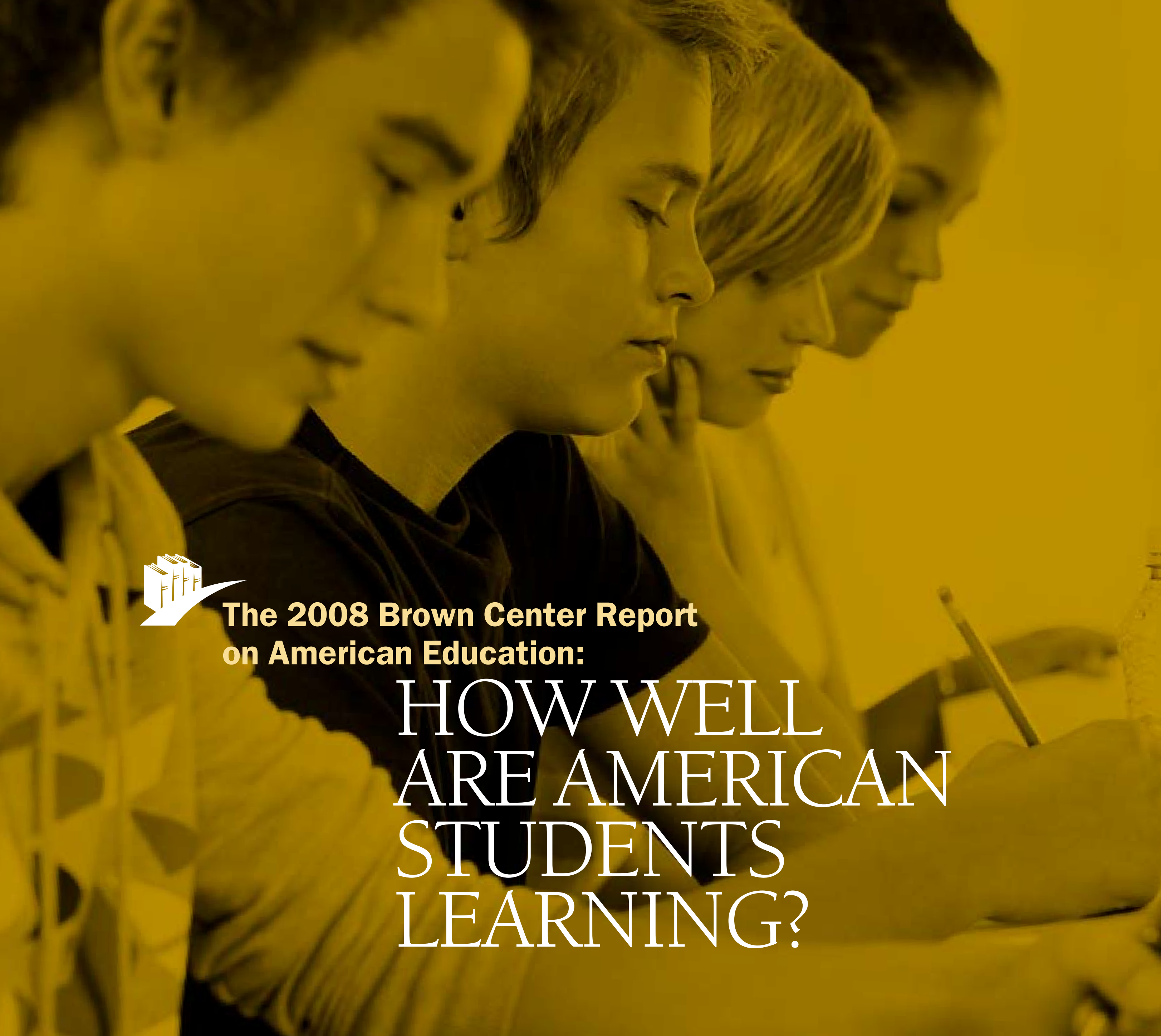


The 2008 Brown Center Report on American Education:

HOW WELL ARE AMERICAN STUDENTS LEARNING?

With sections on international assessments, the misplaced math student, and urban schools

BROOKINGS

ABOUT BROOKINGS

The Brookings Institution is a private nonprofit organization devoted to research, education, and publication on important issues of domestic and foreign policy. Its principal purpose is to bring knowledge to bear on current and emerging policy problems. The Institution maintains a position of neutrality on issues of public policy. Interpretations or conclusions in Brookings publications should be understood to be solely those of the authors.

