

Failure of U.S. Public Secondary Schools in Mathematics

Michael Marder, PhD
Professor of Physics
Co-Director of UTeach
University of Texas, Austin
Austin, TX

Abstract

Metaphors play a powerful role in arguments about education. It is common to say schools are broken, and that the school system is failing. Here I take the metaphor seriously and briefly review an historical episode where airplanes failed seemingly for no reason at the dawn of the jet age. The responses to these failures at first with public relations and eventually with science have lessons to teach about our complex system of schools. I review evidence, mainly in graphical form, that poverty is the primary threat to schools, as, a generation ago, cracks were the primary threat to airplanes.

Key Words

school reform, poverty, school failure

The loss of Yoke Peter and Yoke Yoke presented a problem of unprecedented difficulty, the solution of which was clearly of the greatest importance to the future, not only of the Comet, but also of Civil Air Transport in this country and, indeed, throughout the world. Lord Cohen, Civil Aircraft Accident, Report of the Court of Inquiry, 1955

The Problem

They say the U.S. public school system is broken, its failure an ongoing disaster for our students and country. Failure is not a metaphor to be taken lightly; it is a subject of intense scientific study. Cases in which deadly failures were studied and overcome have lessons for school administrators and education policy makers.

One of the most dramatic lessons comes from the dawn of jet aviation in the United Kingdom. Passenger planes started falling out of clear skies. Early theories to account for avian failure were wrong. The problem was not quality of the pilots, as initially believed, but the quality planes themselves. The first jets were not flaw-tolerant systems and large stresses formed in rare locations on airframes. The British aviation authorities were too slow to understand the real causes for the crashes and they lost civil aviation for decades.

In the U.S. today, theories for why schools fail are wrong. The problem is not quality of the teachers. Schools fail under stresses of concentrated poverty. We stand to lose something larger than aviation -- technical leadership -- unless we understand and remedy the problems.

The Comet

In the middle of the twentieth century, Britain appeared poised to dominate the jet age. In 1952, the de Havillands Comet began commercial service, triumphantly connecting London with the farthest reaches of the Empire. The jet plane was years ahead of any competitor, gorgeous to look at, and set new

standards for comfort and quiet in the air. Then things went horribly wrong.

No one can say the problems could not have been imagined. In 1948, Nevil Shute published *No Highway*, a best-selling novel about a fictional new aircraft with a defect that made it crash after 1440 hours of service. Jimmy Stewart and Marlene Dietrich starred in a film version released in 1951. The novel is filled with realistic details about aircraft design and safety because Shute was trained as an aeronautical engineer. In *No Highway*, a plane falls out of the sky and the crash is attributed to bad weather and pilot error. In 1953 a Comet fell out of the sky, and the crash was attributed to bad weather and pilot error. In the novel an engineer determines that the cause of the crash is fracture due to metal fatigue, and succeeds after great difficulties involving a child with a concussion, backstabbing politicians, and affections of a movie star and a stewardess to have the whole fleet of jets grounded and repaired. In 1954 two more Comets fell out of clear skies; their failure attributed to metal fatigue and fracture. In the real life case, the fleet was not grounded and saved. A whole plane was placed in a testing tank and stressed to failure to find the cause of disaster, resulting in a brilliant accident report. But it was too late.

The Paris Law

As the Comet accident report was being released in 1955, a little-known military contractor in the northwest corner of the United States was completing its prototype for a civilian jet airplane. The company's engineers knew that cracks had brought down the Comet, and they had better understand them before they brought down the Boeing 707.

Boeing brought in a researcher, Paul Paris, a mechanical engineering graduate student. Paris knew nothing about fracture when he arrived, and eager to prevent his employers from discovering it, he read about 120 papers on the subject in a few weeks. None of them made sense to him or were of any use, except for papers by a researcher at the Naval Research Laboratories, George Irwin. They were not widely accepted. An expert commented at the time that Irwin's theory "could not be mechanics because it had dimensions never seen in mechanics."¹¹

The view of fracture that Paris brought to Boeing was dramatically different from the one that had guided construction of the Comet. Cracks were the centerpiece of the investigation. They could not be eliminated. They were everywhere, permeating the structure, too small to be seen. The structure could not be made perfect and it was inherently flawed. The goal of engineering design was not to certify the airframe free of cracks but to make it tolerate them. Part of the tolerance was achieved by eliminating points of stress concentration. The corners of the Comets' rectangular windows had been such points. Angular windows had been chosen for aesthetic appeal but it was a deadly mistake. At the corner of a perfectly rectangular hole in an airplane skin, stresses rise far above their values in other places, and a crack triggered at a window corner in fact felled the frame. This much had been determined during the Comet inquiry, but Paris went further. Every particular metal alloy chosen for use in the Boeing planes was tested according to a new concept of "toughness." Cracks were deliberately introduced into a metal sample of special shape and size, and the metal was stretched and shaken in specific fashion until it broke. Paris found that some types of metal sheet were much tougher than others, and that for some the

toughness improved greatly as thickness increased while for others it did not. The goal was to make the airframe fail-safe, meaning if a large crack was introduced at any point and the maximum possible flight stresses applied, the crack would scarcely advance.

By 1957 the engineering work was done, and Boeing was ready. They produced a brief documentary film, *Operation Guillotine*. A steel blade the size of a shovel falls on a pressurized aircraft hull in a laboratory, the top bursts off, and it explodes. The message was not subtle. That was the Comet. Next the blade falls on a new Boeing hull which absorbs it, hull intact. It was the end of the Comet, the end of de Havillands, the end for a long time of British civil aviation.

The Educational System

Now we come to a system larger and more complex than an airplane, and more important to the United States than aviation: the public education system for our children. There is a theory of how and why the school system is failing that guides federal and state governments. We are in the position of the British in the wake of watching Comets fail. We carried out tests on an enormous scale to diagnose the situation. But the theory of failure is wrong, the repair regime misguided, and unless correct theories are rapidly developed and employed, schools will crash in growing numbers.

The dominant school repair program has three parts. The first is deregulation of public education through creation of charter schools, and promoting choice. The second is deregulation of teaching through creation of alternative pathways to teacher certification and weakening influence and control of unions. The third is accountability, measuring the performance of all students, and holding

teachers and administrators accountable for the educational results. These reforms reinforce each other.

