Failure of U.S. Public Secondary Schools in Mathematics:
Poverty is a More Important Cause than Teacher Quality

Michael Marder¹

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

Many say the US public school system is broken, its failure is an ongoing disaster for our students and country. Once when systems of great consequence failed, the accidents were investigated with all needed resources, causes identified, mechanisms understood, solutions devised tested and implemented. I have been studying aircraft accidents that took place at the birth of jet aviation. Mistaken concepts employed by British engineers, and an illusion that confident public pronouncements could substitute for sound science helped move world leadership from the United Kingdom to the United States. The path the British followed to disaster bears an unsettling resemblance to the path the US now follows to reform our schools.

The Comet

Britain was set 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 makes 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, for Shute was trained as an aeronautical engineer. He worked for de Havillands and Vickers Aircraft in the 1920s, and in 1930 founded his own company, Airspeed Ltd which he ran until 1938. In the novel 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

¹ Michael Marder is Associate Dean for Science and Mathematics Education and Professor 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.
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 a second Comet fell out of clear skies near Rome. The fleet was grounded for two months while repairs were made. Flights then resumed with the declaration, "Although no definite reason for the accident has been established, modifications are being embodied to cover every possibility that imagination has suggested as a likely cause of the disaster. When these modifications are completed and have been satisfactorily flight tested, the Board sees no reason why passenger services should not be resumed." Four days after these words were written a third Comet fell into the sea out of clear skies near Naples and the fleet was grounded again indefinitely. Now there were no limits on the resources that would be devoted to resolving the problem. An exhaustive investigation released in 1955 determined that the last two Comets failed because of fracture due to metal fatigue. The conclusion was unequivocal. It resulted from a test in which a whole airplane was submerged in a water tank and subject to pressure cycles until it came apart (Figure 1), and from meticulous examination of airframe segments recovered from underwater such as in Figure 2.

The sequence of events once must have been shocking; now it is puzzling. One of Britain’s leading engineers, with specialty in stress analysis of aircraft, explains to millions that aircraft can crash from fatigue fracture. Almost immediately the first jet aircraft, upon which the whole future of British aviation depends, fails repeatedly in a manner identical to that described in novel and film. Engineers completely anticipated the problem. How could they not prevent it? No source better explains the thinking of British engineers about aircraft safety than No Highway itself. Shute explains that “An aluminium alloy which has stood up quite well to many thousands of hours in flight may suddenly become crystalline and break under quite small forces, with most unpleasant consequences to the aeroplane. That is the general story of the effect that we call fatigue in aircraft structures, and we don’t know a great deal about it.” His explanation of the science behind fatigue fracture is fundamentally wrong. The metals in the Comet were weak without having to become crystalline, partly because of particular shapes into which the metals were formed. British engineers failed to understand failure, and it made all the difference in the world.

2 Lord Cohen, Civil Aircraft Accident (1955) p. 16
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. Boeing had had little success with civilian craft in the past. The company 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 for the summer, Paul Paris, a mechanical engineer who had just finished a Master’s degree and was pursuing graduate studies at Lehigh University. 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.”

The view of fracture 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, it was inherently flawed, and 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 in Figure 2. 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” that had been introduced by Irwin. Cracks were deliberately introduced into a metal

---

3 Frank McClintock, Irwin Symposium volume, p. 9, Chan editor (TMS, 1997)
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, only 10 minutes long. In the opening minutes of 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 is not subtle. That is the Comet. The calm narrator discusses three sets of design modifications that prevent cracks from running, and shows the results of fatigue fracture tests, how cracks slowly elongate over time in a stressed and vibrating plate. Now the blades come down again in slow motion towards the new Boeing hull. They pass in and through like knives plunged into butter and the hull remains intact. In the final scene a shiny 707 soars into the air as the narrator flatly explains that passengers are now guaranteed safety. It was the end of the Comet, the end of de Havillands, the end for a long time of British civil aviation.

Not that the new science of fracture mechanics was immediately and universally accepted. When Paul Paris submitted his new theory on the motion of cracks in fatigue for publication in 1959 it was rejected by three top journals in turn. Today it is known as the Paris Law. The crash of an F-111 in 1969 finally led the US Air Force to mandate the new methods for all its planes, and soon afterwards the US Federal Aviation Administration mandated Damage Tolerance Analysis based on fracture mechanics to validate the safety of all commercial aircraft. In the space of 15 years concepts that had lain outside the bounds of possible imagination came to be used by government regulators to judge the safety of every airplane.

