# Visualizing Educational Data

Michael Marder  
Department of Physics, The University of Texas at Austin  
February 9, 2011

## 1 Preliminaries

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well-Off</td>
<td>Student has never been eligible for free and reduced lunch</td>
</tr>
<tr>
<td>Low-Income</td>
<td>Student who is or has been eligible for free and reduced lunch</td>
</tr>
<tr>
<td>Poverty Concentration</td>
<td>Percentage of students in school currently eligible for free and reduced lunch</td>
</tr>
<tr>
<td>Free and Reduced Lunch</td>
<td>Students are eligible for reduced price meals when their family income is 185% of the poverty guideline, and for free meals when their family income is 130% of the poverty guideline. The poverty guidelines for 2010 appear below</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number in Family</th>
<th>Poverty Guideline</th>
<th>Number in Family</th>
<th>Poverty Guideline</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10,830</td>
<td>5</td>
<td>25,790</td>
</tr>
<tr>
<td>2</td>
<td>14,570</td>
<td>6</td>
<td>29,530</td>
</tr>
<tr>
<td>3</td>
<td>18,310</td>
<td>7</td>
<td>33,270</td>
</tr>
<tr>
<td>4</td>
<td>22,050</td>
<td>8</td>
<td>37,010</td>
</tr>
</tbody>
</table>

| NAEP             | National Assessment of Educational Progress, “The Nation’s Report Card.” Assessments of reading and mathematics administered to a sample of around 150,000-200,000 US students in 4th and 8th grade roughly every two years. More infrequently there are exams at 12th grade, and exams in science and social studies. |
|                  |                   |
| TIMSS            | Trends in International Mathematics and Science Study.                                                                                                                                                    |
| PISA             | Programme for International Student Assessment                                                                                                                                                    |
| TAKS             | Texas Assessment of Knowledge and Skills. The Texas high-stakes exam that provides data both for the Texas school accountability system and for the federal accountability system under No Child Left Behind. Mathematics and Reading are tested every year for students from third to eleventh grade, with exceptions for students with language and learning disabilities. |
| Texas SAT Criterion | A score of 1110 on the SAT or 24 on the ACT. Texas uses for SAT only mathematics and critical reading                                                                                       |
| Commended        | A score on the mathematics TAKS exam that varies by grade level and year but is approximately 90%                                                                                                          |

Table 1: Definitions of terms
2 International Comparisons

PISA—Programme for International Student Assessment

Among 30 developed countries, the United States is ranked twenty-fifth in math and twenty-first in science. When the comparison is restricted to the top 5 percent of students, the United States is ranked last.

- Davis Guggenheim (2010), Waiting for Superman

Figure 1: Mathematics proficiency of 15-year-olds comparing various countries. Family economic status is indicated by whether father is low-skill blue-collar worker or high-skill white collar worker.

**Comments:** This exam focuses on students of a specific age ~ 15 years old ~ rather than on a specific grade level. It emphasizes practical application of knowledge over familiarity with procedures and vocabulary. The US indeed ranks very low.

Michael Marder, The University of Texas at Austin 2010. The conclusions of this research do not necessarily reflect the opinions or official position of the Texas Education Agency, the Texas Higher Education Coordinating Board, or the State of Texas.
TIMSS-Trends in International Mathematics and Science Study

Figure 2: Data on eighth grade mathematics performance from TIMSS in 1995 and 2007

Comments: TIMSS more closely resembles US high-stakes exams than PISA. In 2007 the US tied for sixth place in eighth grade mathematics.
Physics

Figure 3: Comparison of US and Chinese freshmen in college physics taking conceptual physics exams. The graphs give a more complete picture than can be obtained just by reporting average scores. For example, in Electricity & Magnetism, there are almost no U.S. students who match a typical Chinese student. After Bao, L. et al “Learning and Scientific Reasoning,” Science, 2009, 323, 586-587

Comments

In much of the rest of the world it is customary for all students to have three or four years of physics in secondary school, while in the United States only 37% of high school students have as much as one year. The difference is apparent. The topics “Mechanics” and “Electricity & Magnetism” refer to the two core subjects in introductory physics. In a test of scientific reasoning where Chinese students receive no more explicit instruction than US students, students in the two countries have nearly identical performance.
3 National Comparisons

New York State’s fourth and eighth graders made no notable progress on federal math exams this year, according to test scores released on Wednesday, sharply contradicting the results of state-administered tests that showed record gains. Across the country, many states posted disappointing results, with fourth-grade students stagnant nationally for the first time in nearly two decades.