The reforms are based upon concepts of why schools fail. Reformers say “[r]esearch tells us that teacher quality is the single-most important factor in determining student achievement ... [T]he impact of a teacher (for good or for bad) is cumulativeⁱⁱ.” “Looking at the range of quality for teachers within a single large urban district, teachers near the top of the quality distribution can get an entire year’s worth of additional learning out of their students compared to those near the bottom ...ⁱⁱⁱ.”

Student achievement means scores on high-stakes tests, such as TAKS in Texas,

FCAT in Florida, or Regent’s exams in New York. *Teacher quality* refers to whatever quality is possessed by teachers whose students obtain large gains in test scores. The argument is not circular, however. The assertion is that some teachers consistently get larger gains than other teachers when working with the same students.

Surprisingly, reformers say, good teachers cannot be identified by knowing which courses they took or workshops they attended or degrees they obtained; only when they actually teach can the strong be separated from the weak. In this view, the main problem with schools is that they have many inferior teachers, and if enough of the worst teachers were eliminated and replaced by much better ones, educational difficulties would be cured.

Math Performance in Third Year by Ranking after First Two Years

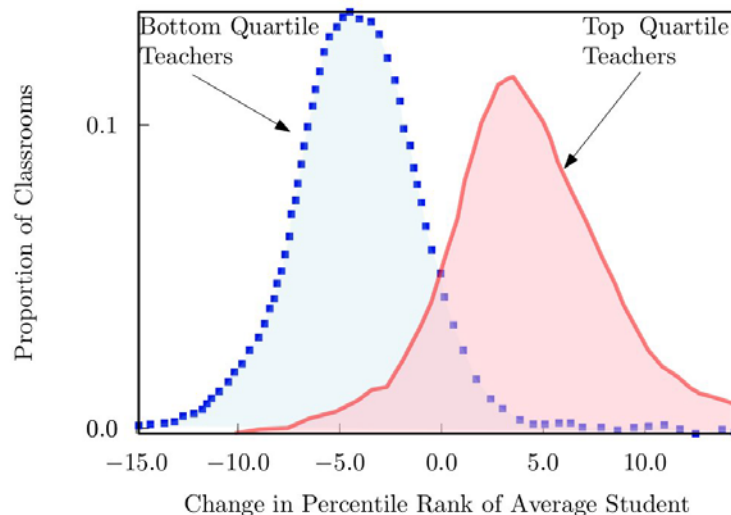


Figure 1: Classroom-level impact on average student performance, controlling for baseline scores, student demographics, and program participation. LAUSD elementary teachers with less than 4 years teaching experience. Gordon, Kane, and Staiger, 2006.

The Teacher Quality Evidence

Is teacher quality the main cause of failing schools? No more than pilots were the main

cause of Comet crashes. Yes, the quality of pilots and teachers matters, but it is very hard to measure when failing vessels travel through turbulent weather.

The most persuasive and influential evidence for the persistence of teacher quality appears in Figure 1.^{iv} This figure shows that if elementary school teachers have students with small mathematics score gains in two consecutive years, the odds are great their students will have small score gains in a third year. Conversely, for elementary school teachers whose students had large gains two years in a row, their odds of large gains the third year are high.

That is, there are good and bad teachers, and they can be identified by monitoring their

students' performance in as short a time as two years.

These findings do not settle the case for teacher quality. Some teachers might regularly do better than others because their classes are typically different. For example, one teacher might get a large number of struggling students every year. One teacher might mainly get students whose aggressive and supportive parents sought her out for their children. One might get students with scores so high to begin with it is hard for them to make significant gains.

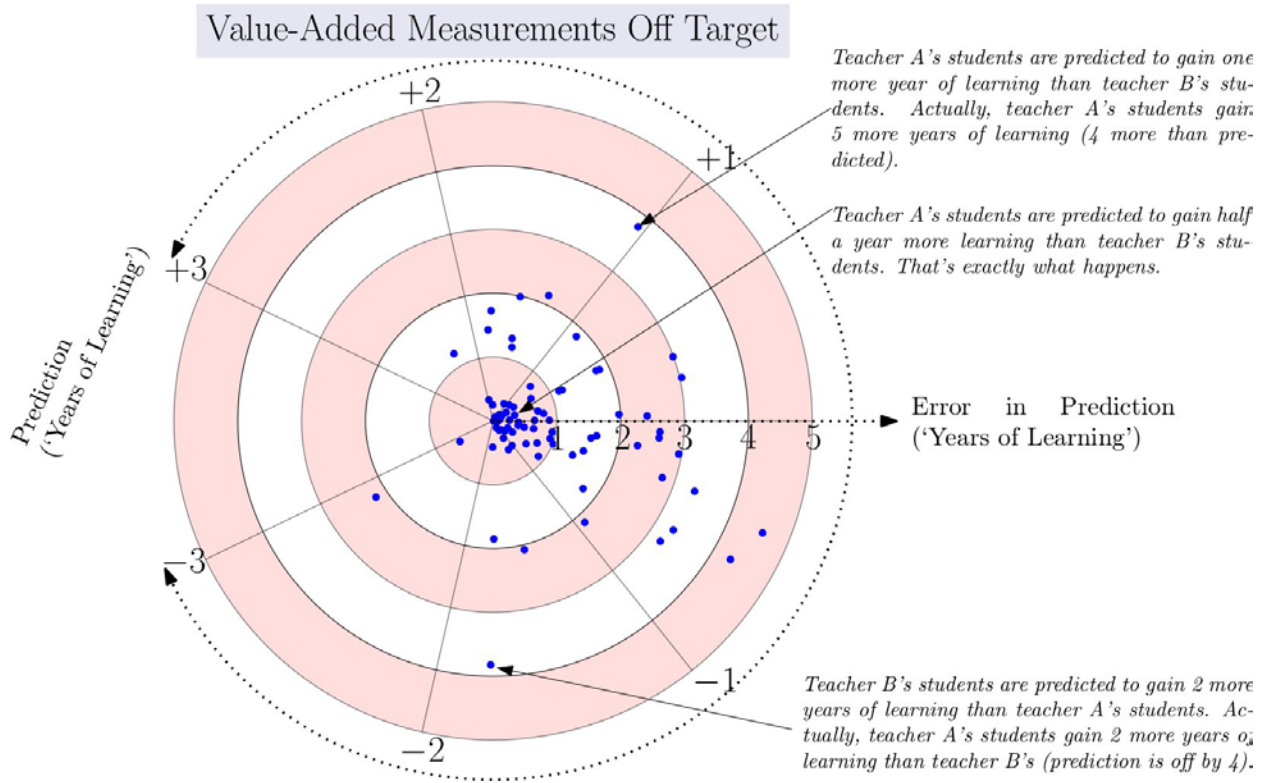


Figure 2: Value-added Measurements Off Target

To test whether teacher quality can be measured and that it does not depend upon the particular set of students teachers are given,

