



MAGAZINE

The Story Behind the SAT Overhaul

By TODD BALF MARCH 6, 2014

In July 2012, a few months before he was to officially take over as president of the College Board, David Coleman invited Les Perelman, then a director of writing at M.I.T., to come meet with him in Lower Manhattan. Of the many things the College Board does — take part in research, develop education policy, create curriculums — it is perhaps most recognized as the organization that administers the SAT, and Perelman was one of the exam’s harshest and most relentless critics. Since 2005, when the College Board added an essay to the SAT (raising the total possible score from 1,600 to 2,400), Perelman had been conducting research that highlighted what he believed were the inherent absurdities in how the essay questions were formulated and scored. His earliest findings showed that length, more than any other factor, correlated with a high score on the essay. More recently, Perelman coached 16 students who were retaking the test after having received mediocre scores on the essay section. He told them that details mattered but factual accuracy didn’t. “You can tell them the War of 1812 began in 1945,” he said. He encouraged them to sprinkle in little-used but fancy words like “plethora” or “myriad” and to use two or three preselected quotes from prominent figures like Franklin Delano Roosevelt, regardless of whether they were relevant to the question asked. Fifteen of his pupils scored higher than the 90th percentile on the essay when they retook the exam, he said.

Right around the time Coleman was appointed as the board’s next president, he read an article about Perelman’s research in The New York Times and decided to reach out to him. “Somebody takes a whack at the SAT, so what?” Coleman said when I met him in his office at the College Board headquarters near Lincoln Center last month. “They get some media coverage, it’s not that interesting. But

this was a guy who devoted his lifetime to work you care about” — teaching students how to write — “and then looks at an instrument meant to celebrate writing and — ” Coleman’s words trailed off. “I wanted to go beyond the news presentation of his claim,” he finally added, “to the *depth* of his claim.”

Over the course of their two-hour conversation, Perelman told Coleman that he wasn’t opposed to an essay portion of the test, per se; he thought it was a good idea, if done well. But “when is there a situation in either college or life when you’re asked to write on demand about something you’ve never once thought about?” he asked. “I’ve never gotten an email from a boss saying: ‘Is failure necessary for success? Get back to me in 25 minutes?’ But that’s what the SAT does.” Perelman said that tutors commonly taught their students to create and memorize an all-purpose essay that contained the necessary elements for a top score — “a personal anecdote, a few historical references; Florence Nightingale seems a strangely popular reference.” When test day comes, they regurgitate what they’ve committed to memory, slightly reshaping depending on the question asked. But no one is actually learning anything about writing.

Perelman was surprised, he told me, by the productive nature of their conversation, but ultimately he couldn’t imagine that much would come of it. The College Board was a huge nonprofit organization, generating hundreds of millions of dollars in annual revenue (in part from the nearly three million SAT tests it administers to high-school students each year), and despite intense criticism in the past, it had done little, in Perelman’s estimation, to bring about meaningful change. “His heart is in the right place,” Perelman recalled thinking at the time. “David Coleman actually believes in education.” But trying to change the way the College Board does business, Perelman said, is “like trying to turn around the Titanic.” There was no way an institution as notoriously slow and defensive as Coleman’s was going to do that, no matter who was at the helm.

By the time he took over in October 2012, Coleman was well versed not just in Perelman’s critiques but also in a much wider array of complaints coming from all of the College Board’s constituencies: Teachers, students, parents, university presidents, college-admissions officers, high-school counselors. They all were unhappy with the test, and they all had valid reasons.

Students despised the SAT not just because of the intense anxiety it caused — it was one of the biggest barriers to entry to the colleges they dreamed of attending

— but also because they didn't know what to expect from the exam and felt that it played clever tricks, asking the kinds of questions they rarely encountered in their high-school courses. Students were docked one-quarter point for every multiple-choice question they got wrong, requiring a time-consuming risk analysis to determine which questions to answer and which to leave blank. Teachers, too, felt the test wasn't based on what they were doing in class, and yet the mean SAT scores of many high schools were published by state education departments, which meant that blame for poor performances was often directed at them.