![Figure 4: Eighth grade scores in mathematics and reading from the National Assessment of Educational Progress (NAEP) averaged over the period 2000-2009. In each state the average score of students eligible for free and reduced lunch is compared with the score of those who are not.](image-url)
Figure 5: Time development since 2000 of 8th grade NAEP mathematics scores, displaying separately well-off and low-income students in each state. In every jurisdiction scores have risen since 2000.

Comments Eighth grade mathematics scores in every state have risen rather steadily in every state since 2000 for both well-off and low-income eight graders. The magnitude of the rise is comparable to the differences among states, although not yet to the difference between well-off and low-income students. Data at the twelfth grade are not available. For a dynamic representation of these data, see this Motion Plot.
Poverty and Educational Outcomes

Figure 6: Mathematics scores on 8th grade mathematics, averaged over the four administrations of the National Assessment of Educational Progress since 2003. Each disk is a state. For each state students are divided between well-off and low-income.

Comments  Poverty is associated with mathematics test results, both at the individual and at the state level. Within any given state, low-income students get scores around 25 points less than well-off students. The states with the lowest concentration of poverty have scores 20 points higher both for well-off and low-income students than states with the highest concentration of poverty.
Union Status and Educational Outcomes

I believe that what is wrong with our schools in this nation is that they have become unionized in the worst possible way... This unionization and lifetime employment of K-12 teachers is off-the-charts crazy.

– Steve Jobs, Feb 16, 2007

Figure 7: States with weak unions (Right to Work states) are highlighted for both well-off and low-income students. Each disk is a state.

Comments  States with and without strong unions are intermingled. Well-off students in wealthy states with strong unions have the highest outcomes. For low-income students the states with highest outcomes have weak unions. Differences in state performance that might be attributed to unions are small compared with effects of poverty.


Michael Marder, The University of Texas at Austin 2010. The conclusions of this research do not necessarily reflect the opinions or official position of the Texas Education Agency, the Texas Higher Education Coordinating Board, or the State of Texas.
Value-Added Methods for Assessing Teachers

Although it can be difficult to know with much certainty who is likely to be an effective teacher during a job interview, we have shown that school districts can learn a lot about teachers’ future effectiveness simply by scrutinizing their record during their first few years on the job...because most districts have never assembled the data required to calculate the “value-added” by individual teachers, the payoff to beginning to do so could be enormous.

Gordon, Kane, and Staiger (2006) Identifying Effective Teachers Using Performance on the Job

Figure 8: Data of Kane and Staiger for 78 pairs of teachers in Los Angeles, (Kane and Staiger, 2008)

Comments  Advocates of Value-Added methods for assessing teacher performance commend this as the most careful study. It differs from other studies because pairs of teachers were randomly switched between classrooms before the school year began. Kane and Staiger report scaled scores for which they say that 0.25 corresponds to about a year of learning; this conversion was used to obtain the scale in this figure. They used four years of teacher and student data to form predictions about relative performance of paired teachers in a fifth year. Although measured student performance is correlated with past performance, conclusions about individual teachers were inaccurate.

- Seventeen percent of the time the prediction is too low by half a year of learning (±.125) or more
- Thirty-five percent of the time the prediction lies within half a year of learning of the measured value.
- Forty-eight percent of the time the prediction is too large by half a year of learning or more.