assign them to random classes of students. It is hard to arrange, but Kane, and Staiger did it.^v They assembled data on 78 pairs of elementary

school teachers over four years, and then randomly swapped their classrooms at the beginning of the fifth year. The researchers predicted performance targets the teachers would reach. But as shown in Figure 2 the predictions were often off the mark. When value-added computations predict student performance, points hit the bullseye.

The more predictions are wrong, the further away points lie from the bullseye. Distance units are .25 standard deviations, which the authors suggest correspond to around a year of learning. The actual value of the predicted student learning difference is indicated by the angle around the circle

Following the random switching, supposedly lower-quality teachers obtained better student performance than supposedly higher-quality teachers around one third of the

time. On average, the teachers said to be better did get better student performance, but fluctuations in predictions were large.

A reliable measurement of teacher quality would need to hit the target most of the time. Rewarding or punishing individual teachers when the best measurements of their quality look like Figure 2 might improve schools on average, but it would be unjust.

Unfortunately, there are no more careful tests of the predictability of teacher quality than appear in Figures 1 and 2. Furthermore, these experiments concern elementary school, where the U.S. does not fare so badly in international comparisons. For secondary schools, where U.S. performance sinks by international comparisons, it is not clear how to measure teacher quality.

The Poverty Argument

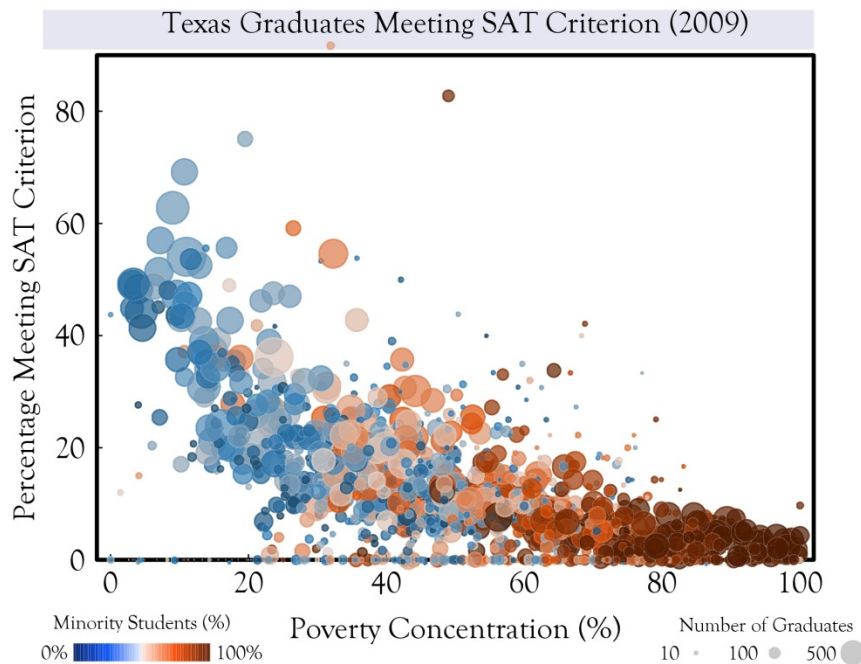


Figure 3: Texas Graduates Meeting SAT Criterion

When Comets failed, the Royal Aircraft Establishment put an entire airplane in a testing tank. Persuaded its schools were failing, the United States put the whole public school system in a testing tank. Since 2002, public school students in the country were tested in mathematics and reading or language arts in Grades 3-8, and once in Grades 9-12. In states, such as Texas, (where these accountability laws originated) mathematics and reading have been tested every year in Grades 3-11. Data for student performance on these exams are publicly available for almost every school in the country on state education agency websites.

The collection of nationwide data do point to a primary cause of school failure, but it is poverty, not teacher quality. As the

concentration of low-income children increases in a school, the challenges to teachers and administrators increase so that ultimately the educational quality of the school suffers. Challenges include students moving from one school to another within the school year, frequency of illness, lack of stable supportive homes with quiet places to study, concentration of students who are angry or disobedient, probability of students disappearing from school altogether, and difficulty of attracting and retaining strong teachers. Most people who see the connection between poverty and educational outcomes are confident that low-income students in a sufficiently supportive environment will learn as much in a school year as students in well-off communities.