An even more serious charge leveled at the test was that it put students whose families had money at a distinct advantage, because their parents could afford expensive test-prep classes and tutors. Several years ago, an exasperated Mitch Kapor, a founder of Lotus Software, co-wrote an op-ed in *The San Francisco Chronicle* suggesting colleges should require mandatory disclosure by students and parents of “each and every form of purchased help,” as a way to level the playing field.

When the Scholastic Aptitude Test was created in 1926, it was promoted as a tool to create a classless, Jeffersonian-style meritocracy. The exam, which purported to measure innate intelligence, was originally adapted from the World War I Army I.Q. test and served as a scholarship screening device for about a dozen selective colleges throughout the 1930s. It was assumed that there was no way to effectively prep for a test geared to inborn intelligence, but as early as 1938, Stanley Kaplan began offering classes that promised higher scores. Today the company Kaplan founded and its main competitor, the Princeton Review, are joined by innumerable boutique firms (not to mention high-priced private tutors), all part of a \$4.5-billion-a-year industry that caters largely to the worried wealthy in America who feel that the test can be gamed and that their children need to pay to learn the strategies.

Coleman conducted a “listening thing” with his organization's various frustrated constituencies. For the College Board to be a great institution, he thought at the time, it had to own up to its vulnerabilities. “Unequal test-prep access is a problem,” he said. “It is a problem that it's opaque to students what's on the exam. It is a problem that the scoring is too complex. I knew some of the science behind the SAT and actually admired a lot of it. On the other hand, I felt that something really had to happen, because what had grown up around it” — the

way in which the test evolved from a vehicle to encourage meritocracy to a reinforcement of privilege in American education — “threatened everything.”

It was clear, Coleman said, that no parents, whatever their socioeconomic status, were satisfied. The achievements of children from affluent families were tainted because they “bought” a score; those in the middle class cried foul because they couldn’t get the “good stuff” or were overextended trying to; and the poor, often minority students, were shut out completely. A paper prepared in 2009 by Derek Briggs, the chairman of the Research and Evaluation Methodology program at the University of Colorado, Boulder, emphasized another cost to test prep beyond the \$1,000-plus classes and the personal tutors: He called it an opportunity cost, meaning that time spent in the narrow pursuit of beating the test meant time away from schoolwork and extracurricular activities that are actually designed to prepare students to succeed in college.

In addition to these educational (and moral) quandaries, Coleman had to grapple with what it meant for the College Board as a business to have the credibility of the SAT called into question. A growing number of colleges and universities, frustrated by the minimal change to the SAT when it was revised in 2005 and motivated by a report issued in 2008 by the National Association for College Admission Counseling (Nacac), began to eliminate the SAT and its competitor, the A.C.T., as admission requirements, following the lead of several small, liberal-arts colleges that did so years before. The authors of the Nacac report cited a University of California study, which characterized the SAT as a “relatively poor predictor of student performance” and questioned the tendency of colleges to rely on the SAT as “one of the most important admission tools.” (Many of the schools that dropped test requirements saw spikes in their applications, at least in the first year.)

Around the time the report came out — and following the publication of “The Power of Privilege,” by the Wake Forest University sociology professor Joseph A. Soares, an account of the way standardized tests contributed to discriminatory admissions policies at Yale — Wake Forest became the first highly rated institution (it regularly appears as a Top 30 university on the U.S. News & World Report college rankings) to announce a test-optional admissions policy. Follow-up studies at Wake Forest showed that the average high-school G.P.A. of incoming freshmen increased after the school stopped using standardized-test scores as a factor.

Seventy-nine percent of its 2012 incoming class was in the top 10 percent of their high-school classes. Before going test-optional, that figure was in the low 60s. In addition, the school became less homogeneous. “The test highly correlates with family income,” says Soares, who also edited a book that, in part, examines the effects of making the SAT optional at the University of Georgia, Johns Hopkins University and Wake Forest. “High-school grades do not.” He continued, “We have a lot more social, racial and lifestyle diversity. You see it on campus. Wake Forest was a little too much like a J. Crew catalog before we went test-optional.”