Michael Marder, The University of Texas at Austin 2010. The conclusions of this research do not necessarily reflect the opinions or official position of the Texas Education Agency, the Texas Higher Education Coordinating Board, or the State of Texas.
Figure 9: Average year-to-year changes in Texas student mathematics scores averaged over 2003 through 2007. Students are grouped according to the poverty concentration in their school, and according to their score the year before.

Comments For secondary school mathematics, there are factors in addition to poverty that are strongly associated with changes in student scores. The students with the highest score gains are eleventh-graders in schools with little poverty who got 20-30% of the problems right in tenth grade, and the students with the lowest score gains are ninth-graders in schools with high poverty who got 90-100% of the problems right in eighth grade. Student gains in a teacher’s class can therefore strongly depend upon the grade level and prior score of the students.
4 Texas Data

There is no doubt that poverty presents real challenges, and it is harder to be a principal or teacher in an urban district than it is in the suburbs. But even in the toughest of neighborhoods and circumstances, children excel when the right adults are doing the right things for them.

--Michelle Rhee (2010), Waiting for Superman

SAT College-Readiness Criterion

Figure 10: 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.

Comments Figure 10 is a scatter plot showing Texas high schools as a function of the fraction of students eligible for free and reduced lunch. The vertical axis shows percentage of graduates meeting the SAT/ACT College Readiness criterion in 2008. Above a poverty concentration of 80% there are no examples of Texas schools where more than 20% of the graduates meet the SAT Criterion.
### SAT Histograms, 2003-2009

<table>
<thead>
<tr>
<th>Percentage Graduates Meeting SAT Criterion</th>
<th>Poverty Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>60.70%</td>
<td>100K</td>
</tr>
<tr>
<td>50.60%</td>
<td>50K</td>
</tr>
<tr>
<td>40.50%</td>
<td>100K</td>
</tr>
<tr>
<td>30.40%</td>
<td>50K</td>
</tr>
<tr>
<td>20.30%</td>
<td>100K</td>
</tr>
<tr>
<td>10.20%</td>
<td>50K</td>
</tr>
<tr>
<td>0.10%</td>
<td>100K</td>
</tr>
</tbody>
</table>

**Minority Percent**

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Color Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0%</td>
<td>#0000FF</td>
</tr>
<tr>
<td>20.0%</td>
<td>#00FF00</td>
</tr>
<tr>
<td>40.0%</td>
<td>#FF6347</td>
</tr>
<tr>
<td>60.0%</td>
<td>#FF00FF</td>
</tr>
<tr>
<td>80.0%</td>
<td>#FF0000</td>
</tr>
<tr>
<td>100.0%</td>
<td>#0000FF</td>
</tr>
</tbody>
</table>

Figure 11: Histogram showing numbers of students in schools where 0-10%, 10-20%, etc. students met the Texas SAT/ACT College Readiness Criterion, grouped by fraction of students in the school eligible for free and reduced lunch. The data come from 2003-2009, and show the average numbers of students per year in each category.

**Comment** Data such as in Figure [10](#) are grouped together for the period 2006-2009, showing numbers of students in every category of poverty concentration and SAT achievement rather than individual schools. Each bar is color coded to indicate the fraction of minority students in that category. In Texas schools where 80% to 100% of the children are poor, the odds that more than 20% will be College-Ready by the SAT criterion are zero.

---

Michael Marder, The University of Texas at Austin 2010. The conclusions of this research do not necessarily reflect the opinions or official position of the Texas Education Agency, the Texas Higher Education Coordinating Board, or the State of Texas.
Time Dependence of SAT College-Readiness Criterion

**Figure 12:**

**Comments** Changes in the level of school poverty over time are correlated with changes in the Texas SAT/ACT College Readiness Criterion. The effect is most dramatic for schools with low levels of poverty, as shown in Figure 12. The plot shows all schools where both the change in poverty concentration and the change in student fraction meeting the SAT criterion over the period 2006-2009 are described by a linear relation with correlation coefficient $r$ of 0.5 or greater. The lower the concentration of poverty, the more rapid is the change of student outcomes with changes in poverty level. This finding is consistent with the fact that the slope of the curve in Figure 10 is greatest on the left hand side of the plot.
Change of Individual Students over Time

