

**When Learning no Longer Matters – standardized testing and the creation of inequality.**

*Phi Delta Kappan* (2003), March, 84 (7), 502-506.

Jo Boaler, Stanford University.

Abstract: This article considers the inequities that are created by standardized tests. It describes the story of Railside – an urban low-income high school in which the students made incredible achievements, only to be labeled ‘under performing’ by the state of California. It is argued that the standardized test upon which this judgment was based, and similar tests used throughout the country, stack the deck against language learners, and students from minority ethnic and cultural groups and low-income homes.

This is a story about a remarkable school, that has been labeled “under-performing” by the State of California. It is a true story and it draws from a combination of research data, collected as part of a Stanford University research project on mathematics learning, and the lived realities of teachers and students working hard to achieve success in an urban, low-income school. At the heart of this story lies the conflict between learning and SAT-9 success, a conflict that has impacted the lives of students and teachers in this school in profound ways.

As a new professor at Stanford, recently arrived from England, I was considering schools to include in a study of mathematics teaching and learning. Railside High<sup>1</sup>, I soon learned, had a mathematics department that worked in very unusual ways. Some years ago they de-tracked their classes, in response to the low performance of some students. The mathematics department plans lessons collaboratively, and teachers meet every week to discuss and improve their lessons. They visit each others’ classes frequently, in order to learn from each other, and every new teacher is given the opportunity to watch every lesson s/he will teach being taught by an experienced colleague, first. The algebra curriculum that all students take on entry to the school was designed by the department and it draws from a variety of different curriculum materials. Students from a range of ethnicities – mainly Latino, African-American, White, Asian, and Filipino – work together in groups, solving complex mathematics problems. All the teachers in the department are mathematics specialists and they all regularly attend professional conferences, as a department. This would be unusual for any school, but this is a school in a low-income area with few resources. Lessons are accompanied by the steady hum of cars zipping past on the two freeways that surround the school, and interrupted at frequent intervals by the sound of trains that pass just feet away from the school yard. Financial resources are low – in the school and in the students’ homes. Yet qualified mathematics teachers are queuing up to join this department and after one year of

---

<sup>1</sup> pseudonym

studying and monitoring the mathematics teaching and learning in this school, we have discovered some unequivocally positive facts.

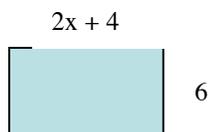
As part of an NSF-funded research project on mathematics teaching and learning, my research team and I gave incoming students at three high schools a mathematics examination. The students at Railside High scored at a significantly lower level than those in the other – more wealthy – schools in our study. However, after one year at Railside High, the students attained a higher average score on the end-of-year algebra examination that all students took, than the students at the other schools in the study. Their “improvement,” or learning over the year, was significantly higher than students at the other schools.

There is one other high school in the district—the other is in a more wealthy area. The two high schools give students the same final examinations at the end of each course (algebra, geometry, trig, etc) that carefully assess the mathematics in the Californian standards. These exams are constructed and graded by the two departments and overseen by the district. Last year Railside high school students significantly outperformed their more wealthy counterparts at every mathematics level.

In questionnaires, the Railside students are significantly more positive about mathematics than other students in our study. Indeed, there are many indicators that the mathematics teaching at Railside High is unusually effective and the vast amounts of time teachers spend working together and preparing lessons that will challenge and motivate students, pays significant benefits, in terms of student engagement and learning. The students at this school appear to learn more mathematics than most, develop more positive attitudes towards mathematics, and take more mathematics courses.

But all is not well at Railside High. The hard working teachers and students have been dealt a devastating blow, as the State has decided they are ‘under-performing’. Despite out-performing the other district high school, and the other schools in our study, on varied mathematics assessments, the Railside students scored significantly lower on the SAT-9 than the other high schools. In addition, SAT-9 scores at Railside did not ‘improve’ sufficiently over the one-year time period set by the state. But the ‘under performing’ label conferred upon this school bears no relation to the learning we see in classrooms. This mis-match - between the students' achievements and the state's label - is unfortunate, for many reasons, but it also provides us with an important opportunity to consider what is being assessed in the SAT-9 and what is not. Consider, for example, two of the questions from our assessment, directly assessing the mathematics in the California standards. In these questions the Railside students performed at a significantly higher level than students from other schools.

1. Here is a rectangle. The sides are  $2x + 4$  and 6 units.



- a. Find the perimeter of the rectangle. Simplify your answer if possible.
- b. Find the area of the rectangle. Simplify your answer if possible
- c. Draw and label a rectangle with the same area that you found in part b, but with a different length and width.

