

## Assessing the Assessments

Has Oregon “Dumbed Down” Its Tests?

By Greg Perry

March, 2003

You ever had one of those hunches about something? You dismiss it at first because it doesn't seem possible. But then it keeps coming back again and again until, finally, you have to check it out.

Well, I had one of those hunches about two years ago. It had to do with the Oregon State Assessment Tests (OSAT)--the tests that Oregon public school students have to take as a part of Vera Katz's School Reform Act. My hunch was that Oregon's Department of Education has wasted a lot of money developing an assessment system little different from standardized tests that could be purchased much more cheaply from any one of a number of private firms.

Oregon's state-created assessments provide very little useful information to parents and teachers. But they cost a lot of money to develop and implement. As a fiscal conservative and taxpayer, I wondered whether we as citizens were getting the “best bang for our buck.”

When I started looking into it, little did I realize that I would uncover a much more controversial problem—Oregon's tests have apparently been “dumbed down.” The tests in recent years appear to be significantly easier than earlier versions.

But let me start at the beginning. I am the parent of a fifth grader in a terrific public school in Corvallis called Franklin School.

Franklin School is a thriving public school of choice that started in the early '90s and now enrolls more than 400 kids in a demanding, content-rich curriculum. Everything everybody says they want in a public school we have at Franklin: high standards, involved parents, excellent teachers, and high achieving students. I've been involved with the school since 1995, and I work at Oregon State University as an agricultural economist.

An important part of our school was annual assessments of all students. We wanted assessments in every grade not limited to reading and mathematics, but also to include language use, spelling, social studies and science. The assessments would be used to track the progress of students over time and also to help teachers identify strengths and weaknesses in their programs.

Our principal consulted with the Corvallis district assessment expert, Bill Auty, and together they decided that the best tool was the CTBS, now known as the Terra Nova, perhaps the mostly widely used assessment program in the United States. We decided to administer the CTBS each spring to students in grades 2-8.

Like other public schools, we also take part in Oregon's statewide assessments, which are required for students in 3<sup>rd</sup>, 5<sup>th</sup> and 8<sup>th</sup> grades. So our students have taken both tests for several years. Thus, we are one of the only schools in Oregon with data for the same

student from both Oregon's tests and a nationally-normed, standardized test of academic achievement.

As you might imagine, the time needed to conduct both of these assessments is significant. Teachers expressed a lot of frustration with the time spent, in particular on the OSAT. They questioned the validity of the mathematics problem-solving exam and the need to develop reading and mathematics standardized tests when the Terra Nova exams seemed to provide the same information. They commented on how well the state paid teachers in the summer to grade the problem-solving and writing assessments, but wondered if the state might better use that money by passing it down to the school districts to pay for books and supplies. And, in examining the scores over time, they began to wonder aloud whether the state might be "dumbing down" the newer exams.

In the late 90's, I too noticed something rather strange from our test scores. As an economist, I pay attention to empirical data.

Our students' scores on the Terra Nova were consistently excellent, although they had the normal year-to-year fluctuations one would expect. But our students' scores on Oregon's statewide assessment had a much different pattern. There was a sharp rise in scores starting in the late 90's that was not evident in our Terra Nova scores.

What was going on? If scores on one test were going up quickly while scores on another did not, there had to be a reason. Since I am trained to analyze trends, correlations and anomalies in sets of data, I decided to take a closer look.

Was this same trend—a sharp rise in student scores on Oregon's tests in the late 90's—happening statewide? Was there a good explanation for it? Was it because achievement levels were actually rising? Or had Oregon's tests gotten easier?

The scores from Franklin School suggested the latter, but until I put the data through the grinder to test for statistical significance, that's all it was—a suggestion.

### **So I tested the hypothesis: had Oregon's tests gotten easier?**

Obviously, this is not a trivial question. If true, it would reflect pretty poorly on the state's claim that Vera Katz's School Reform Act was working. The Oregon Department of Education website tells us that achievement is way up: far more kids are meeting the reading and math testing standards than in the early nineties.

To investigate this I needed statewide test scores from 1991 to present, data only available from the Oregon Department of Education (ODE). After months of hassling and many futile phone calls, I finally filed a Freedom of Information Act request and they were required to give me the data.