A report released last month by William C. Hiss, a former dean of admissions at Bates College, and Valerie W. Franks, a former Bates assistant dean of admissions, supports Wake Forest’s experience. They reviewed 33 colleges and universities that did not require SAT or A.C.T. scores and found no significant difference in college G.P.A. or graduation rates between those who had submitted tests and those who had not. Specifically, they saw that students with good high-school grades did well in college, even if they had weak SAT scores. But students with weaker high-school grades — even with strong SATs — did less well in college. Those who didn’t submit SATs were more likely to be minority students, women, Pell grant recipients or the first in their families to go to college.

While more colleges are choosing to opt out of standardized testing, an estimated 80 percent of four-year colleges still require either SAT or A.C.T. scores, according to David Hawkins at Nacac, and admissions officers report feeling bound to the tests as a way to filter the overwhelming numbers of applicants. Robert Sternberg, a celebrated author and Cornell professor, told “Frontline” that when he was at Yale and reviewed admissions applications, the scores were hard to ignore. “I know that when I’m reading applications and as the night goes on and I’m reading more and more, it gets more and more tempting to count the SATs,” he said. “It’s easier than reading these long essays and teacher recommendations. It’s human nature.” On top of the pressures to winnow the applicants, the Nacac report cited the problems resulting from the use of SAT and A.C.T. scores by U.S. News & World Report to create its rankings, stating that the scores “were not a valid measure of institutional quality.” In addition, it criticized the use of the SAT and A.C.T. by bond-rating companies to help assess the financial health of a school as creating “undue pressure on admission offices to pursue increasingly high test scores.”

Coleman said that many of the admissions officers he spoke with made it clear that they were uncomfortable being beholden to the test, at least to *this* test, but there was no consensus about what an exam that was fair and acceptable to all would look like.

Hard questions have always interested Coleman. In 1994, after earning a bachelor's degree in philosophy at Yale, a bachelor's in English literature at Oxford (where he was a Rhodes scholar) and a master's in ancient philosophy at Cambridge ("three degrees that entitled you to zero jobs"), Coleman intended to come home to New York City and work as a public-school teacher. But when he realized he wouldn't find a job teaching high-school English, he ended up instead as a consultant at McKinsey & Company, where for five years he became increasingly obsessed with evidence-based solutions. During that time, he did pro bono work for school districts trying to improve performance, and in 1999, he left McKinsey and helped create a company called the Grow Network, which focused on assisting students and parents, including non-English-speaking families, in navigating an educational system that was increasingly dictated by standardized tests. His immersion in the world of standardized testing — talking to educators as well as students — convinced him that the standards those tests were supposedly measuring had to change. They were too vast and vague, and they produced textbooks that suffered from the same lack of purpose.

"When you cover too many topics," Coleman said, "the assessments designed to measure those standards are inevitably superficial." He pointed to research showing that more students entering college weren't prepared and were forced into "remediation programs from which they never escape." In math, for example, if you examined data from top-performing countries, you found an approach that emphasized "far fewer topics, far deeper," the opposite of the curriculums he found in the United States, which he described as "a mile wide and an inch deep."

In 2008, Coleman helped start a nonprofit organization called Student Achievement Partners, which was dedicated to "acting on evidence" when making decisions about education policy. While at Partners, Coleman was integral in helping shape the Common Core, a set of academic standards that has subsequently been implemented in more than 40 states. While not without its critics — many parents and educators believe it deeply roots schools and teachers in a problematic "teach to the test" mind-set — Coleman talks about it not just as a

bulwark against the declining standards of American public education but as a rare bipartisan success. At the Strategic Data Project conference last May in Boston, he challenged those in the audience to cite a “significant domestic policy area where Republicans and Democrats have gotten together and gotten something done.” The Common Core, he said, was a galvanizing idea that “swept the country during a period when all ideas seemed to stop.”

