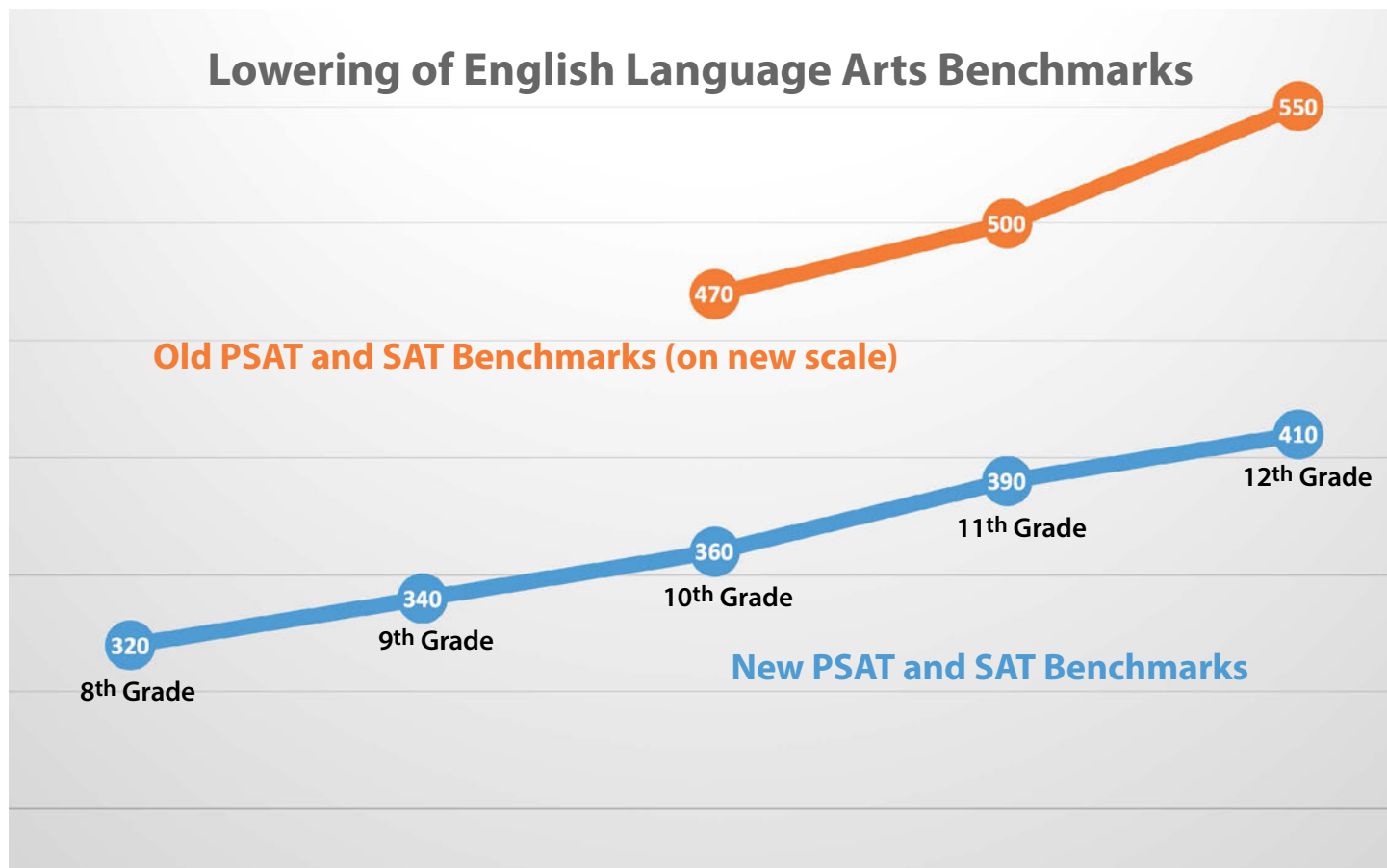
Math Readiness Benchmarks are Virtually Unchanged

Math Readiness Benchmarks			
	Old P/SAT*	New P/SAT	Change
10 th Grade	470	470	+0
11 th Grade	490	500	+10
12 th Grade	530	520	-10

*Benchmarks have been concorded to the new P/SAT scale.



