

March 12, 2003

## When a Passing Grade Defies Laws of Physics

By MICHAEL WINERIP

THIS is the era of make-or-break testing, and no one has been more aggressive in leading that charge than New York State's education commissioner, Dr. Richard Mills. He has put in effect a system that requires high school students to pass five state Regents examinations to graduate, and has fought anyone in his way.

A few years ago, when principals from 30 highly regarded alternative public high schools asked to continue using a portfolio assessment system, rather than the state tests, Dr. Mills rejected their request and successfully battled them in court.

Such alternative assessments, Dr. Mills argued, were not rigorous or scientific.

Well, anyone who thinks there is anything scientific about the way the state decides who will pass its high school exams needs to take a close look at the behind-the-scenes shenanigans that went into determining the passing grade for the state's Regents physics test introduced in June 2002.

The state hired a testing consultant with 35 years of experience, Gary Echternacht, then ignored his findings on what the passing physics grade should be. The state hired a panel of 28 veteran physics teachers, then discarded their recommendations. State officials were worried that these hired experts were setting a physics passing score that was too easy. They were worried that they would have a repeat of 2001, when the state was criticized for setting too low a passing score on the earth science and biology Regents tests.

So Gerald DeMauro, the state's top testing official, took matters into his own hands. Under his direction, the passing score was raised far beyond what Mr. Echternacht and the panel of 28 physics teachers thought reasonable. "Way too high," said Mr. Echternacht, who has worked on setting the scoring scale for most New York Regents examinations in recent years.

Jim Guido, a physics teacher from Williamsville with 31 years of experience who was part of the state panel, said, "It was pure voodoo, the way the state picked that passing score."

And Heather Dunshee of Sherburne, another teacher on the panel, said, "They asked our opinion, and when they didn't get what they wanted, they disregarded it and came up with a completely different number."

The results of this "voodoo" scaling? The loudest outcry since Dr. Mills brought in the new testing system. Physics students are typically the best and brightest, and yet in June 2002, 33 percent failed the state test, compared with 11 percent the year before. Only 16 percent scored distinction (85 or better), compared with 33 percent the year before. Did New York physics students suddenly get stupid? Not according to their teachers, who reported that the same students scored as well as ever on other physics exams — like the SAT II and AP tests.

"It's obvious the tests were scaled wrong," William Johnson, Rockville Centre's superintendent, said.

For weeks, parents and school officials flooded the state with complaints. Some districts, including Hewlett-Woodmere, omitted the physics score from student transcripts. Last fall, for the first time ever, the state's Council of School Superintendents sent a letter to college admissions offices urging them to disregard a Regents examination, saying, "We believe the physics Regents exam grade to be suspect."

Through it all, Dr. Mills has maintained that the state made no mistakes, that the higher failure rate was a result of tough new academic standards. In response to pages of questions I submitted, Tom Dunn, a state spokesman, sent a four-paragraph response by e-mail.

"The development of the physics examination, including setting of the cut scores, was accomplished in accordance with nationally accepted standards," Mr. Dunn said in his e-mail message. "The process was conducted in a statistically sound manner."

He said state officials would not answer questions because of pending litigation. (Fifteen districts have sued to have the test rescaled. A lower court recently dismissed the suit on the ground that the districts did not have standing to sue, but the districts may appeal.)

People may assume that a state test has 100 questions and if you score 65, you pass; score 85, you make distinction. The scoring is nothing like that. The state uses a scaling method known as bookmarking. On March 4, 2002, at 8 a.m., 28 physics teachers gathered in a conference room in Albany to help set the passing and distinction marks.

The group was given several pages of physics questions that had been field-tested on students in recent years. The questions were ordered from easiest to hardest. Gary Echternacht, the testing expert hired by the state, instructed teachers to think of a "mythical Gary," a student who deserved to score a 65. They then were to study the questions, from easiest to hardest, and find the last question that this mythical 65 student ought to be able to answer. That question would mark the cut between passing and failing.

For eight hours, through six rounds of tabulations, the 28 teachers debated which question should mark the cut line. Mr. Echternacht says it is natural when thinking of a mythical student that teachers envision their own classes, and so he warns state officials that the teacher panel should be representative of New York.

It was not. Although 34 percent of state high school students are from New York City, only two panelists were. No teachers were from Buffalo, Rochester, Syracuse, Albany or Yonkers. And 6 of 28 were from nearby, suburban Albany. When I told this to Mr. Echternacht, he said: "I didn't know that. That's not good."

I was able to reach 10 of the 28 teachers, and almost all (including Mr. Guido, Ms. Dunshee, Richard Freyer of Voorheesville, Libby Hokanson of Niskayuna, Charlene Rydgren of Malone, Joyce Kelly of Briarcliff Manor) said they thought the process and the results were reasonable and would be used to set the scale for the June test. They also said that state officials present that day were clearly upset that the panel had arrived at too low a passing score. What to do when you do not get the numbers you want the first time? Try again. At the end of April, state officials convened a second panel made up of 9 of the original 28. Precisely how this panel was picked is unclear, but several of the original 28 were not invited back. Nor was Mr. Echternacht. He says he was never even told there was a second panel. Asked if he felt this undermined the process, he said, "I wouldn't do that."

This time, Dr. DeMauro himself led the teachers, and several said he steered them to picking higher cut marks. "The process the second time was different," said Andrew Telesca Jr., a teacher from Johnson City. "I knew the cut score would be higher."

At the end, Mr. Telesca said, teachers asked what the passing score would be, and Dr. DeMauro would not tell them. "He just said it's going to be higher," Mr. Telesca said.

In mid-June when panel members saw the scoring sheets for the new physics exam, they were stunned at how much higher the cut marks had been set. Many had previously been hired by the state to help develop the curriculum that the new physics test was based on. They had worked hard to prepare their students and still reported an unprecedented number of failures.

"I've taught 18 years and had more fail that test than in all my previous years combined," Ms. Dunshee said. "The passing score was so far above what we had picked."

Ms. Kelly, a 15-year veteran who sat on the first panel, said she felt used and manipulated. "Why did they ask my opinion if they weren't going to use it?" she asked. "I will never go to another thing run by the State Education

Department."

Mr. Echternacht, who for 30 years worked for the Educational Testing Service, which developed the SAT's, said the root problem was that the state was introducing too many new tests too quickly, resulting in unreliable pass/fail scores.

"It's naïve to think you'll create a new test that comes out of the box fine and dandy," he said. "Setting a pass score is not science. It's a messy, subjective process. I find 20 to 30 percent of the time, the passing cut mark is too hard or too easy. A new test always has problems."

Mr. Echternacht said he recommended that the new state tests be phased in over several years.

In the late 1980's, the last time New York introduced a physics test, state officials spent three years phasing it in, said Mr. Telesca, a past president of New York's science teachers' association. The new version was given in a few school districts for the first year, then scores were curved to correct scaling errors. In the second year, the new version was given to a larger group, and scores were again adjusted afterward. In the third year, districts were given the option of taking the old test or the new. Not until the fourth year did everyone take the new version.

In 2002, the state mandated that the new physics test be used for everyone right away and refused to make scoring adjustments.

Mr. Echternacht said he told state officials that they could avoid the wild fluctuations in their scaled scores, from too easy (biology and earth science) to too hard (physics), if the tests were phased in like the physics test of the 1980's. Their response? "They don't respond," he said.

They did not respond to me either.