When Coleman attended Stuyvesant High in Manhattan, he was a member of the championship debate team, and the urge to overpower with evidence — and his unwillingness to suffer fools — is right there on the surface when you talk with him. (Debate, he said, is one of the few activities in which you can be “needlessly argumentative and it advances you.”) He offended an audience of teachers and administrators while promoting the Common Core at a conference organized by the New York State Education Department in April 2011: Bemoaning the emphasis on personal-narrative writing in high school, he said about the reality of adulthood, “People really don’t give a [expletive] about what you feel or what you think.” After the video of that moment went viral, he apologized and explained that he was trying to advocate on behalf of analytical, evidence-based writing, an indisputably useful skill in college and career. His words, though, cemented his reputation among some as both insensitive and radical, the sort of self-righteous know-it-all who claimed to see something no one else did.

Coleman obliquely referenced the episode — and his habit for candor and colorful language — at the annual meeting of the College Board in October 2012 in Miami, joking that there were people in the crowd from the board who “are terrified.”

The lessons he brought with him from thinking about the Common Core were evident — that American education needed to be more focused and less superficial, and that it should be possible to test the success of the newly defined standards through an exam that reflected the material being taught in the classroom. This was exactly how the College Board’s Advanced Placement program worked (80 percent of teachers surveyed in a study by the Fordham Institute said that the A.P. exam was a good indication of their and their students’ work). It was also one of the main suggestions in Nacac’s 2008 report, that a college admission exam should be redesigned as an achievement-style test — like the A.P. exams — that would send a “message to students that studying their course material in high

school, not taking extracurricular test-prep courses that tend to focus on test-taking skills, is the way to do well on admission tests and succeed in a rigorous college curriculum.”

The question for Coleman was how to create an exam that served as an accurate measure of student achievement and college preparedness and that moved in the direction of the meritocratic goals it was originally intended to accomplish, rather than thwarting them.

More than a year ago, Coleman and a team of College Board staff members and consultants began to try to do just that. Cyndie Schmeiser, the board’s chief of assessments, told me that their first order of business was to determine what the test should measure. Starting in late 2012 and continuing through the spring of 2013, she and her team had extensive conversations with students, teachers, parents, counselors, admissions officers and college instructors, asking each group to tell them in detail what they wanted from the test. What they arrived at above all was that a test should reflect the most important skills that were imparted by the best teachers. Schmeiser explained that, for example, a good instructor would teach Martin Luther King Jr.’s “I Have a Dream” speech by encouraging a conversation that involved analyzing the text and identifying the evidence, both factual and rhetorical, that makes it persuasive. “The opposite of what we’d want is a classroom where a teacher might ask only: ‘What was the year the speech was given? Where was it given?’ ”

The team then set about trying to create test questions that lent themselves to this more meaningful engagement. Schmeiser said that in the past, assembling the SAT focused on making sure the questions performed on technical grounds, meaning: Were they appropriately easy or difficult among a wide range of students, and were they free of bias when tested across ethnic, racial and religious subgroups? The goal was “maximizing differentiation” among kids, which meant finding items that were answered correctly by those students who were expected to get them right and incorrectly by the weaker students. A simple way of achieving this, Coleman said, was to test the kind of obscure vocabulary words for which the SAT was famous (or infamous). The answer pattern was statistically strong, he said — a small percentage of the kids knew them, most did not — but it didn’t adequately reflect the educational values Coleman believed in. In redesigning the test, the College Board shifted its emphasis. It prioritized content, measuring each

question against a set of specifications that reflect the kind of reading and math that students would encounter in college and their work lives. Schmeiser and others then spent much of early last year watching students as they answered a set of 20 or so problems, discussing the questions with the students afterward. “The predictive validity is going to come out the same,” she said of the redesigned test. “But in the new test, we have much more control over the content and skills that are being measured.”

When I met with Coleman in his office last month to talk about the remaking of the SAT, he periodically leapt from his chair when he became excited about an idea. At one point he jumped up and drew a dividing line down the middle of his whiteboard (he’s a very enthusiastic user of the whiteboard), then scrawled, “Evidence-based reading and writing” on one side and “Math” on the other. He was unveiling, at least in broad strokes, the results of those many months of rethinking and testing.