In contrast to the continuity of the math benchmarks, the ELA benchmarks have been reset far lower than before. In the chart below, the old and new benchmarks are compared after setting them both on the scale of the redesigned P/SAT.



Among the takeaways on the new EBRW benchmark compared to the old Critical Reading and Writing benchmark:

- ELA benchmarks have been reduced by at least 110 points.
- The new mark is more than a standard deviation away from the old mark.
- Almost half (49%) of National Representative sophomores would miss the old benchmark. Only 10% of National Representative sophomores miss the new benchmark.
- The new 12th grade benchmark is 60 points lower than the previous 10th grade benchmark.

It is hard to grasp the extent of this change in an area so vital to college and career readiness. It makes little sense when attempting to bring more rigor to an admission test to lower the standards this much without strong reason. Soon before he assumed his duties as College Board president, David Coleman said, “The engine of social justice at the College Board is the presumption that more rigor draws students to a higher level. It would be the utmost betrayal of it to water that down.”



Changing Benchmark Definitions

For the new PSAT and SAT, College Board changed how it would define college and career readiness standards. Although such standards are almost universally referred to by testing organizations as “college and career readiness,” the benchmarks are usually defined by college success alone. Below is a comparison of College Board’s old and new benchmark definitions.

Old Readiness Standard

The college readiness benchmark was calculated as the SAT score associated with a 65 percent probability of earning a first-year GPA of 2.67 (B-) or higher.

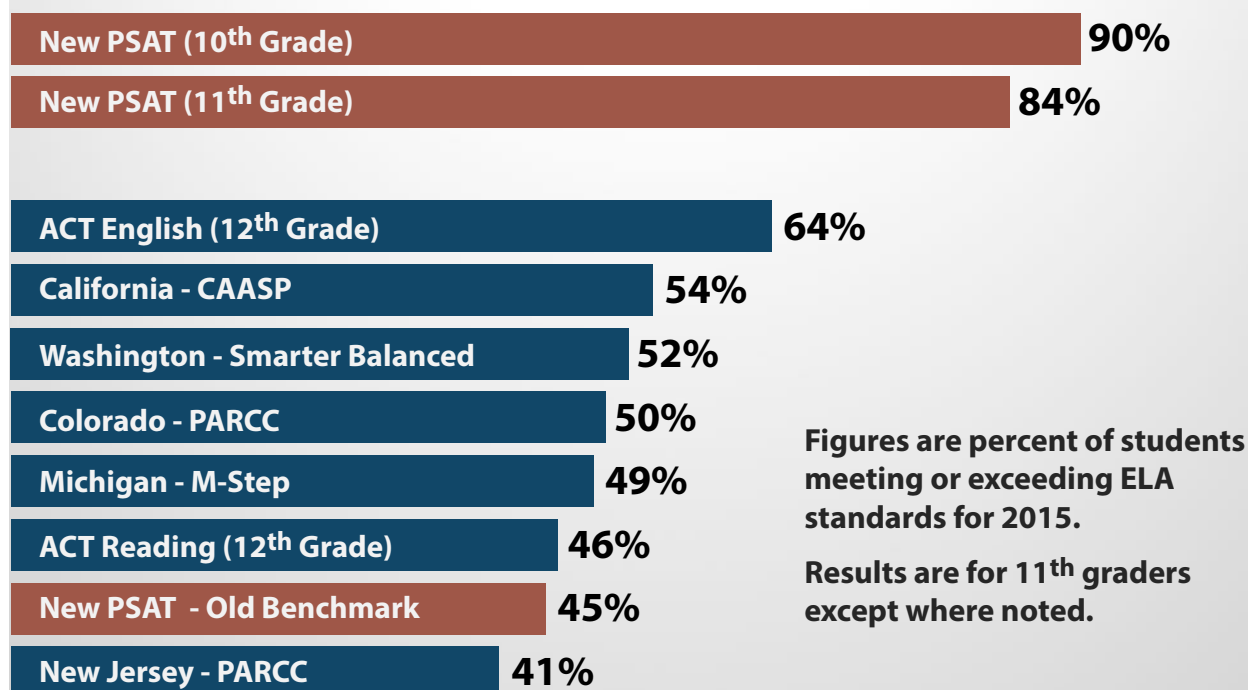
New Readiness Standard

The college and career readiness benchmarks for the SAT predict a 75 percent likelihood of achieving at least a C in a set of first-year, credit-bearing college courses.