The Educational System

Now we come to a system larger and more complex than an airplane, more important to the United States than aviation: the education of 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 have 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 use of vouchers. The second is deregulation of teaching though 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. For example, it is much easier to hold teachers accountable when unions cannot automatically protect them on the grounds of seniority, or when they work at charter schools with innovative reform-minded administrations.

The reforms are based upon concepts of why schools are fail. The concepts come from appealing research claims that do not stand up to inspection. Reformers say “[r]search 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 cumulative."  “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.... [B]eing taught for four years by a teacher in the top...
quarter of ability versus a teacher in the bottom quarter can help eliminate the achievement gap in test scores for African-American students. A bit of translation is useful here. “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 would get 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. The reform analysis proceeds as follows: Measuring student test scores is critical for determining teacher quality, since teacher quality is defined by changes in student test scores. Frequent tests of all students and data systems to compile the results are essential. Unions provide professional protection to teachers, strong and weak alike, and need to be resisted and contained. Charter schools are nimble and independent, with the freedom and willingness to hire teachers who raise test scores, improve or dismiss those who don’t, and therefore are important to validate the model of improved instruction and provide parents choice. Certification of teachers should be made as simple and quick, with minimal barriers to entering the teaching profession, since good teachers can only be identified after they begin to work.

---

Fig. 3: 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.

Teacher A’s students are predicted to gain one more year of learning than teacher B’s students. Actually, teacher A’s students gain 5 more years of learning (4 more than predicted).

Teacher A’s students are predicted to gain half a year more learning than teacher B’s students. That’s exactly what happens.

Teacher B’s students are predicted to gain 2 more years of learning than teacher A’s students. Actually, teacher A’s students gain 2 more years of learning than teacher B’s (prediction is off by 4).

Fig. 4: Data of Kane and Staiger, comparing predictions of student learning based on teacher quality to measured student performance. Teacher quality was defined by measuring student test score gains for four years in their classrooms. Teachers were grouped in pairs, and teacher quality was used to predict the relative learning gains of students in their classrooms. The teachers within each pair were randomly switched before classes began, and student gains measured near the end of the year. When teacher quality predicts student performance, points hit the bullseye. The more predictions are wrong, the further away points lie. Distance units are .25 standard deviations, which the authors say corresponds to around a year of learning. The actual value of the predicted student learning difference is indicated by the angle around the circle.

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 36. 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, the their odds of large gains the third year are high. Put more simply, the claim is that there are good teachers and bad teachers, and they can be identified simply by monitoring their students’ performance in as short a time as two years.

These findings do not settle the case for teacher quality. There is another possible explanation. Some teachers might regularly do better than others because their classes are regularly different. For example, one teacher might get a large number of struggling students every year. One 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.

To test whether teacher quality can be measured and does not depend upon the particular classroom teachers are given, assign them to random classrooms. It is hard to arrange, but that is exactly what Kane, and Staiger did\textsuperscript{7}. They studied 78 pairs of elementary school teachers for 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 4 the predictions were often off the mark. 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; this means measurement points landed above or below the target equally often. A reliable measurement of teacher quality would actually need to hit the target most of the time. Rewarding or punishing individual teachers when the best measurements of their quality look like Figure 4 not only would be very unreliable it would be unjust. I know of no more careful check of the concept of teacher quality than appears in Figures 3 and 4. And these experiments concern elementary school, where the US does not fare so badly in international comparisons. For secondary schools, where US performance sinks by international standards, test-based measurements of teacher quality have not been made. Data on teacher quality are too weak to justify a national reform program.

The Poverty Argument

When Comets failed, the Royal Aircraft Establishment put an entire airplane in a testing tank (Figure 1). Persuaded its schools were failing, the United States put the whole educational system in a testing tank. Since 2001, every public school student in the country was tested in mathematics and reading or language arts at least once in grades 3-5, once in grades 6-9 and once in grades 10-12. In some states such as Texas, the state where these accountability laws originated, mathematics and reading have been tested every year in all grades 3-11. Data for student performance on these exams are publicly available for almost every school in the country. Data for all the graphs here were obtained from openly accessible websites of various state education agencies.

The masses 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 of 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 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 in schools presenting these problems. Most people who see the connection between poverty and educational outcomes are confident that low-income students in a sufficiently supportive environment will do as well as students in well-off communities.

The concentration of poverty in a school can be defined precisely although not perfectly, by the concentration of students eligible for Free and Reduced-Price Meals. Eligibility is determined by family income, and uniform data are available for every public school in the United States. Many different measures of student performance are available and they 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 or ACT and score 1110 or more on the SAT or 24 or more on the ACT.