BROWN CENTER ON EDUCATION POLICY

Established in 1992, the Brown Center on Education Policy conducts research on topics in American education, with a special focus on efforts to improve academic achievement in elementary and secondary schools. For more information, see our website, www.brookings.edu/brown.aspx.

To order copies of this report, please call 800-537-5487, fax 410-516-6998, e-mail hfcustserv@press.jhu.edu, or visit online at www.brookings.edu/brown.aspx.

This report was made possible by the generous financial support of The Brown Foundation, Inc., Houston.

We are grateful to the National Center for Education Statistics and the Educational Testing Service for supplying the NAEP restricted-use data and answering our questions.

The 2008 Brown Center Report on American Education:

HOW WELL ARE AMERICAN STUDENTS LEARNING?

With sections on international assessments, the misplaced math student, and urban schools

January 2009
Volume II, Number 3

by:
TOM LOVELESS
Senior Fellow, Brown Center
on Education Policy

TABLE OF CONTENTS

Research assistance by:

KATHARYN FIELD-MATEER
Brown Center on Education Policy

THE 2008 BROWN CENTER REPORT ON AMERICAN EDUCATION

The watchword of this year's Brown Center Report is caution—caution in linking state tests to international assessments—"benchmarking" is the term—caution in proceeding with a policy of "algebra for all eighth graders," caution in gleaning policy lessons from the recent progress made by urban schools. State and local budget woes will restrain policymakers from adopting costly education reforms, but even so, the three studies contained herein are a reminder that restraint must be exercised in matters other than budgets in governing education well. All too often, policy decisions are based on wishful thinking rather than cautious analysis. As education evolves as a discipline, the careful analysis of high-quality data will provide the foundation for meaningful education reform.

The report consists of three sections, each discussing a separate study. The first section looks at international testing. Powerful groups, led by the National Governors Association, are urging the states to benchmark their state achievement tests to an international assessment, the Programme for International Student Assessment (PISA). After comparing PISA to the Trends in International Mathematics and Science Study (TIMSS), the other major international assessment in which the United States participates, the Brown Center analysis examines findings from a chapter of the 2006 PISA report that addresses student engagement. The chapter presents data on students' attitudes, values, and beliefs toward science.

Benchmarking proponents argue that PISA offers policy guidance to American school officials by identifying the characteristics of successful school systems around the world. The Brown Center analysis calls that claim into question. The PISA report makes causal claims from cross-sectional data that cannot support such inferences. The chapter on student engagement presents inferences based on selective treatment of data, with policy recommendations going beyond the evidence adduced to support them.

Moreover, PISA poses questions that contain ideological bias. To define scientific literacy as encompassing beliefs as well as knowledge—a definition also embraced by skeptics of evolution—is a dubious position for any science assessment to take. PISA wants to assess whether students are capable of applying science to public policy. Fair enough. That capacity can be evaluated, however, without making a judgment about students' political beliefs. PISA asks students whether they support several environmental policies and then creates an index of "responsibility for sustainable development" from the responses. Responses in favor of the policies are responsible; those opposed are not. That kind of questioning is inappropriate on a science assessment. Without serious reform, PISA is inappropriate for benchmarking.

The second section tackles another hot topic in policy circles—whether all eighth graders should take an algebra course. California recently adopted a universal eighth grade algebra policy that will be implemented in 2011, joining a Minnesota policy with the same objective and implementation date. Are all eighth graders prepared to take an algebra class? National data are examined from eighth grade math classes in 2005 to answer that question.