Starting in spring 2016, students will take a new SAT — a three-hour exam scored on the old 1,600-point system, with an optional essay scored separately. Evidence-based reading and writing, he said, will replace the current sections on reading and writing. It will use as its source materials pieces of writing — from science articles to historical documents to literature excerpts — which research suggests are important for educated Americans to know and understand deeply. “The Declaration of Independence, the Constitution, the Bill of Rights and the Federalist Papers,” Coleman said, “have managed to inspire an enduring great conversation about freedom, justice, human dignity in this country and the world” — therefore every SAT will contain a passage from either a founding document or from a text (like Lincoln’s Gettysburg Address) that is part of the “great global conversation” the founding documents inspired.

Coleman gave me what he said was a simplistic example of the kind of question that might be on this part of the exam. Students would read an excerpt from a 1974 speech by Representative Barbara Jordan of Texas, in which she said the impeachment of Nixon would divide people into two parties. Students would then answer a question like: “What does Jordan mean by the word ‘party’?” and would select from several possible choices. This sort of vocabulary question would replace the more esoteric version on the current SAT. The idea is that the test will emphasize words students should be encountering, like “synthesis,” which can

have several meanings depending on their context. Instead of encouraging students to memorize flashcards, the test should promote the idea that they must read widely throughout their high-school years.

The Barbara Jordan vocabulary question would have a follow-up — “How do you know your answer is correct?” — to which students would respond by identifying lines in the passage that supported their answer. (By 2016, there will be a computerized version of the SAT, and students may someday search the text and highlight the lines on the screen.) Students will also be asked to examine both text and data, including identifying and correcting inconsistencies between the two.

“Whenever a question really matters in college or career, it is not enough just to give an answer,” Coleman said. “The crucial next step is to support your answer with evidence,” which allows insight into what the student actually knows. “And this change means a lot for the work students do to prepare for the exam. No longer will it be good enough to focus on tricks and trying to eliminate answer choices. We are not interested in students just picking an answer, but *justifying* their answers.”

To that end, the question for the essay portion of the test will also be reformulated so that it will always be the same, some version of: “As you read the passage in front of you, consider how the author uses evidence such as facts or examples; reasoning to develop ideas and to connect claims and evidence; and stylistic or persuasive elements to add power to the ideas expressed. Write an essay in which you explain how the author builds an argument to persuade an audience.” The passage will change from test to test, but the analytical and evidentiary skills tested will always be the same. “Students will be asked to do something we do in work and in college every day,” Coleman said, “analyze source materials and understand the claims and supporting evidence.”

The math section, too, will be predicated on research that shows that there are “a few areas of math that are a prerequisite for a wide range of college courses” and careers. Coleman conceded that some might treat the news that they were shifting away from more obscure math problems to these fewer fundamental skills as a dumbing-down the test, but he was adamant that this was not the case. He explained that there will be three areas of focus: problem solving and data analysis, which will include ratios and percentages and other mathematical reasoning used to solve problems in the real world; the “heart of algebra,” which

will test how well students can work with linear equations (“a powerful set of tools that echo throughout many fields of study”); and what will be called the “passport to advanced math,” which will focus on the student’s familiarity with complex equations and their applications in science and social science.

Last June, Coleman spoke at the Harvard Summer Institute’s multiday seminar for college-admissions and counseling professionals. Before the talk, he met with William Fitzsimmons, the longtime dean of admissions and financial aid at Harvard and the primary author of the 2008 Nacac commission report. Coleman brought along an outline of the SAT redesign to get Fitzsimmons’s impressions.

Fitzsimmons told me he was stunned by what he saw, the ways in which the exam read like a direct response to his commission’s most serious recommendations. “Like any other truly significant change, there will be debate,” he added. But then he went on: “Sometimes in the past, there’s been a feeling that tests were measuring some sort of ineffable entity such as intelligence, whatever that might mean. Or ability, whatever that might mean. What this is is a clear message that good hard work is going to pay off and achievement is going to pay off. This is one of the most significant developments that I have seen in the 40-plus years that I’ve been working in admissions in higher education.”