The GPA cut-off has been dropped from 2.67 to 2.00, while the probability for success has been moved from 67% to 75%. However, the definitional change does not lead directly to the lower benchmark scores. In fact, the change brings College Board’s definition more in line with those used by ACT and by Partnership for Assessment of Readiness for College and Careers (PARCC), two important competitors in readiness assessments. A key difference is that ACT adds a piece to the equation: “*a 50% chance of obtaining a B or higher* or about a 75% chance of obtaining a C or higher [emphasis added].” ACT and College Board also have differing interpretations of the college courses included in calculating first-year GPA.

ACT’s ELA benchmarks are 18 for English and 22 for Reading. Due to content differences between the tests, these scores cannot be directly translated into SAT values. We can, though, look at performance against the benchmarks. Of graduating seniors in 2015 who took the ACT, 64% reached the English standard and 46% met or exceeded the Reading standard. This compares with 84-90% of PSAT/NMSQT Nationally Representative students meeting the grade-appropriate benchmarks. The figures become 85-92% if the User sample is considered. The “meets expectations” rate for the PSAT is almost double that of its previous standard and well above the rates in a basket of other ELA assessments.

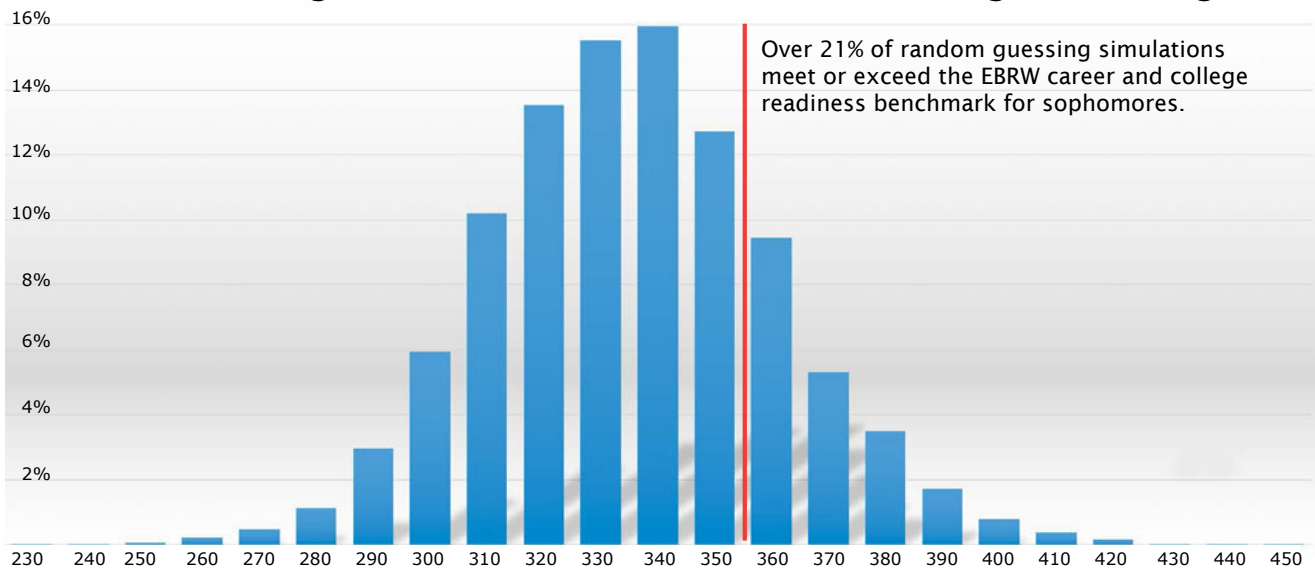
Do High Pass Rates Reflect Lower Standards?