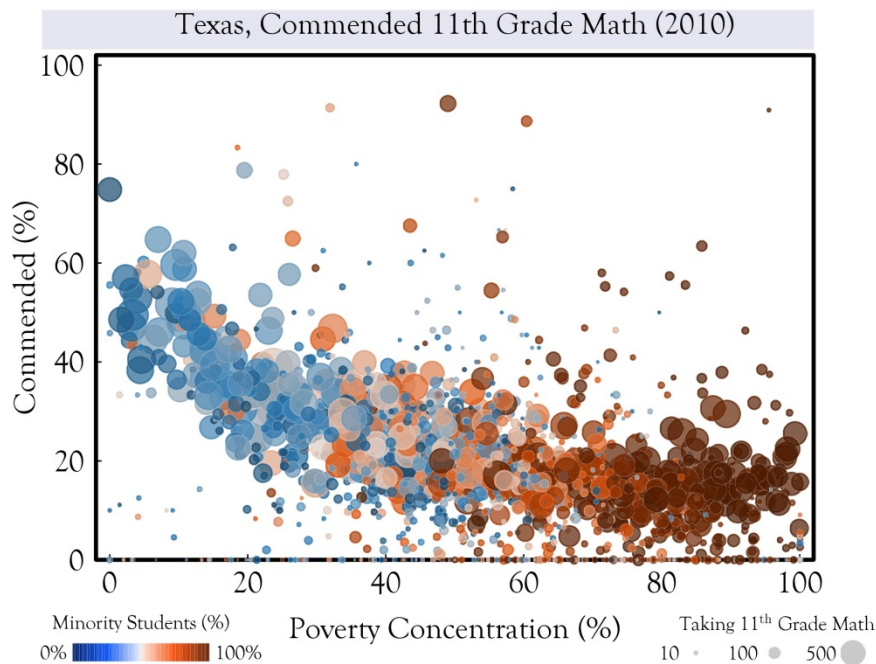


Figure 4: Texas Commended on Grade 11 Math versus Poverty Concentration

The concentration of poverty in a school can be defined precisely, although not perfectly, by the concentration of students eligible for free and reduced-priced meals. Eligibility is determined by family income, and uniform data are available for every public

school in the U.S. Many different measures of student performance are available and tell a consistent story. I will begin with a college-readiness indicator provided by Texas, which counts the fraction of students in each high school who take the SAT and score 1110 or

more or ACT and score 24 or more. In Figure 3 I plot the percentage of high school graduates meeting Texas' SAT/ACT College Readiness Criterion as a function of concentration of poverty. Every disk is a high school, with the area of the disk proportional to the number of graduates. Colors indicate the percentage of minority students in school. Figure 4 depicts the percentage of Grade 11 students in Texas who receive Commended scores in mathematics as a function of concentration of poverty. The area of each disk is proportional to the number of 11th graders.^{vi}

The association between poverty concentration and educational performance is very strong. For example, among schools where less than 15% of the students are eligible for free and reduced meals, there are virtually none where **fewer** than 20% of the students graduate college-ready. Conversely, among schools where more than 85% of the students are eligible for free and reduced meals, there are none where **more** than 20% of the students graduate college-ready. In short, the least successful schools serving the wealthy do better than the most successful schools serving the poor.

The same data are available for 2006, 2007, and 2008 with very similar findings. If teacher quality were indeed the most important

factor that impacts student achievement, logic would dictate that every Texas high school with more than 85% poverty concentration has retained a staff of largely inferior teachers for as long as data have been collected, whereas virtually every single Texas high school with less than 15% poverty concentration has managed to acquire superior teachers.

The SAT college readiness criterion in Texas brings out an especially high contrast between schools of well-off and low-income students because taking the SAT and ACT are not mandatory. The tests reflect hopes to attend out-of-home-state colleges. But all Texas students take a mathematics exam in Grade 11 and although obtaining a commended score (around 90%) is not as demanding as obtaining 1110 on the SAT, it sets a reasonably high bar and the results display a similar pattern.

Only a few schools where the poverty rates exceed 85% match performance of schools where poverty concentration is below 15%. Poverty concentration also strongly associates with fractions of students in schools simply passing mathematics exams (scores of better than around 60%, Figure 5). The graphs here are only a small sample from an abundance of evidence. In every state or year I examined to date, poverty is tightly connected to high school performance in mathematics.

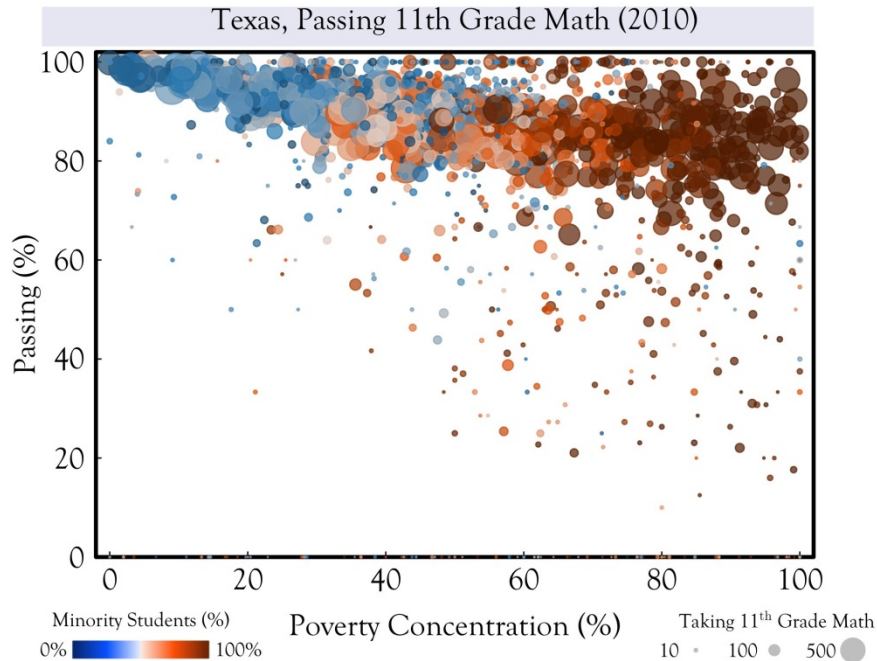


Figure 5: Texas Passing Grade 11 Math versus Poverty Concentration

The Evidence on Charter Schools

Adherents of the theory that “quality teachers are the key ingredient to a successful school and to improved student performance” place special emphasis on charter schools (“... schools that enjoy public financial support but that operate outside the controls that hamper traditional public school systems.”)^{vii}. Charter schools are supposed to provide laboratories in which to test new ideas. In Figure 6, I show the percentage of high school graduates who meet the Texas SAT/ACT College Readiness Criterion plotted as a function of concentration of poverty. Every disk is a high school, with the area of the disk proportional to the number of graduates. Charter schools are highlighted; non-charters are grey. Figure 7 shows the same, but for commended scores in eleventh grade mathematics.

Fourteen or 15 of the charter schools stand out positively, but the rest are comparable to regular public schools or worse.

