PUBLIC POLICY
RESEARCH CENTER

JANUARY, 2001

POLICY BRIEF 3

VALUE ADDED TESTING:
A MODEST PROPOSAL FOR EDUCATION REFORM

By: **David C. Rose**, Associate Professor, Department of Economics and Public Policy Research Center Fellow

All people learn over time. What schools are supposed to do is increase the rate of learning.

University of Missouri St. Louis

8001 Natural Bridge Road
St. Louis, Missouri 63121-4499

POLICY BRIEFS AND OTHER PPRC MATERIALS ARE PUBLISHED ON AN OCCASSIONAL BASIS

FOR MORE INFORMATION

(314) 516.5273 TEL
(314) 516.5268 FAX

pprc@umsl.edu EMAIL
www.umsl.edu/services.pprc/index.htm

Other policy briefs include:

1. THE RISE IS DUAL EARNER FAMILIES
2. GENDER DIFFERENCES IN PAY AND PAY RAISES

As concern over failing schools continues to grow there are increasing calls for more accountability of educational performance in the form of standardized testing. There is little doubt that more accountability is desirable. Understandably, however, teachers' organizations have expressed concern over the use of standardized tests since teachers do not control the sample over which they are judged. Teachers also argue that standardized tests produce pressure to "teach to the test." Obviously, prudent public policy requires that much thought be given to how we administer standardized tests and how we interpret their results.

The purpose of this policy brief is to suggest to policy makers that many of the problems associated with standardized tests are products not of the tests themselves but how they are interpreted. What follows is a testing policy proposal that can implemented at the school, district, and even state level. This proposal provides an overview of a new way of interpreting standardized test data that will provide teachers and administrators with information that is both easy to interpret and directly pertinent to achieving the true goal of education, which is to maximize the academic achievement of every child regardless of innate ability, socioeconomic status, or ethnic background. In a nutshell, the key is shifting focus away from average test scores over the entire school or school district sample and focus instead on the change in scores over time on a student-by-student basis.

All people learn over time. What schools are supposed to do is increase the rate of learning. The ultimate standard by which schools should be judged is therefore how much they increase the rate of learning. Schools that achieve high rates of learning are better schools than those that achieve low rates of learning. Put another way, a grade school that starts with 95th percentile children (on average) and finishes, five years later, with 90th percentile children has nothing to be proud of, despite being a school comprised of 90th percentile fifth graders (on average). This is because the rate of learning needed only be that of an average school to maintain the initial percentile score of 95. Indeed, going from the 95th percentile to 90th percentile means that the rate of learning is actually below that which occurs in average schools. On the other hand, a different grade school that starts with 50th percentile children and finishes with 60th percentile children should be commended, since its students have learned at a rate that exceeds that of average schools.

Social scientists, statisticians, and educators know that unadjusted average test scores can be very misleading. We all understand that it is unfair to judge a school or a teacher by absolute academic achievement when student abilities are subject to significant cross-sectional variation, so comparative assessments normally attempt to control for these factors through the use of socioeconomic proxies, COGAT scores, and the like. While this is clearly a step in the right direction, why not go further? Instead of using proxies to control for the aforementioned factors, why not compute the year-to-year change in standardized test scores on a student-by-student basis to directly measure whether students are learning at a rate that is faster, slower, or about the same as the average child in the reference sample? With such information it would be easy to compare the rate of improvement of students in any given school to the rate of improvement in other schools, both inside and outside of the district.

Consider a 4th grade class filled with 20 students, where 15 were in the district in the previous year. Each of these students will have a standardized test score on file that will tell us his/her percentile score on the test from the previous year. When the results of the 4th grade test are computed, we can compare each child's percentile score to the previous year's score. If, for a given student, the percentile score rose by a statistically significant amount, then we can be reasonably sure that this child has learned at a faster rate than the average child in the reference sample. Many children, of course, will have a percentile score that is not statistically significantly different from the previous year's score. In this case we should infer that there is no evidence that the child is learning at a faster or slower rate than children do on average *in the reference sample and, hence, in average performing schools*. Some students, however, will have scores that have increased by a sufficient amount to constitute a statistically significant change, and some will have scores that decreased by a sufficient amount to constitute a statistically significant change.

Suppose that in our hypothetical class 8 of the students have percentile scores that are statistically indistinguishable the last year's score, 5 have percentile scores that are (statistically) significantly above last year's score, and only 2 have percentile scores that are (statistically) significantly below last year's score. From this information the principal could easily tell that this class is doing quite well *even if we had no idea what the absolute scores are*. This is because relative performance is rising for many more students than for whom it is falling. There is no need to transform the data with proxies for socio-economic status or cognitive ability because we are looking at each child in isolation relative to his/her own previous scores. In statistical terms, because we have first-differenced the data, cross-sectional correlates of these proxies (fixed effects attributes) vanish.