I graphed the average state scores in reading and mathematics in third, fifth and eighth grade from 1991 to 2001. Eighth grade scores trended upward in a modest zigzag pattern, similar to the pattern exhibited by Oregon average scores on the SAT. Nothing out of the ordinary there.

Then a strange pattern emerged. The scores for third and fifth grades varied up and down from 1991-95, with a small upward trend. But beginning in 1996, the scores jumped up

every single year, at an accelerating pace. An odd result, not consistent with the zigzag pattern one would expect.

What could explain this result? Maybe student achievement was in fact suddenly increasing at a much faster rate. Maybe Vera Katz's school reforms are working. But the reforms began in 1991, so why would test scores suddenly accelerate after five years? And why such robust gains after 1996?

### **Had the tests gotten easier?**

There are several ways to determine whether or not assessment exams are being "dumbed down" over time. One way is for assessment experts to examine the test questions over several years to see if questions have been simplified. This was done in Texas. Experts there found that the difficulty of the reading assessments had in fact declined from 1995 to 1997, which accounted for almost all of the test score increases.

### **Had the same thing happened to Oregon's tests?**

I sent a formal request for copies of the OSAT third, fifth grade and eighth grade exams from 1996 to 2001 under the Freedom of Information Act.

Bill Auty (who was now in charge of the Oregon Department of Education's assessment and evaluation program) refused my request, saying that the state reuses questions from past exams to save money. He said I would be allowed to view the exams at the offices of ODE, but I could not make copies.

I countered with a proposal: They could release copies of the test to me with permission to share them with selected experts in the assessment field. I, in turn, would sign a document that subjects me to legal action if I release the tests to anyone other than those experts.

My request was again denied. Mr. Auty conceded however, that if we used the same techniques as the Texas analysis to look at Oregon's reading tests, "...you would probably see differences in the difficulty of reading passages. I suspect that you would find the difficulty has varied randomly. However, even if you discovered a trend in difficulty variation, it would be inaccurate to conclude that the student or school results are affected by that trend..."

I shared this response with an expert involved with assessments in another state who was very skeptical of Auty's explanation and of the agency's refusal to release exams to the public. In her view, releasing the tests each year was a good way to keep their state's assessment program "on its toes."

My suspicion deepened that some type of mischief was afoot in the state's assessment program. But without access to the old tests, how could I determine if the OSAT scores were being inflated? This question led me back to the Franklin data.

My first step was to confirm from a same-year statistical analysis that the OSAT reading and mathematics assessment tracked closely with their Terra Nova counterparts, which proved true. Most of the time, Franklin gave the exact same Terra Nova test year in and year out, so we knew its difficulty had not changed.

Since we had several years of data in which students took both tests, if the OSAT was getting progressively easier, the inflated scores would be revealed by a multiple regression analysis.

After pooling the data for both the Terra Nova and the OSAT for the years 1996-2001, with special variables representing each year, I ran the regressions. If there was any inflation in the scores, these variables would be both positive and statistically significant.

The results for third and fifth grade reading exams showed a small amount of inflation in the OSAT over time, but not enough to be able to say with a high degree of probability that the effect was statistically significant. The results on the mathematics test, however, were consistent, and statistically significant.

At the third grade level, the OSAT was apparently inflated by five points from 1997 to 2001. At the fifth grade level, the scores were inflated by almost 15 points from 1996 to 2001.

To put these results in perspective, note that the OSAT mathematics average score statewide increased by seven points from 1996 to 2001. If the Franklin results are an accurate prediction of the results statewide, it suggests that the state scores actually declined by eight points from 1996 to 2001.

What might explain this result? One possible explanation is that teachers are teaching to the test, in this case the OSAT. However, the two tests track each other so closely that any effort by teachers to boost OSAT scores would almost certainly boost the Terra Nova scores.

Another explanation is that the assessment experts at ODE have done a poor job of equating tests from year to year. This might explain the results for the reading assessment, because the differences do vary from year to year. However, for the mathematics assessment, the gap between the OSAT and Terra Nova widens each year for third and fifth grade students. This kind of trend seems deliberate, not random.