Figure 8 displays passing scores at 11th grade for Texas charters. Secondary charter schools in Texas are worse than those in some other states, but in no states do secondary charter schools look strikingly better than regular public schools. Plots of educational performance highlighting charter schools in various states appear in Figures 9-12. All data come from data sets available for public download. I chose the highest-level measure of mathematical performance I can find in each case. In every state educational outcomes depend strongly on the concentration of poverty in schools, and charter schools are either not distinguishable from other public schools (Florida, California), are too few to draw conclusions (New York) or are markedly worse (Texas, New Jersey).

The same achievement patterns are strong in New Jersey, a state in which the Commissioner of Education, Christopher Cerf has openly doubted that poverty provides a good way to identify students at risk.^{viii}

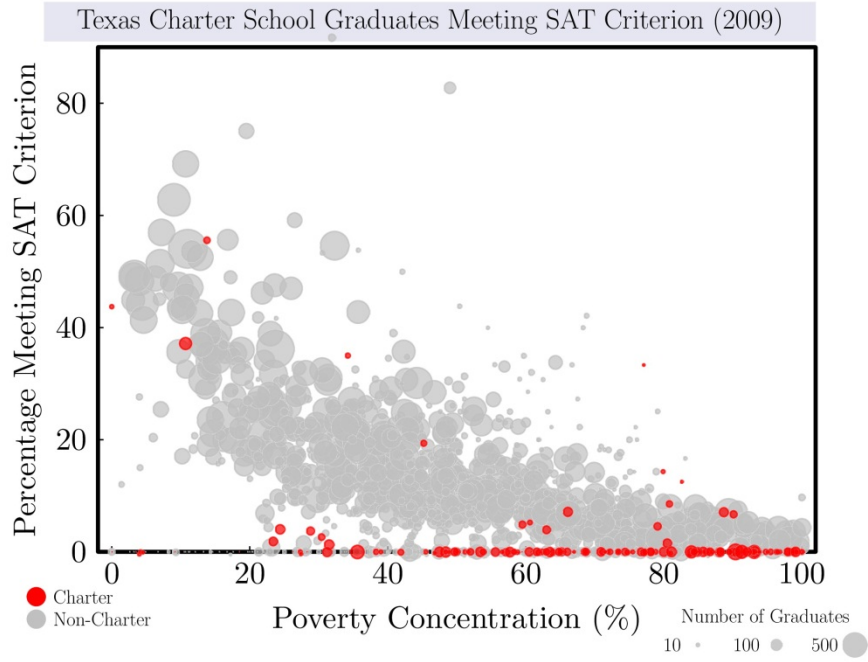


Figure 6: College Readiness of Texas Charter School Graduates

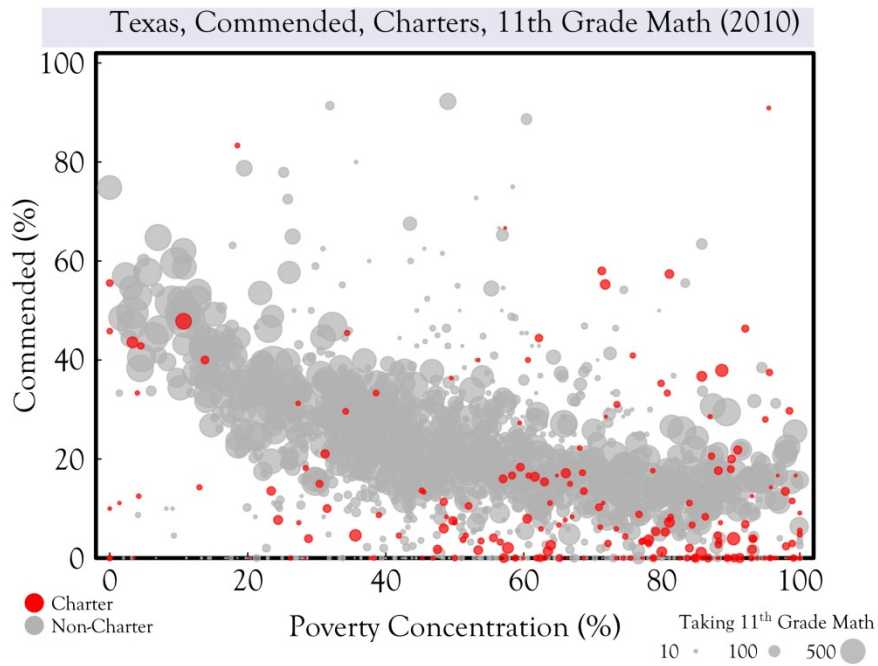


Figure 7: Concentration of Poverty and Commended Scores for Texas charter schools

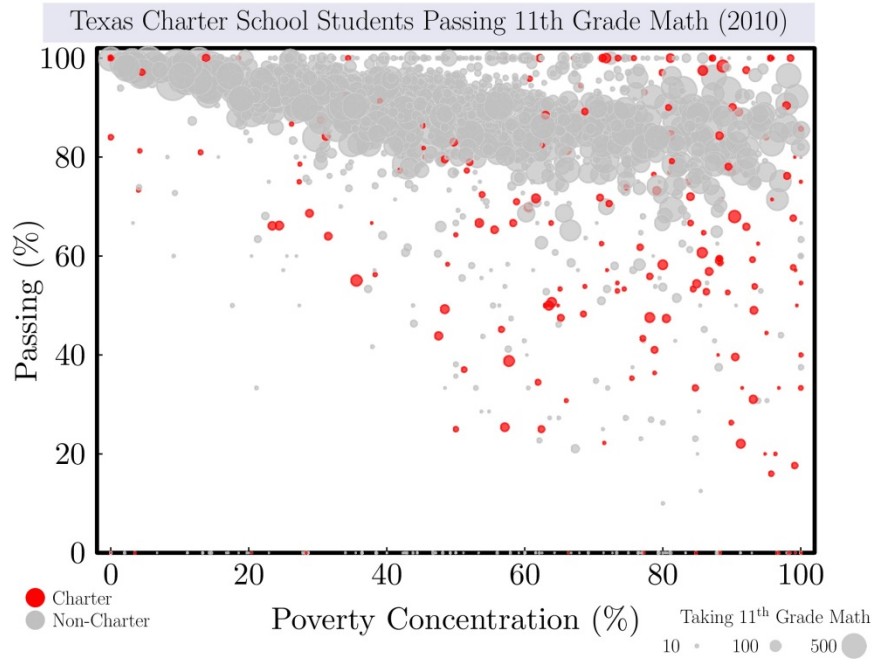


Figure 8: Concentration of Poverty and Passing Scores for Texas charter schools.

