

Sunday, Sep. 3, 2006

How I Did on the SAT

In 2003 I predicted dire consequences from a massive test redesign.
What I got right--and wrong

By JOHN CLOUD

The new SAT scores are out, and buried in them is a sign of hope for American education. True, the scores are actually a bit lower than last year's; the combined average for the SAT's math and reading sections fell 7 points, to 1021, the biggest single-year decrease since 1975, when the score dropped 16 points, to 1010. But statistically speaking, a 7-point decline (out of a possible 1600 on those two sections) isn't much. It's less than the value of a single question, which is about 10 points. Also, **the SAT was radically changed last year. The College Board made it longer and added Algebra II, more grammar and an essay. Fewer kids wanted to take the new 3-hr. 45-min. test more than once, so fewer had an opportunity to improve their performance. Scores were bound to slide.**

But tucked into the reams of data the College Board included with the new scores was some wonderful news: I was wrong. In 2003 I spent six months tracking the development of the new SAT. I sat through hours of test-development sessions and even learned how to grade SAT essays. TIME ran my resulting story on its cover that October.

The story did make some predictions that turned out to be right. For instance, **the new test favors girls** more than the old one did. It is a long-standing tenet of testmaking that girls outperform boys on writing exams. For reasons I am not foolish enough to speculate about in print, girls are better than boys at fixing grammar and constructing essays, so the addition of a third SAT section, on writing, was almost certain to shrink the male-female score gap. It did. **Girls trounced boys on the new writing section, 502 to 491. Boys still outscored girls overall, thanks largely to boys' 536 average on the math section, compared with girls' 502. But boys now lead on the reading section by just 3 points, 505 to 502; the gap was 8 points last year. What changed? The new test has no analogies ("bird is to nest" as "dog is to doghouse"), and boys usually clobbered girls on analogies.**

My story also predicted that the addition of the writing section would damage the SAT's reliability. Reliability is a measure of how similar a test's results are from one sitting to the next. The pre-2005 SAT had a standard error of measurement of about 30 points per section. In other words, if you got a 500 on the math section, your "true" score was anywhere between 470 and 530. But the new writing section, which includes not only a multiple-choice grammar segment but also the subjective essay, has a standard error of measurement of 40 points. That means a kid who gets a 760 in writing may actually be a perfect 800--or a clever-but-no-genius 720. In short, the College Board sacrificed some reliability in order to include writing.

Finally, I was right about one other thing: that the **graders would reward formulaic, colorless writing over sharp young voices.** The average essay score for kids who wrote in the first person was 6.9, compared with 7.2 for those who didn't. **(A 1-to-12 scale is used to grade essays. That score is then combined with the score on the grammar questions and translated into the familiar 200 to 800 points.) As my editors know well, first-person writing can flop. But the College Board is now distributing a guide called "20 Outstanding SAT Essays"--all of them perfect scores--and many are unbearably mechanical and clichéd ("smooth sailing always**

comes after the storm"; "they say that history repeats itself").

Still, there's good news. The central contention of my 2003 story was that the SAT's shift from an abstract-reasoning test to a test of classroom material like Algebra II would hurt kids from failing schools. I was worried that the most vulnerable students would struggle on the new version. Instead, the very poorest children--those from families earning less than \$20,000 a year--improved their SAT performance this year. It was a modest improvement (just 3 points) but significant, given the overall slump in scores. And noncitizen residents and refugees saw their scores rise an impressive 13 points. **It was middle-class and rich kids who account for the much reported decline.**

What explains those wonderfully unpredictable findings? The College Board has no firm answers, but its top researcher, Wayne Camara, suggests a (somewhat self-serving) theory: **the new SAT is less coachable.** When designing the new test, the board banned analogies and "quantitative comparisons" (flummoxing math questions that asked you to compare two complex quantities). "I think those items disadvantaged students who did not have the resources, the motivation, the awareness to figure out how to approach them," says Camara. "By eliminating those, the test becomes much less about strategy." **Because it focuses more on what high schools teach and less on tricky reasoning questions, the SAT is now more, not less, egalitarian.**

Sometimes it's nice to be wrong.