Another explanation is that the results may be simply a fluke, a product of a very small data set in a special situation. However, when a result is statistically significant, it means that there are some relationships in the data that are consistently pointing in the same direction. The chances of these results being entirely random are pretty small.

To me, the most plausible explanation is that Oregon's math assessment has simply gotten easier.

Another way to check the gains on Oregon's test scores is to see if they are validated by other tests Oregon students take. There are two such tests: 1) the SAT—a college entrance exam, and 2) the NAEP (National Assessment of Education Progress). If Oregon student achievement has really increased as much as scores on Oregon's state tests claim, we would expect to see similar increases in our scores on NAEP and SAT.

Since 1995, when the SAT test was "recentered," Oregon's average SAT score has risen by 4 points on the math test and 3 points on the verbal test. But nationwide, average scores climbed by 8 and 5 points respectively.

On the NAEP, we get similar results. Oregon's gains lag the national average gains in both reading and math. The conclusion: the dramatic gains on Oregon's statewide assessment since the mid 1990s are not confirmed by either the SAT or the NAEP.

Oregonians have been told that student academic achievement is on the rise, that CIM-CAM reform has improved the schools. But if the rise in test scores is an illusion—either purposeful or accidental—if the tests have been dumbed down, it is a shocking betrayal of public trust.

The state has spent more than a decade and millions upon millions of dollars in an effort to implement a far-reaching reform of Oregon's school system. Student test scores are the "audit"—the primary instrument that measures whether or not this reform is working.

We don't allow corporations to audit themselves, yet here in Oregon nobody blinks an eye when the state is given the authority to run its own assessments. Enron should have taught us something. The state should not design and administer the tests that constitute the audit of their own reform efforts. There are plenty of private, independent organizations that not only do a better job of developing tests, but they also do it more cheaply, and without the inherent conflicts of interest.

Did the state dumb down the tests? They had both motive and opportunity.

The analyses I conducted with the data do not constitute absolute proof—statistics cannot prove a hypothesis, they can only rule out possible explanations. And the assessment experts at the state will surely deny it. But when they do, there is a simple and definitive way for them to prove their case to the taxpayers and to students and parents in Oregon.

Select 500-1,000 students in the 3<sup>rd</sup>, 5<sup>th</sup> and 8<sup>th</sup> grades that reflect the diversity of Oregon students. On Monday give them Oregon's reading and math tests, 1996 version. On the following Monday give them Oregon's reading and math tests, 2002 version. The results should be very close. If the 2002 scores are substantially higher, the tests have been dumbed down.

Just be sure that someone other than the Oregon Department of Education conducts the experiment.

Greg Perry is a co-founder of Franklin School. He lives in Corvallis and is the father of six children.

The full text of Greg Perry's paper, "A Longitudinal Analysis of the Oregon State Assessment Tests 1991-2001," can be found at [www.oregoneducation.org](http://www.oregoneducation.org)

### **SIDEBAR: What is wrong with Oregon's Assessment System?**

- 1) High Costs: There are large fixed costs associated with developing assessments, but the variable costs of administering and scoring most tests is quite low. Why would Oregon want to develop its own assessments from scratch when many outside firms have already borne those fixed costs?
- 2) No Quality Control: Other than assurances from ODE employees, there is no way to independently determine the quality of the OSATs.

- 3) Conflict of Interest: Increases in test scores are the primary evidence for ODE assertions that the education reforms, which they are charged with implementing, are working. Good scores imply good job performance by ODE staff. Obviously, they have a direct interest in making sure scores improve—a clear conflict of interest.
- 4) Lack of Accountability: The state assessment program has an appalling lack of public oversight. The state develops all its own tests, makes no attempt to track them against other common assessment tools, hand picks any outside reviewers of their work, and greatly restricts public scrutiny of the tests.
- 5) Subjectivity: In a quest for the holy grail of assessment—measuring nebulous skills such as “critical thinking” and “problem solving,”—Oregon has developed a myriad of assessments such as writing sample tests, math problem-solving tests, and portfolio assessments. These faddish assessments are costly, cumbersome, lack reliability and validity, and take too much time away from instruction.