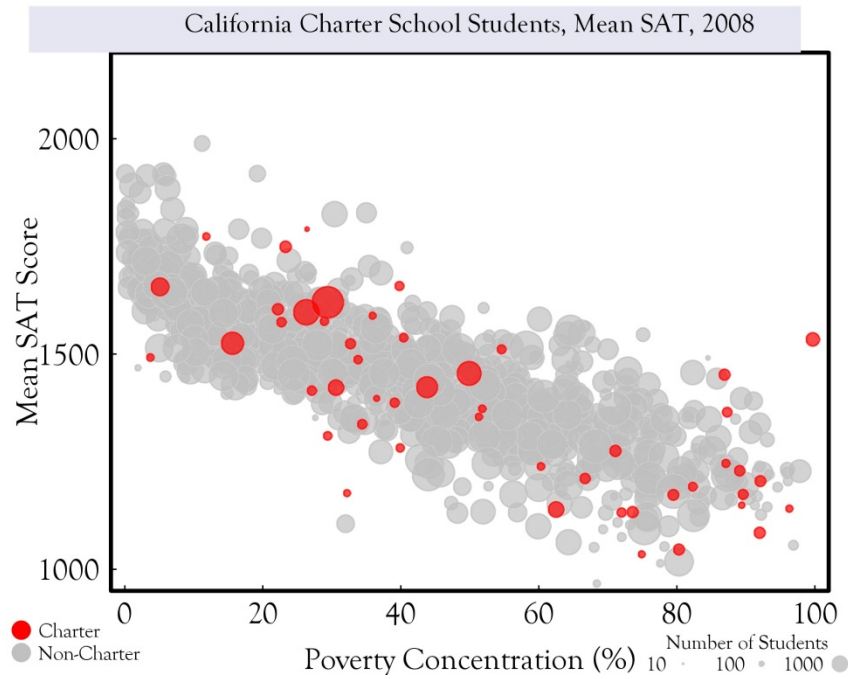


Figure 9 California High School Student SAT Scores and Poverty, highlighting charter schools

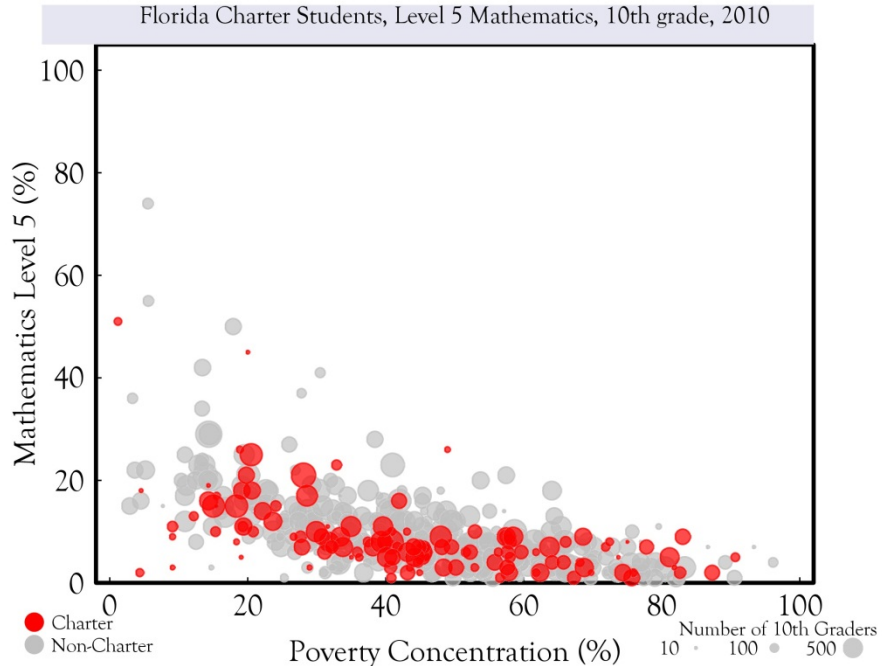


Figure 10 Florida Grade 10 Student Mathematics Assessment Scores and Poverty, highlighting charter schools

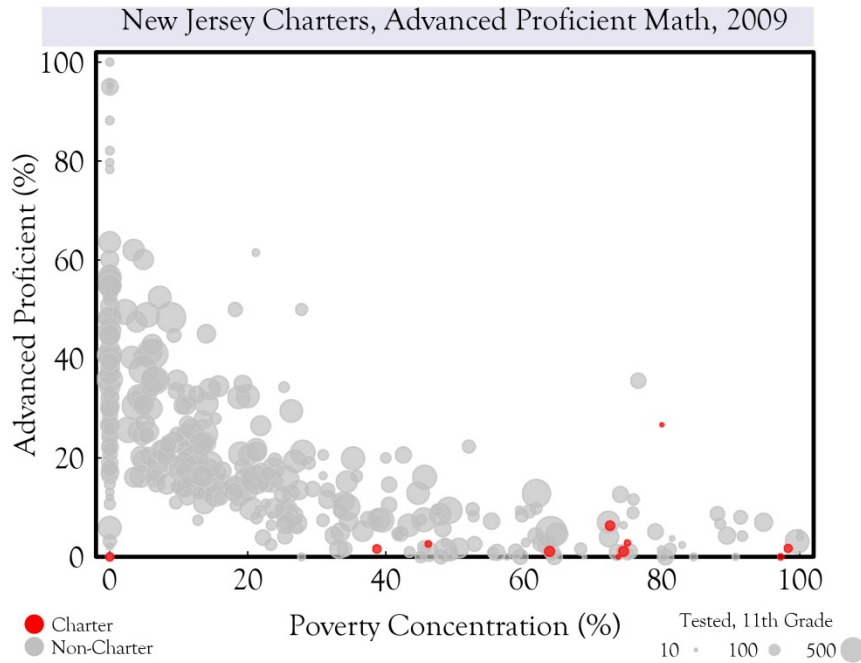


Figure 11 New Jersey Student Achievement in Mathematics Grade 11 and Poverty, highlighting charter schools