An additional problem with the low benchmarks set on the PSAT — especially for sophomores — is that the standards start bumping into the limitations of the test’s construction. Since the scale of the new PSAT is 160 to 760, it would appear that even a score of 360 (the sophomore benchmark) is 200 points above the bottom of the scale. The effective bottom of the scale, though, is closer to 300-320. A key feature of the new PSAT is the elimination of a “guessing penalty.” Every student should answer every question, because no points are deducted for incorrect answers. Given that there are only 4 answer choices per question, a random guesser would get 25% of questions right over the long term. This “tailwind effect” takes many scaled scores out of play. A simulation of 10,000 students randomly guessing on every Reading and Writing question on the released PSATs from October 14, 2015 and October 28, 2015 found that the average score for those random guessers would be 335; 50% of students would get 340 or higher. In fact, 21% of random guessers would meet the sophomore college and career readiness benchmark! It is hard to understand how an important learning benchmark can be reached by 21% of students doing nothing but bubbling random patterns on their answer sheets. If the benchmark really should be that low, then the test is not well designed to measure the cut-off. The noise from random guessing creates a type of score pollution.



Random Guessing Simulation on Evidence-Based Reading and Writing



It is difficult to reconcile the EBRW benchmarks with one of the prime goals stated in the founding document for the redesigned SAT: “[T]he test must better reflect, through its questions and tasks, the kinds of meaningful, engaging, rigorous work that students must undertake in the best high school courses being taught today, thereby creating a robust and durable bond between assessment and instruction.” To date, no research reports have been published on how the new benchmarks were set. College Board’s most recent publication on readiness, [Research Foundations: Empirical Foundations for College and Career Readiness](#), does not yet deal with scores for the redesigned SAT and includes only descriptive statistics about college attainment. Did what is expected of students change so much so quickly that 90% of them will now meet expectations when only half as many met them last year? Were the previous expectations wrong? Do the new expectations better measure and expand opportunity? Will students respond positively — or at all — to the many different benchmarks provided? These are the types of questions that should be addressed.

Accountability and Transparency

While College Board long ago outgrew its official name — College Entrance Examination Board — the PSAT and SAT are still in transition from an earlier time. The public thinks of the PSAT and SAT primarily as college admission tests, but the exams are being tasked with an increasing number of duties — assessment, alignment, benchmarking, merit scholarships, gatekeeping, and placement. — for an increasing number of students and educators. States such as Connecticut, Michigan, Illinois, and Colorado have recently made multimillion dollar commitments to the SAT because of the promise of the redesign. Later this year, thousands of colleges will start receiving the new scores and slotting them into millions of applicant files (along with ACT scores). There is no slow build in standardized testing — change happens all at once. The inevitability of the new test, though, does not mean that attention should be limited to operational aspects. The rapid rise in public funding for the PSAT and SAT and the increasing number of tasks for which the exams claim competency require an increased level of scrutiny, accountability, and transparency.

Compass Education Group's three part series on the PSAT is meant to illustrate several of the ways in which the new test must find its bearings. That is not to say that percentile inflation or the rise of the sharp-witted sophomore or even the resetting of the ELA benchmarks is the most fundamental question around the new PSAT or SAT. There are many questions and challenges to be faced. Implementing an entirely new college admission exam is a Herculean undertaking, and College Board set ambitious delivery goals for both validity and rollout. We hope that the efforts to meet deadlines have not pushed aside the original impetus for the redesign. Trust in the new exam should not come from the fact that its name reuses three letters; students deserve an SAT that judges fairly and openly.





COMPASS
COMPASSPREP.COM

Southern California Office

9100 Wilshire Blvd, Suite 520E
Beverly Hills, CA 90212
(800) 925-1250

Northern California Office

1100 Larkspur Landing Circle, Suite 280
Larkspur, CA 94939
(800) 620-6250