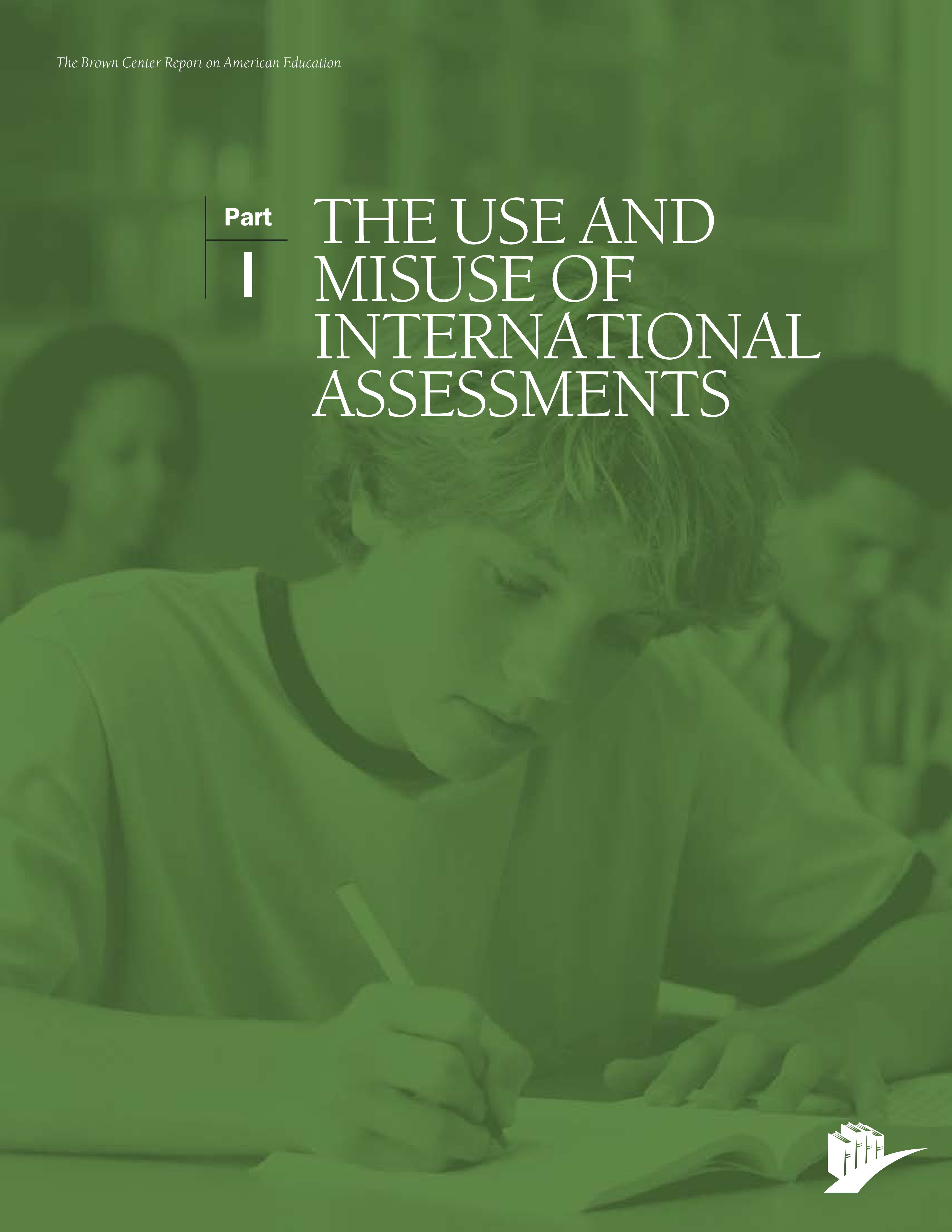
Low achievers in mathematics, those scoring in the bottom tenth of all students, function several years below grade level. A shocking percentage of these low achievers, 28.6 percent, were enrolled in advanced math courses—Algebra I, Algebra II, or Geometry—in 2005. A policy of algebra for all eighth graders will dramatically increase the proportion of these misplaced math students. Sample math items are presented to illustrate the large gaps in the misplaced students' mathematical knowledge, in particular, their poor grasp of fractions, decimals, and percentages. The misplaced students are described in terms of demographic characteristics, the schools they attend, and the teachers who are instructing their math classes. The portrait is deeply troubling. The misplaced students are some of the nation's most vulnerable youngsters. The analysis raises questions about the feasibility of an "algebra for all" policy until we know how to reduce the number of underprepared students and how to effectively teach algebra to students who struggle with basic arithmetic.