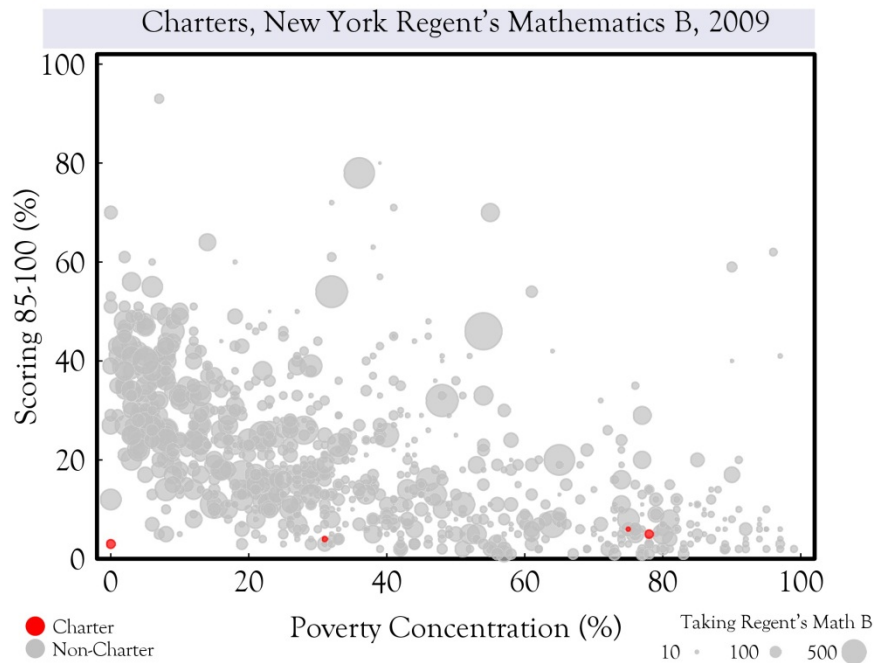


Figure 12 New York Regent's Math B Scores and Poverty, highlighting charter schools

The Rising Bar for Public Schools

The comprehensive measurements of public school performance are accompanied with provisions intended to force their improvement. The provisions seem inspired by Milo of Croton who is said to have lifted and carried a calf every day as it grew to a bull so that Milo acquired superhuman strength. Applying progressive challenge to educational systems to improve them has a long history.

Abraham Flexner in a survey of medical education at the start of the twentieth century noted approvingly that *The state of Texas has taken a sound and yet conservative position. Beginning with 1909, it has decreed a gradual annual rise of standard that will shortly result in making its four-year high school the legal basis of medical education. Cautious elevation thus avoids all danger of breaking with the school state system.*^{ix}

This same principle decades later became part of the Texas public school accountability system and then was exported to the rest of the United States with the signing of No Child Left Behind in 2002. The law requires larger and larger fractions of students to achieve acceptable scores until by 2014, 100% of students are proficient in language arts and mathematics.

Each of these subjects must be tested at least once in elementary, middle, and high school. Since 5% of students can be exempted from testing, that means 95% of students are supposed to pass the exams. In Texas, 60% of students had to pass mathematics in 2010, and to reach the goal of 100% proficiency by 2014 as the law requires, the passing standard will need to rise on average 10% each year after 2010.

When a required fraction of students fail to pass the exams in any given year, the school is labeled Academically Unacceptable; if it remains unacceptable for 5 years strong remedies are required, usually dismissal of administrators and teachers. As the standards rise towards the demand of 95% proficiency, what is likely to happen?

Figure 13 provides a rough estimate of which Texas secondary schools may be declared Academically Unacceptable because of mathematics year by year as standards rise; unacceptable schools are depicted in red. It underestimates the difficulty of rising above Unacceptable. Students not only have to score well in mathematics, but in language arts, and to meet graduation targets. Subgroups of students such as low-income students or African American students have to achieve good scores. The estimates of school performance were produced by measuring the rate at which passing fractions of students increased for every school from 2008 through

2010, and by assuming that the rate of increase will continue (although leveling off at 100%) until 2014, when the mandatory achievement bar hits 95%. A difficulty in making these predications for Texas is that the current state exam has just been replaced by another. But the new exam is supposed to be more difficult.

The hero of *No Highway* finds himself in the sky—in an airplane that he himself has predicted will fail within a few hours. The failure of the public school system is scheduled to begin in two years. There is still time to set the plane down. It is almost taken for granted that the Federal law mandating schools reach 95% levels of proficiency will change before too many schools are affected. Congress will have to agree and vote to change the law.^x

Should necessary consensus not be reached in time, an unprecedented wave of dismissals will begin to sweep through the public school system, largely affecting teachers in schools of low-income children.