The association of poverty concentration with educational performance is very strong (Figure 5). 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

\textsuperscript{7} Thomas J. Kane and Douglas O. Staiger, Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation (2008), \url{http://www.nber.org/papers/w14607}
Fig. 5: Percentage of high school graduates meeting 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. Colors indicate the percentage of minority students in school.

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 and 2007, and the pictures look very much the same. If teacher quality really were the most important factor impacting student achievement, one would have to conclude 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, while 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, and the tests reflect hopes to attend out-of-home-state colleges. But all Texas students take a mathematics exam in 11th grade, and while obtaining a commended score (around 90%) is not as demanding as obtaining 1110 on the SAT, it sets a reasonably high bar and displays a similar pattern (Figure 6). Only a few schools where poverty concentration exceeds 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 7).

The graphs here are only a small sample from a flood of evidence. In every state or year I have examined so far poverty is tightly connected to high school performance in mathematics.
Fig. 6: Percentage of eleventh graders in Texas receiving Commended scores in mathematics 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.

Fig. 7: Percentage of eleventh graders in Texas receiving Passing scores in mathematics 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.
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.” Charter schools are supposed to provide laboratories in which to test new ideas. One cannot expect greatness from all, but some of them should succeed in hiring nothing but highly qualified teachers, and showing what can be accomplished when all elements of the school reform program are brought to bear.

Let us see.

**Fig. 8:** Percentage of high school graduates meeting 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.

Figures 8–10 show fractions of students reaching the SAT college-readiness threshold, commended on 11th grade mathematics, and passing 11th grade mathematics in Texas. The charter schools are highlighted. There are around 140 charter high schools in Texas with 11th grade data, and all but around seven of them are comparable to regular public schools or considerably worse. Most are considerably worse. There are a few exceptional schools where nearly 100% of eleventh grade students are commended in mathematics, but none is a charter school.

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 appear in Figures 11 - 14. I have chosen 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).

---

8 Lesley Chilcott, *Waiting for Superman*
Fig. 9: Percentage of eleventh graders in Texas receiving Commended scores in mathematics 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.

Fig. 10: Percentage of eleventh graders in Texas receiving Passing scores in mathematics 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.
Fig. 11: Mean SAT score of California high school graduates 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 students. Charter schools are highlighted.

Fig. 12: Percentage of Florida high school tenth graders reaching level 5 (Commended) in mathematics, 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 tenth graders. Charter schools are highlighted.
Fig. 13: Percentage of New Jersey 11th graders scoring Advanced Proficient in mathematics, 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 eleventh graders. Charter schools are highlighted.

Fig. 14: Percentage of New York high school students passing Regent’s Mathematics B, 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 students. Charter schools are highlighted.
Fig. 15: Estimates of Texas high schools to be rated unacceptable based on mathematics performance if the current rates of school improvement continue and mathematics standards rise as currently scheduled towards 100% proficiency in 2014. Schools in light red are likely to receive a rating of Academically Unacceptable if judged according to the standard Texas accountability system.

**The Rising Bar for Public Schools**

The comprehensive measurements of public school performance are accompanied with provisions intended to force them to improve. 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*.  

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

---

Fig. 16: Fractions of eleventh-grade students passing TAKS mathematics in 2009 at all the Academically Unacceptable public schools (light red) and all secondary Texas charter schools, with charter schools color coded according to their accountability status and system. Schools colored dark blue are charter schools rated acceptable under Alternative Education Accountability.

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 fails 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 15 provides a rough estimate of which Texas secondary schools may be declared Academically Unacceptable because of mathematics year by year as standards rise. 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 is about to be replaced by another. But the new exam is supposed to be more difficult.

The hero of No Highway finds himself in the sky on 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 four 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. 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.

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 Figure 16, when their mathematics scores fall below the level at which regular public schools would be rated unacceptable, almost all Texas secondary charter schools successfully change their status to be judged by a second system that right now has considerably lower standards. Thus it is quite plausible 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 much lower.

The Future

The purpose of an accident report is to determine the cause of the 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.

Having shown that poverty and not teacher quality is the main element connected to low student performance, and that the reform program brought to bear so far in charter schools shows no signs of providing superior education in secondary mathematics, I will conclude without laying out a complete alternative program, but a few remarks.

School reform cannot succeed if it tries to certify that poverty does not matter. Poverty permeates schools as cracks permeate planes. The schools must be made fail-safe enough to tolerate it. Finding effective and affordable combinations of 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. I believe defining teacher quality mainly 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 themselves studied physics, preferably 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 or not has no bearing on her 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 US, devise solutions, and 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 United States can avoid a similar error, that proponents of teacher quality and charter schools will recognize the weakness of the evidence before it is too late, that we will not damage public education, let down our most vulnerable students, and lose technical leadership we take for granted.