2. Solve the following equations:

a)  $5x - 3 = 101$

b)  $3x - 1 = 2x + 5$

These questions differ from those in the SAT-9 in a number of ways. First, they are not set in contexts that are confusing to linguistic minority and low-income students. Second, they reward all students who attain the correct answers, rather than only those who have answered the questions in the same form as the acceptable multiple-choice answer. Third, they do not use long and confusing sentences. By contrast, consider this question from the SAT-9 test for students of the same grade:

'A cable crew had 120 feet of cable left on a 1000-foot spool after wiring 4 identical new homes. If the spool was full before the homes were wired, which equation could be used to find the length of cable (x) used in each home?

F  $4x + 120 = 1000$

G  $4x - 120 = 1000$

H  $4x = 1000$

J  $4x - 1000 = 120$

The most obvious difference between our questions and this one is that the SAT-9 question is set in a context with which only some students will be familiar. In addition it uses long sentences and words unknown to many students new to the country (eg spool, cable crew, wired). Importantly, the expression that would sensibly be used to represent the length of cable used:  $x = (1000-120) \div 4$  does not appear as an acceptable answer. This question, as with many others in the SAT-9, assesses many things - confidence in the face of unfamiliar answers, context knowledge, and language – but none of these are indicators of mathematics knowledge. Importantly, they are all likely to stack the deck against language learners, students from low-income homes, students who are from minority ethnic and cultural groups, and girls (Boaler, 1994; Cooper & Dunne, 1998; Zevenbergen, 2000). In other questions students are asked to consider a students' bank balance, and the possible values of combinations of nickels and dimes. Students who have a bank account will undoubtedly be advantaged by questions that reference them. I arrived in this country from England three years ago and I am still thrown on the rare occasions when I come across the terms 'nickel' and 'dime', because they are rarely used in modern-day American society and I had no cause to learn them before I came to the country. The publishers of the SAT-9 questions have used contexts in response to recommendations from NCTM and other groups, that mathematics be taught through

realistic problems and situations, but teaching situations in which students are learning, are very different from standardized assessments in which they are being tested. Contexts may be useful and motivating in classroom activities and questions, but they are minimized in the standardized assessments used in most other countries because it is known that they present barriers to some groups of students and not others, and they contribute to inequalities.

In interviews the students at Railside reported that they found the SAT-9 totally confusing, mainly because of the language and contexts used in the mathematics questions. But our research has found that there was yet another, more insidious, factor impacting students success on the SAT-9 at Railside. The students had been told by the State that their school was ‘under-performing’, so they did not expect to do very well on the tests. Claude Steele has shown the importance of ‘stereotype threat’ (1997). He found that when students were told that the test they were about to take tended to produce achievement differences, with women and minority students scoring at a lower level, this is exactly what happened. In the control groups where students took the same test, but were not told about any expected performance differences, there were no performance differences among different groups of students (Steele, 1997). Educational research – a field that often produces different results – shows remarkable consistency on this issue. If you tell students they are low achievers, they achieve at a lower level than if you do not. In one of our first visits to the school a young boy asked us why we were looking at the mathematics department. When we replied that it was interesting, he frowned quizzically and said “but we are a 3”. He was referring to the API 3/10 ranking the school had been given. In recent interviews the students all told us that they go to a “ghetto” school. Students in other schools had told them this. The Railside students struggled to make sense of the label, as they believe that the teaching at Railside is good and the teachers really care about them, but they have been seriously affected by these different labels that have emanated from the SAT-9.

In addition to the previous years' school label that appeared to impact the students at Railside, when they took the SAT-9 they received their own, individual 'label'. Parents and students at this school, as in all others in California, were sent the results of the students' SAT-9 tests shown on a graph divided into 3 sections marked 'above average', 'average' and 'below average'. This label tells us nothing of what students have learned over any period of time. One of the students we interviewed, Simon, had arrived in this country from Nicaragua as a young boy, and told us that elementary school was a time of constant failure, as he couldn't understand what the teachers were saying. But he has since caught up and is excelling in school. He told us that the teachers at this school told him that he was smart and he started to believe in himself and achieve. He now loves mathematics, is very appreciative of the teaching at the school, and in our assessments he performed extremely well. Despite all of this, when the SAT-9 result arrived at his door he started to question his ability:

“My parents, they saw in the SAT 9 graph thing I was *below average* in the majority of the things, and especially math. I was like - *below average*. Right there. The thing is like - *below average* - you want it to be a little bit above average.”

I asked him whether that affected how he thought of his abilities as a mathematics learner, he told me that it did, because:

“You say - you tried so hard and then suddenly they give you a paper where it says you’re *below average* and you’re like - what? I did so much work.”

Simon had reasonably assumed that the result he was given should tell him something about how hard he had worked, or what he had learned in mathematics, but it did not.