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Well Off</td>
<td><img src="image1" alt="Plot" /></td>
<td><img src="image2" alt="Plot" /></td>
<td><img src="image3" alt="Plot" /></td>
</tr>
<tr>
<td>Low Income</td>
<td><img src="image4" alt="Plot" /></td>
<td><img src="image5" alt="Plot" /></td>
<td><img src="image6" alt="Plot" /></td>
</tr>
</tbody>
</table>

Figure 13: Math flow plots. The vertical axis separates students according to whether or not they have ever been eligible for free and reduced lunch. The horizontal axis shows grade levels.

Comment Figure 13 provides a graphical representation of the year to year changes of student mathematics scores in Texas, separating out well-off and low-income students. Each arrow shows the average score one year later of students who shared a certain starting score in the previous year. The plots provide snapshots from single year and do not show the movement of a single cohort of students. The area of an arrow is proportional to the number of students making the transition, with large arrows corresponding to tens of thousands of students. [Methods described by Marder and Bansal, *PNAS*, 2009]

---

Michael Marder, The University of Texas at Austin 2010. The conclusions of this research do not necessarily reflect the opinions or official position of the Texas Education Agency, the Texas Higher Education Coordinating Board, or the State of Texas.
Figure 14: (A) Flow streamlines showing the different predicted paths of well-off and low-income students who start with the same score in third grade. (B) The rate at which the performance gap between well-off and low-income students increases at each grade level, for 2006-2007 through 2008-2009.

Comment  By comparing the changes at every score and grade level in Figure 13, one can construct the graph in Figure 14. In part (A) note that well-off students who have a 60% score at third grade flow to a score of 67% by eleventh grade, well above the eleventh grade score of low-income students who started at 90% at third grade. Part (B) shows the rate at which the gap between well-off and low-income students increases each year. The largest rates of increase occur at sixth and ninth grades.
Value-Added Measures for Schools

Figure 15: Schools are separated into four quartiles according to score gains in 9th grade mathematics, 2004-2006. These gains are then used to predict the score gains in 2007. The lowest-quartile schools obtain on average a 5-point smaller gain than the highest-quartile schools. Poverty is more concentrated in the lowest-quartile schools than the highest-quartile schools.

Comments Figure 15 separates schools into four quartiles depending upon the magnitude of student score gains in ninth-grade mathematics, averaging over 2004-2006. For each of the four quartiles, the probability of score changes in 2006-2007 is plotted as a result. The right hand plot compares lowest-quartile schools with highest-quartile schools, color coded for the concentration of poverty. The lowest-quartile schools have a higher concentration of poverty.
Figure 16: Lower-performing and higher-performing schools in Houston and Dallas.

Comment Figure 16 shows the geographic distribution of schools in the Dallas and Houston areas, color coded according to whether score changes in ninth-grade mathematics were positive on average during 2004-2006 (higher-performing) or negative (lower-performing). The background colors indicate per capita income by zip code. Lower-performing schools are clustered together in low-income areas, with some exceptions.
School Size

*In small schools, student achievement increases, particularly for minority and low-income students.*

— http://www.smallschoolsproject.org

Figure 17: Graduates meeting SAT/ACT College Readiness Criterion [1100 on SAT or 24 on ACT] versus school size and poverty concentration for (almost) all Texas high schools in 2009. Data from the Texas Education Agency Academic Excellence Indicator System (AEIS).

Comments For every level of poverty, increasing school size is correlated with increasing fractions of students meeting the criterion. The effect is strongest for low concentrations of poverty. For each concentration of poverty, there are some relatively high-scoring schools that are also small.
5 Charter Schools

Charter schools give children the opportunity to achieve their dreams. Across America, parents are clamoring for charter schools—boosting waiting lists at America’s 5,000 charter schools to an all-time high, enough to fill another 5,000 charters.

–Center for Education Reform

Figure 18: Texas secondary charter schools and SAT college readiness criterion. All charter schools are colored in. Every disk is a school, and the area of the disk is proportional to the number of graduates.