But changing the test didn’t solve all the problems that preoccupied Coleman. He was still troubled by the inequalities in education opportunity and believed that the College Board should play a role in ameliorating them. For some time, the College Board had been aware of the work of Caroline Hoxby, a professor of economics at Stanford, and Christopher Avery, a professor of public policy and management at Harvard’s John F. Kennedy School of Government, who had been studying what is sometimes called undermatching — the tendency of poor students to pick a school that is closer to home and less rigorous, in spite of evidence that they could succeed elsewhere. Hoxby first became aware of the problem in 2004, when she was on the faculty at Harvard and the university announced to great fanfare that it would recruit top-performing, financially challenged kids by offering free tuition if their parents made less than \$40,000. Yet despite the offer, enrollment numbers for those students remained stubbornly low. When Hoxby began to research the issue, she hypothesized that there was a large population of high-achieving, low-income students yet to be identified. She

and Avery began working with the College Board and the A.C.T. to develop new techniques to find out how many students were in this low-income, top-performing pool and where they lived. By piecing together data from census reports, I.R.S. income data broken down by ZIP code, real estate valuations and other sources, they pinpointed some 35,000 students whose grades were in the top-10 percent nationally and whose family income was in the bottom quarter of families with a 12th-grader.

When they tracked where those kids applied to school, they found a number that would later shock Coleman. Fifty-six percent didn't apply to a single selective college or university.

The researchers surmised this was a problem of communication, more than anything else; the information wasn't reaching the students and families it needed to reach, and in the cases when it did, it wasn't as clear and useful as it could be. Hoxby and Sarah Turner, an economics professor at the University of Virginia, tested whether they could change enrollment patterns. From 2010 to 2012, Hoxby's team sent out personalized, detailed packets encouraging the high-achieving, low-income students to apply to several schools and providing application-fee waivers and financial-aid information about scholarships. In many cases these students would be able to see that they could get a better deal financially at more highly selective schools that wanted to attract them. The intervention resulted in those students' applying to more colleges and closing "part of the college behavior 'gap' between low-income and high-income students with the same level of achievement," Hoxby and Turner wrote.

When Coleman became College Board president, he was briefed on the supporting role the board had played to date. He agreed with those who saw an opportunity and decided, he says, to "take it from a small experiment to implement it nationwide." He called for additional research to find low-income students whom the board refers to as "college-ready," meaning they scored 1,550 or above on the SAT (the top 43 percent of U.S. test-takers). Ultimately the board mailed out nearly 100,000 packets to top-performing and college-ready students. The packets included four or more application waivers to allow students to apply immediately to any of the more than 2,000 schools that agreed to participate in the program. One requirement of those colleges, Coleman said, was that they agreed to rely on the financial determination made by the College Board and didn't

make the students requalify for aid or special tuition dispensation. Instead, the waiver was designed to look like a ticket — “Here’s your ticket, go!” Coleman said — simplifying the process and encouraging the students to jump at the opportunity. Research on the initial effects of the program won’t be released until next month, but the speed with which the deployment happened fulfilled Coleman’s promise to accelerate the agenda.

In January 2013, at a College Board-sponsored conference in Florida, Coleman met Daniel Porterfield, the president of Franklin & Marshall College, whose surprisingly effective work in bringing high-achieving, low-income students to his small liberal-arts school in Lancaster, Pa., gained national attention. Porterfield agreed with Hoxby’s team’s conclusions that the problem wasn’t that the students weren’t out there; the problem was that colleges weren’t looking hard enough to find them, and that this commitment was a big part of Franklin & Marshall’s success. Porterfield, who is now a board member of the College Board, told me he saw Coleman as uniquely “using the College Board to serve society as opposed to the College Board serving its own position.” He also said that when the two of them first talked, Coleman promised that the College Board could help F&M find talented, high-striving, high-achieving students and that he has “exactly delivered on that promise.”

