

U.S.

Tutors See Stereotypes and Gender Bias in SAT. Testers See None of the Above.

By ANEMONA HARTOCOLLIS JUNE 26, 2016

In an annual ritual, hundreds of thousands of students took the SAT this spring as they made their first steps toward applying to college. But they were not the only ones being tested.

Sprinkled among them in May, when the SAT was given for the second time since a much-ballyhooed revamping, were a number of people long past college — members of the test-prep industry who took the exam to see how those changes played out in practice so that they could improve their tutoring services. Armed with perhaps sharper pencils and a more jaundiced eye than the typical 17-year-old, they noticed two questions that some thought could throw off the performance of girls.

Their concerns are fueling a debate in the industry — on Facebook and in private emails — over whether the test items were sufficiently vetted for gender bias, and whether the exam was unfair to female students.

The two items, one in the verbal portion and one in the math section, posed what some test-prep experts considered a textbook example of “stereotype threat.” When people are reminded during a test of a negative stereotype about their race or sex, psychologists say, it creates a kind of test anxiety that leads them to underperform.

The math question involved a chart showing more boys than girls in math classes over all.

The verbal section asked students to analyze a 19th-century polemic arguing that women's place was in the home.

Officials from the **College Board**, which administers the test, said it had not presented negative stereotypes. They said the revamped SAT had been, if anything, more thoroughly vetted for fairness than older iterations. Content had been reviewed by experts and pretested on a representative sample of students.

The results of the May 7 test have also been analyzed. No differences in the scores of boys and girls of comparable ability were found on the questions in dispute, College Board officials said.

“This means the questions did not present an unfair advantage to either group,” Kate Welk, the director of assessment communications, said in an email.

But some test-prep tutors do not see it that way. “I thought, ‘Wait a minute: This test is really trying women in a way that’s slightly different than it’s testing men,’” said Sheila Akbar, the education director for Signet Education, a test-prep company. “Here I am, a seasoned test taker, a 36-year-old woman, being distracted by this material. I wonder what 17-year-olds are thinking.”

The standardized testing industry has long grappled with questions of racial, socioeconomic and gender fairness. On average historically, whites and Asians do better than blacks and Hispanics on the SAT; wealthier children do better than poorer ones; and boys do better than girls — slightly better in verbal skills and considerably better in math. It is unclear why these gaps persist, test designers and social scientists say, but the differences are the subject of much study, including the research on negative stereotypes.

There have been hundreds of studies on the effects of stereotypes in test situations, but the testing industry says that effects on performance have not been well replicated in real life.

Board officials have said that the test's overhaul, the **biggest in a decade**, was

done to make the test more relevant to what children learned in school — and more fair.

The reading item paired 1837's "Essay on Slavery and Abolitionism" by Catharine E. Beecher with an 1838 reply from Angelina E. Grimké, an abolitionist. It asked questions about how to interpret the passages.

The Beecher essay argues, beginning around Section 99, that by divine law, women have a lower station than men and wield their influence through the domestic sphere.

Grimké argues that no one's rights should be diminished because of her sex.

Christian references in the Beecher passage were edited out, suggesting to some critics that the board did not want to offend religious sensibilities. But Zachary Goldberg, a board spokesman, said the religious references had been edited out for "length and focus." He described Beecher's essay as "separate-spheres feminism."

The questions related to the passage were not the problem, critics said. It was the placement of the passage at the beginning of the test, allowing it to linger in students' minds for the rest of the time.

The math question was identical to one on a practice test, which showed a chart of 10th graders in which there were more boys than girls taking math.

The two SAT items appeared to be classic examples of situations that might trigger stereotype-driven test anxiety, said Joshua Aronson, an associate professor of applied psychology at New York University who, with Claude Steele, pioneered the research on the subject in the mid-1990s.

"I'm not saying we should put everybody in a rubber room so they couldn't possibly be touched by controversy," Dr. Aronson said. "But why would you go out of your way to couch a percentage problem as a girls-in-math problem?"

The reading passage, he said, could induce "cognitive fatigue" for girls who were bothered by it. "You could imagine one girl really ruminating on it, and she would pay for it down the road," he said.

Andrew McGlothlin, the owner of Summit Tutoring and Test Prep in Los Angeles, compared the reading item to pairing a segregationist and an integrationist. “We can’t have George Wallace being given equal voice to Thurgood Marshall,” said Mr. McGlothlin, who took the test in May. “That would be beyond the pale.”

“I think you can defend it as a historical argument, but there was no historical perspective on it,” Mr. McGlothlin said.

Pamela Donnelly, the founder of GATE, a college consulting and test prep company in Los Angeles, said her daughter, Lily Berthold, took the test in May and remembered the Beecher passage as “weird.”

“I’m a mother of three daughters,” she said, “and to present something like that without historical context creates a kind of cognitive dissonance between that and what they believe is true as they try to wrestle their way through the rest of the test.”

Cynthia Cowan, a Boston-based tutor, said she had a male student who had found Beecher’s argument so alien that he was sure he had misunderstood it. “This is a good student who read the piece right and convinced himself that he had read it wrong,” she said. “If it has that strange effect on a guy, who knows what effect it can have on women?”

Both positive and negative reinforcements can work in subtle ways, experts say. For example, some research has found that boys got a boost just from being asked to mark down their sex before taking the test, as if it helped to be reminded that they were boys, Dr. Aronson said.

In the 2013-14 school year, the College Board stopped asking students to mark their sex on the SAT answer sheet; a spokesman said this was done to save time.

In an intriguing twist, Dr. Aronson said, recent studies have found that girls are not as aware of female stereotypes as women were when the original research was done. “What I think is in part responsible for that is that in the last 25 years, the gender roles for boys and girls have totally reversed,” he said. “Girls are the good students now; boys are a mess.”

“If there’s no difference, so be it,” he said. “I’m certain that it bothers some girls

and not others. In fact, that's where the stereotype research is going.”

A version of this article appears in print on June 27, 2016, on page A11 of the New York edition with the headline: Tutors See Stereotypes and Gender Bias in SAT. Testers See None of the Above.

© 2016 The New York Times Company