Comment: Almost all Texas charter high schools are considerably worse than comparable public schools by the SAT College-Readiness measure. Only one charter school stands out as comparable to the best public schools.
Comment: When Commended status in eleventh-grade mathematics is used as the measure of performance, five secondary Texas charter schools (YES North, YES Southeast, Idea, Peak Preparatory High School, and KIPP) stand out with strong performance relative to public schools. Another 35 charter schools have performance comparable to public schools, and the remaining 100+ are worse.

Michael Marder, The University of Texas at Austin 2010. The conclusions of this research do not necessarily reflect the opinions or official position of the Texas Education Agency, the Texas Higher Education Coordinating Board, or the State of Texas.
Figure 20: Texas high school charter schools and student gains. Score changes have been averaged over the years 2003-2007. Error bars on charter schools come from the sequence of annual score changes. The figure shows the mean score changes of students in all public schools with the same levels of poverty, compared with individual charter schools.

Comment. For every level of starting performance and every concentration of poverty, the majority of secondary charter schools produces smaller student learning gains than the average result obtained by conventional public schools.
Texas Charter School Accountability

Figure 21: Fractions of eleventh-grade students passing TAKS mathematics in 2009 at all the Academically Unacceptable public schools (red) and all the Texas charter schools, with charter schools color coded according to their accountability system.

Comment The majority of secondary charter schools are subject to the Alternative Education Accountability system, where the cutoffs for acceptable performance are much lower than in the Standard system. The schools colored blue are Acceptable according to Alternative Education Accountability, but most of them have mathematics performance worse than the public schools that are Academically Unacceptable under Standard Accountability.
California Charter Schools

California Students, Mean SAT, 2008

California Charter School Students, Mean SAT, 2008

Figure 22: Mean score of California seniors taking the SAT, color coded by the percentage of graduates who take it. Data from 2008, source, California Department of Education. The lower panel colors in only charter schools.

Comments. California has one exceptional charter school, where all students are low-income, all take the SAT and all do well (Preuss School, UCSD). Fifteen other charter schools encourage large fractions of low-income students to take the SAT but their scores are low. The remaining, roughly 25 high schools with reported SAT averages are comparable to other public schools.

Michael Marder, The University of Texas at Austin 2010. The conclusions of this research do not necessarily reflect the opinions or official position of the Texas Education Agency, the Texas Higher Education Coordinating Board, or the State of Texas.
Figure 23: Fractions of Florida 10th graders with score of 5 on the FCAT mathematics exam; level 3 is passing. Every disk is a school, and charter schools are highlighted in color.

Comment: Florida has around 10 secondary charters with higher mathematics performance at tenth grade than comparable public schools. Most of the remainder are comparable to regular public schools, and around 30 are worse.
New Jersey Charter Schools

Figure 24: Fractions of New Jersey 11th graders rated advanced proficient in mathematics at 11th grade. Every disk is a school, and charter schools are highlighted in color for the lower panel. 

Comment: New Jersey has around 11 secondary charters and only one of them can be compared with the better non-charter public schools for 11th grade mathematics.
New York Charter Schools

Figure 25: Percentage of New York high school students taking Regent’s Mathematics B who score 85-100%. Every disk is a school, and charter schools are highlighted in color for the lower panel. 

Comment: Only 4 New York State charter schools have public data for Regent’s Mathematics B, which indicates readiness for college-level mathematics.

Michael Marder, The University of Texas at Austin 2010. The conclusions of this research do not necessarily reflect the opinions or official position of the Texas Education Agency, the Texas Higher Education Coordinating Board, or the State of Texas.
Maxims

1. (Golden Rule) Provide the education for others’ children you would have for your own.

2. (Dobzhansky’s Law) Nothing makes sense in public education except in light of poverty.

3. (Occam’s Razor) When all other explanations fail, assume professionals are acting in their self-interest.

4. (Uncertainty Principle) Educational data tend to be most sparse where the interests of the country are most at stake.