What Coleman found exciting about the intervention was its use of the standardized tests as a way to reach students who would otherwise not apply to the kinds of colleges that they might assume were out of reach. It transformed an exam that most thought of as a burden — and many low-income students opted not to take at all — into an opportunity. Coleman explained that the moment when students get their test results is a rare instance in which you have their full attention — that’s the moment you have to seize on, connecting the score they’re holding in their hand to the future that they could possibly attain. “When have you ever gotten anything for taking the SAT?” he said, imagining the reaction of students opening up their test results and holding the application waivers in their hands.

For all the good intentions and all the evidence-based ideas brought to bear by Coleman and his colleagues over the past year and a half, there is still a chasm between the educational experiences of children at good schools in wealthier districts and those in lower-income areas. The fact that you could never fully level

the playing field — that good, focused instruction and meaningful preparation would still be unavailable to the students they were most focused on lifting up — nagged at Coleman and his staff as they continued redesigning the test.

They began to consider how they might provide teachers in sixth through 12th grade, particularly in low-performing schools, with broader access to content and resources to help prepare students for the test. Then last July, on a bus to dinner at a staff retreat in upstate New York, two of Coleman’s senior team members, Cyndie Schmeiser and Jeff Olson, threw out an idea: What if the College Board worked with Khan Academy, the free online tutoring service, visited by 10 million students each month, to offer SAT prep classes to anyone who wanted them? At Khan Academy, students logged on to the site and then worked over weeks or months at their own pace answering questions in different subject areas and following their progress. If they needed help, they could watch one of the thousands of casual but engaging videos created by the founder, Sal Khan. Khan holds multiple degrees from Harvard and M.I.T. and serves as the site’s ubiquitous guide, his voice explaining how to do various problems while text and numbers appeared on a digital chalkboard.

Khan started the site in modest fashion, tutoring his young niece over the Internet. When her relatives and friends wanted to be taught by him, too, he began posting videos to YouTube. As the site grew, he worked through all sorts of problems, including some that he took from past SAT exams, which he says now he probably didn’t have permission to use.

Coleman and his team were aware of what was happening at Khan Academy and were intrigued by the idea of a partnership, but they were also wary. “You kind of say, ‘O.K.,’” Coleman said. “‘But is it good enough?’” This kind of partnership had never been done by the College Board and he worried about what it might mean for the brand.

Throughout last fall his staff spent many hours on the Khan Academy website. The idea of creating a transparent test and then providing a free website that any student could use — not to learn gimmicks but to get a better grounding and additional practice in the core knowledge that would be tested — was appealing to Coleman.

He thought about athletics as the corollary for what they were trying to do. In sports, you practiced all the time to prepare for games. But while the stakes were

high on game days, they didn't result in the "suffering" and counterproductive anxiety that was a common reaction to the SAT. The difference, he said, was that in sports there was no mystery as to what would be required of you in a game. The rules were clear and didn't change. On the SAT, the rules were unclear. "Half the anxiety is about what's going to happen on 'game day,'" Coleman said. "It's not really fair. High stakes should not be placed on something that didn't matter before that suddenly matters now. The stakes should emerge because the work is important and your demonstration of that is significant."

The long, deliberate practice required for that type of performance was consistent with the Khan Academy method. In theory, anyway, the partnership made perfect sense.

In January, Coleman met with Wade Henderson, the president and C.E.O. of the Leadership Conference on Civil and Human Rights, who spoke with him about the ill will that had been built up in the minority community over the SAT, how the test has long been viewed not as a launching pad to something better but as an obstacle to hard-working, conscientious students who couldn't prepare for it in the way more affluent students could. Coleman acknowledged "the extent to which the exam recapitulates income inequality." Henderson also expressed concern, Coleman said, that poor SAT scores could block access to jobs. After the hourlong conversation, which Coleman characterized as deeply moving, he decided to add one more element to the redesign. Test information sent to an institution would include a "safe use" warning in red ink: "This data should only be used in combination with other relevant information to make responsible decisions about students."