A valid criticism of using standardized test results to infer performance is that there is more to learning than what is covered on standardized tests. The first-difference approach proposed here directly addresses this criticism. Suppose a school spends 13 hours per week on subjects like creative writing, art, theater, etc., and that this has the effect of reducing the average standardized test score by 10 percentile points because these skills, important as they are, are not measured on standardized tests. When we first-difference the data, we find that this effect vanishes because it is, statistically, a fixed effect. Put simply, the 3rd grade score was marginally lowered by 10 points due to time spent on non-testable academic activity and the 4th grade score was marginally lowered by 10 points as well. Any change in the score from 3rd to 4th grade can therefore not be due to these effects because the minus ten in 4th grade minus the minus ten in 3rd grade nets to a zero marginal effect. This is a common feature of real world data and explains the popularity of fixed effects modeling by statisticians.

From this information the principal could easily tell that this class is doing quite well even if we had no idea what the absolute scores are.

This approach requires no additional test data, no fancy data transformations, and no arbitrary controls that are, at best, mere proxies. Since the methodology works for any percentile score, teachers and administrators can also look at the time path for scores in the math, language arts, science, and social studies sub-sections of the test. Most importantly, test scores that are reported in this way will directly measure what we really want to know: the presence or absence of improvement in performance over time. Below is a summary of the basic methodology:

1. Divide the set of all students in a given (classroom, school, or district) that were present the previous year into the following three groups for the complete battery and each of the four subject areas covered by our standardized test instrument: 1) those students who did statistically significantly worse than last year, 2) those students who did not have a statistically significant change in their scores, and 3) those who did statistically significantly better than last year.

2. Compute for the (class, school, or district) the ratio of students whose scores increased significantly to those whose scores decreased significantly for the complete battery and for each of the four subject areas.

3. Compute for the (school or district) item (2) above using the percentile scores for the first and last years of attendance (e.g., for grade schools compare a child's 1st grade score to his 5th grade score).

One valid complaint about placing a lot of emphasis on standardized test scores is the absence of a level playing field.

A reasonable question at this point is this: why bother with statistical significance? Why not just compute the mean of the first-difference values? The reason for categorization by statistical significance is that this automatically adjusts for the fact that even a small improvement in the percentile score of a child who is already in the 95th percentile is quite difficult to achieve because we are already in the tail of the distribution. Similarly, it is almost inevitable that a child starting at the 2nd percentile will realize a higher score next time. By only counting statistically significant improvements, we account for this issue.

With such an approach, teachers will no longer be penalized for teaching outside the scope of standardized test material because we will no longer focus on the absolute test score average when assessing teachers or schools. In addition, teachers will no longer need no longer worry that their measured performance will be a function of how classes were allocated. The innate ability of students and the quality of instruction they received in prior years will no longer be an issue when assessing teachers since all that matters is the rate at which student performance improves over the year a given student is in his/her class.

One valid complaint about placing a lot of emphasis on standardized test scores is the absence of a level playing field. Schools play an important role in determining which children take the test and the baseline data that is used to norm the test against socio-economic factors. This puts ethical schools at a disadvantage since unethical schools can enjoy higher scores than they deserve. The value added eliminates this problem entirely because there is no "norming" and it doesn't matter who takes the test.

There is much more to this proposal that has been discussed here. The bottom line is that the proposed reporting policy will shift the focus of assessment to where it should be — educational value-added. This approach should not be very expensive and there is no single policy change that would more strongly signal the community that our state is serious about improving its academic standards. Instead of rewarding luck and privilege, it rewards improvement that is generated by what schools do, not how gifted their students are.

FOR FURTHER READING

"How does the pendulum swing on standardized testing?" by Jean Baresic. *The Education Digest*, January 2001; Vol. 66, No. 5, 12.

"The Disciplined Mind: Beyond Facts and Standardized Tests," by Dudley Barlow. *The Education Digest*, December 2000; Vol. 66, No. 4, 75.

"What's at stake in high-stakes testing: Teachers and parents speak out," by Mary Alice Barksdale-Ladd. *Journal of Teacher Education*, November/December 2000; Vol. 51, No. 5, 384-398.

"Selling higher test scores," by Edward Cohn. *The American Prospect*, October 23, 2000; Vol. 11, No. 22, 29-32.

Standardized Minds: The High Price of America's Testing Culture and What We Can Do to Change It by Peter Sacks. Perseus Books, 2000.

Contradictions of School Reform: Educational Costs of Standardized Testing by Linda M. McNeil. Routledge, 2000.

Will Standards Save Public Education? by Deborah Meier. Beacon, 2000.