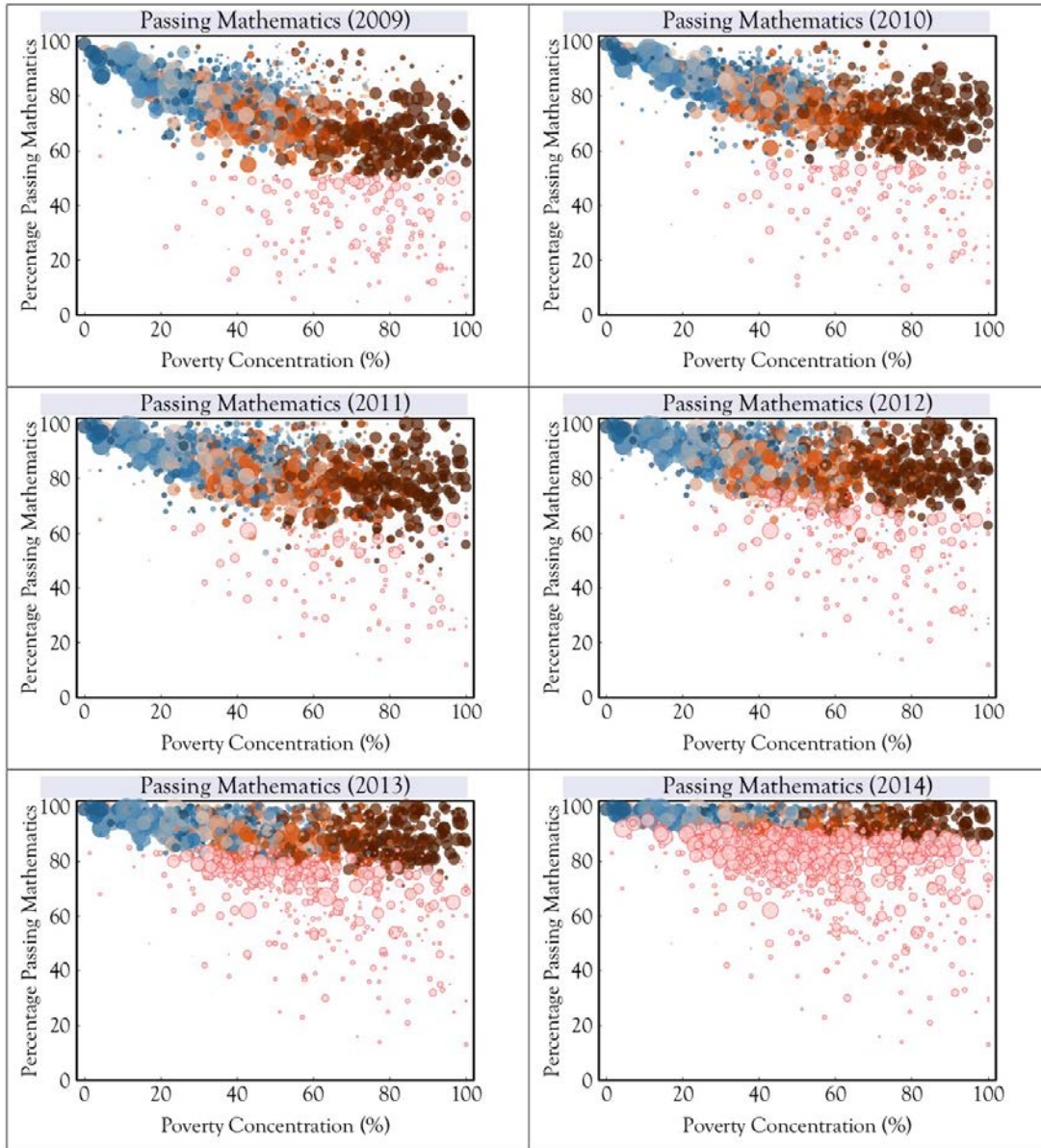


Figure 13: The rising tide of unacceptable schools to come in Texas.

Inspecting the data on charter schools, it is natural to conclude that because of their low scores they will be among the first to close. Interestingly, in Texas, this is not true. Texas has two accountability systems: the standard system by which most public schools are judged, and an Alternative Education Accountability system.

As shown in Figures 3-8, the mathematics scores produced at charter schools fall below the level at which regular public schools would be rated unacceptable. However most Texas secondary charter schools are judged by the second system with lower standards. Thus it is possible that the accountability system could force large

numbers of conventional public schools to close and put the children into charter schools where the levels of performance are lower.

The Future

The purpose of an accident report is to determine the root cause of an accident. Looking back at Lord Cohen's investigation of the Comet, this task was accomplished brilliantly. The least successful portions of the report are those making specific recommendations for the future. Cohen hesitantly suggested that planes could be made safe by stressing them prior to use in such a way as to permanently deform the skin in areas of stress concentration.

The correct engineering solutions turned out to be quite different. Poverty, and not teacher quality, is the main element connected to low student performance. The reform program brought to bear so far in charter schools shows no signs of providing superior education in secondary mathematics except in rare cases. Although appropriate solutions are not as clear as the problem, I close with a few remarks.

School reform cannot succeed if policies do not recognize that poverty is a significant factor. Poverty permeates schools as cracks permeate planes. The schools must be made fail-safe enough to tolerate it. Finding effective and affordable combinations of factors, such as parent education, health care, housing policies, advising, discipline, improved instructional models, and better teachers calls urgently for experimentation. Saying teacher quality is not by itself enough to reform schools does not mean that teacher quality does not matter.

Sudden structural collapse of the Comets could never have been remedied by a program of holding pilots accountable.

However the very first Comet that crashed did so due to bad weather and pilot error. Once planes are structurally sound, the pilots matter very much. Similarly, once problems associated with poverty in schools are addressed, outcomes associated with teachers should become easier to measure and relatively more important to affect.

Correctly defining the problem was the most important task accomplished by fracture mechanics developers such as Irwin and Paris. A major error in certifying the Comet came from errors defining material strength. Defining teacher quality exclusively through rises in student test scores is similarly flawed as a concept. Unfortunately there is no general agreement on an objective replacement. In some cases credentials are a legitimate measure of teacher quality. For example, physics teachers should have studied physics, preferably as a major or minor. No one advocates assigning randomly chosen teachers to physics or biology on the grounds that whether a person has studied science has no bearing on the ability to teach it.

In physics, chemistry, computer science, and engineering there is such a shortage of secondary teachers that schools avoid offering classes. A program of reform based on testing and accountability offers few answers to questions about how new people will be attracted to teach in shortage areas.

For the short term, preparing teachers in mathematics and science is a wise and useful step toward improving schools. As quickly as possible, we must understand the link between poverty and educational outcomes in the U.S., devise solutions, and finally test and implement them. Britain briefly tried to substitute public relations for aircraft safety and paid with the loss of its commercial aviation sector. I hope the U.S. can avoid a similar error. I hope that

proponents of teacher quality and charter schools will recognize the weakness of a single-minded approach before it is too late,

and that we will not damage public education, let down our most vulnerable students, and lose technical leadership we take for granted.

Author Biography

Michael Marder is associate dean for science and mathematics education and professor of physics at the University of Texas at Austin. His most recent books are *Condensed Matter Physics, second edition* (Wiley 2010) and *Research Methods for Science* (Cambridge University Press 2011). He is co-director and co-founder of UTeach, a program that prepares science and mathematics teachers at universities across the United States. E-mail: marder@chaos.ph.utexas.edu