The teachers at Railside are concerned that the SAT-9 results that are sent home to parents will convey negative messages about the students’ mathematics learning. In response, they have started to organize portfolio days, where students show their mathematics work and all that they achieve, to parents. These are very positive, well-attended events, but Simon’s repetition of the term ‘below average’ reminds us of the impact such labels have, and the extent to which they are internalized and remembered. Almost half of all students in the state are told that they are ‘below average’. In the new system that will be used in California, students will not take the SAT-9, but a different test that also comprises only multiple-choice questions and that will confer labels that seem more damaging. In the new system approximately half of all students will be told that their attainment is "basic", "below basic" or even "far below basic". What impact, I wonder, do those who designed this system think that this will have upon students’ confidence, and future mathematics achievement? Research tells us that confidence in one’s ability to succeed in mathematics, is an intrinsic part of success and motivation. The labels the students at Railside received are working against the positive achievements their teachers had brought about.

Now Railside students are being told by students in the other district school that they attend a “ghetto school” because their test scores are low. Yet the different researchers I know who have spent time in the school, agree that it is one of the most professional and dedicated departments they have ever seen. The hard-working mathematics teachers have, understandably, been demoralized by the school’s label, as one of them reported to us in an interview:

“They told us we had been considered an “under-performing” school because of our API scores, and if we wanted to, we would be able to choose this outside consultant to come in and help us raise our test scores. And I left that meeting in tears because I have never in my lifetime worked as hard as I work here to help students learn math and show off what they know about math in a way that is meaningful and makes sense to them. So being told that we were under-performing meant to me that I hadn’t been doing my job. And that was at the heart level.”

The teachers are now thinking that they need to spend more time on test-taking skills, even though they do not believe this will improve the students’ understanding of mathematics.

State officials are now reconsidering the assessment and reporting systems they use and it seems imperative that they consider both what is being assessed, and the impact the associated labels may have, very carefully. Such reflections must include careful examination of the reasons that schools with high proportions of low income and language minority students are at the lowest levels. The correlations between SES and SAT-9 success have been reported to be as high as 0.9 in some districts. The simple fact that the vast majority of ‘under-performing’ schools in California are in areas of high poverty should give us cause to look very carefully at the nature of the test that produces this label. The students at Railside tell us that language and contexts are a huge barrier, as well as the unfamiliarity of the test – as they do not normally face a barrage of short questions with “fill in the bubble” answers, and they are not normally required to work without calculators. Railside is not the only California high school that is producing positive learning environments and being told it is under-performing. We have collected a range of evidence that suggests that the low performance of students in the SAT-9 at Railside is related less to mathematical understanding than it is to language, context interpretation (which relies heavily on language) and test-taking skills. Using such tests – and their associated labeling – as a supposed tool to increase the performance of underachieving students, particularly those from low-income and ethnic minority homes – does not therefore seem to be a wise decision on the part of Californian policy makers.

Even if the SAT-9 was a good test of students’ mathematics achievement, which I strongly believe it is not, no single test can measure students’ learning, only attainment at a certain point in time. When the British government moved from reporting test scores to reporting students’ improvement over their years in school, the list of schools, ordered by “success” completely rearranged. When ‘improvement’ (learning) became the criterion, many of the excellent schools in low-income areas moved to the top of the lists. The designers of the API system have tried to assess learning by focusing upon the improvement of schools, but this is measured by comparing one cohorts' scores to a later cohorts' scores. Such a system not only compares different students, without capturing the learning of any particular students, it is open to considerable abuse. Schools that push needy students away and into other schools will report higher "improvements" than schools who keep their students with special needs and work to help them. Exclusionary practices, aimed at higher improvement ratings, are already being reported. Simon, from Nicaragua, may have worked harder, learned more, and improved more than anyone else in the country over the year, but a test score could never give any indication of that. The SAT-9 does not measure learning and events in this school suggest that its main effect is to lower students’ perceptions of what they can do, and demoralize teachers. The place of this and similar regimes of testing and labeling in an education system that purports to hold higher – and more equitable – achievement as its goal, seems questionable, at best. The students and teachers at Railside have discovered this, at some considerable personal cost.

## **References.**

Boaler, J. (1994). "When do girls prefer football to fashion? An analysis of female underachievement in relation to "realistic" mathematics contexts." *British Educational Research Journal* 20(5): 551-564.

Cooper, B. and M. Dunne (1998). "Anyone for tennis? Social class differences in children's responses to national curriculum mathematics testing." *The Sociological Review* Jan: 115-148.

Steele, C. (1997). "A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance." *American Psychologist* 52(6): 613-629.

Zevenbergen, R. (2000). "Cracking the Code" of Mathematics Classrooms: School Success As A function of Linguistic, Social and Cultural Background. In J. Boaler (Ed.), *Multiple Perspectives on Mathematics Teaching and Learning*. (pp. 201-224). Westport, CT: Ablex Publishing.

Dr Jo Boaler  
Associate Professor  
Stanford University , School of Education  
485 Lasuen Mall  
Stanford, CA 94305-3096  
650-723-4076