A couple of weeks after his talk with Henderson, Coleman flew to Silicon Valley to discuss a partnership with Sal Khan. There was no discussion of financial terms, just an agreement in principle that they would join forces. (The College Board won't pay Khan Academy.) They talked about a hypothetical test-prep experience in which students would log on to a personal dashboard, indicate that they wanted to prepare for the SAT and then work through a series of preliminary questions to demonstrate their initial skill level and identify the gaps in their knowledge. Khan said he could foresee a way to estimate the amount of time it would take to achieve certain benchmarks. "It might go something like, 'O.K., we think you'll be able to get to this level within the next month and this level within

the next two months if you put in 30 minutes a day,'” he said. And he saw no reason the site couldn't predict for anyone, anywhere the score he or she might hope to achieve with a commitment to a prescribed amount of work.

Coleman told Khan that the College Board would invest in an outreach campaign through organizations like Boys and Girls Clubs and Big Brothers Big Sisters groups to reach as many students as possible, especially low-income students who aren't the website's primary users now. He also gave Khan access to actual test questions, and Khan is in the process of creating material for students who will be taking the old exam. (He says that it will be available early this month.) Coleman told me that his confidence in the partnership crystallized when Khan told him they were constantly revising their material based on what was most effectively helping students on the site. Coleman was particularly inspired by Khan's belief that it was possible for any student to achieve better skills with the proper instruction. Khan asked if Coleman was aware that five centuries ago, there was an analogous misperception. Coleman explained, “He said, ‘David, do you realize it used to be believed that most human beings couldn't read?’ ”

At various times in our discussions, Coleman referred to some test-prep providers as predators who prey on the anxieties of parents and children and provide no real educational benefit. (Though there's a debate about how helpful test prep is, much research shows increases of an average of only 30 points.) “This is a bad day for them,” he said about the new test and his Khan Academy partnership.

Still, Coleman concedes that the redesigned SAT won't quiet everyone's complaints, and he doesn't expect there to be a universal celebration of what they've done. You can imagine there will be substantial questions, for instance, about whether any standardized test can be fair across all groups, and whether the College Board is not ultimately creating a new test that somehow, some way, will be gamed as much as the old one.

Coleman's response to those concerns is to say that the new, more transparent test will be tied to what's being taught in high school and will be evidence-based. But his previous work on the Common Core has raised some educators' concerns. “Dave Coleman is not an educator by training,” says Lucy Calkins, the founding director of the Reading and Writing Project at Columbia University's Teachers College and an author of “Pathways to the Common Core.” Calkins has been a

strong defender of the Common Core but thinks Coleman has been too insistent on his own particular method for implementing its standards. She cites a video that Coleman helped create of a “model lesson” for teaching the Gettysburg Address, where he would have students spending several classes “parsing the meaning of each word in each paragraph,” she said. She doesn’t feel there’s evidence that this method works.

With a redesigned SAT, Calkins thinks that too much of the nation’s education curriculum and assessment may rest in one person’s hands. “The issue is: Are we in a place to let Dave Coleman control the entire K-to-12 curriculum?”

William Fitzsimmons, the Nacac chairman and head of admissions at Harvard, for his part, was impressed with the quickness with which Coleman has been able to make these changes. “In the world of education,” Fitzsimmons told me, “this is lightninglike speed.” And Coleman rejects the worries that he might be making changes that are too radical without waiting to see what works. He says that he believes that if you’ve been diligent in gathering the supporting facts, which he has been, then that is your defense against hubris and wrong thinking. In reality, he said, the decisions he has made aren’t all that bold, because they’re all completely supported by research. This is where he and critics like Lucy Calkins disagree, of course, but like any good debater, Coleman seems to know when to marshal hard evidence and when to wrap it in persuasive rhetoric. And what’s at stake, he often makes clear, is not just the fairness and usefulness of an exam but our nation’s ability to deliver opportunity for all, which, really, is about the soul of the country. The rest of us will have to wait for the proof that he has found the answer.

Todd Balf is the author of several books, including “Major,” an account of Marshall Taylor, one of the first African-American athletes to become an international superstar.

Editor: Ilena Silverman

A version of this article appears in print on March 9, 2014, on page MM26 of the Sunday Magazine with the headline: The SAT is Hated by . . . All of the Above.