The final section of the report is a good news story. The 2001 Brown Center Report presented an analysis of academic achievement in big city school districts. That study compared test scores for school districts serving the top fifty cities in the 2000 U.S. Census to the average test score in the cities' respective states. Not surprisingly, the big city districts lagged far behind. This year's report replicates that study using the most recent achievement data. Big city schools have made significant gains. While all school districts have notched achievement gains, the big city districts made even larger gains than other districts. They are closing the gap with suburban and rural districts, slowly, to be sure, but they are clearly making progress.

The analysis does not hazard a theory as to why big city achievement is rising. One possible catalyst is mayoral control, a popular urban reform in recent years. The data neither support nor refute the effectiveness of mayoral control. Another possible influence is No Child Left Behind.

The law targets low-performing students, and studies of test scores at both state and national levels have shown greater progress at the bottom of the achievement distribution than at the top. Having a disproportionate share of low achievers, big city schools benefit from that trend. As noted above, cross-sectional data are limited in what they can reveal about the causes of events, so whether NCLB has played a role in the progress of big city schools is merely speculative. In addition, not all big city districts have made gains.

A daunting obstacle to determining the drivers of academic trends is that there is no authoritative source that documents the policies that local districts have adopted, along with such details as when particular policies were started, when they were modified, what policies they replaced, and how they were implemented. The Brown Center Report ends with a call for a periodic national inventory of district policies across the country. We are getting much better at determining how well students are learning and tracking trends in test scores as they unfold over time. But policy analysis lags behind. Explaining why students are learning more or less—and really pinpointing the causes of trends in achievement—will take much more information about the policies and practices of our schools.



Part I

THE USE AND MISUSE OF INTERNATIONAL ASSESSMENTS

THIS SECTION OF THE BROWN CENTER REPORT USUALLY reviews the latest data from state, national, and international assessments to answer the question: how well are American students learning? The tests given to assess students are also investigated and discussed to give readers a better understanding of the instruments used to measure learning. This year international tests are scrutinized. The section first describes two international testing programs in which the United States participates. Then one of the tests, the Programme for International Student Assessment (PISA), is given special attention.

PISA is influential in Europe but it flies below the radar in the United States. That is about to change. The National Governors Association (NGA), backed by other powerful groups in Washington, would like states to use PISA as an international benchmark of student performance. Especially attractive to PISA supporters are the numerous policy recommendations that PISA officials make based on test results. In September 2008, the NGA, Achieve, Inc., and the Council of Chief State School Officers (CCSSO) announced the creation of an advisory group to pursue benchmarking to PISA. As explained by co-chair Governor Sonny Perdue of Georgia, "As governors, we must have consistent, comparable data in order to make informed decisions about our state's education system. Benchmarking will help us identify the qualities and characteristics that make up the education systems that best prepare students for success. Understanding these policies give us the option of incorporating the best of them into our own educational structure."[1]

Background

The United States participates in two K–12 international testing programs: PISA, which is administered by the Organisation for Economic Co-operation and Development (OECD), and the assessments of the International Association for the Evaluation of Educational Achievement (IEA). The Trends in International Mathematics and

Powerful groups in Washington would like states to use PISA as an international benchmark of student performance.

Comparison of PISA and TIMSS on Key Characteristics

Table **1-1**

Characteristic	PISA	TIMSS
Governance	**Government representatives**	**Researchers and government representatives**
Sample	**Age-based: 15-year-olds**	**Grade-based: 4th and 8th grades**
Philosophy of assessment	**Measures the ability to apply what has been learned to real-world situations (socio-constructivist)**	**Measures what has been learned in the school curriculum**
Scope	**Learning inside and outside of school, including attitudes, values, and beliefs**	**Topics in school curriculum**
Content—Math	**Mathematical literacy: Space and shape, change and relationships, quantity, uncertainty**	**Grade 4 mathematics: Number, geometric shapes and measures, data display** **Grade 8 mathematics: Number, algebra, geometry, data and chance**
Content—Science	**Scientific literacy: Physical systems, living systems, earth and space systems, technology systems**	**Grade 4 science: Life science, physical science, earth science** **Grade 8 science: Biology, chemistry, physics, earth science**
Policy Recommendations	**Numerous**	**Sparse**

Source: Compiled by author. Also see Dougal Hutchison and Ian Schagen (2007).[2]

Science (TIMSS) is the best known of IEA tests. The TIMSS test assesses both math and science and is given every four years. It was last given in 2007 and will be given again in 2011. The PISA program tests reading literacy, mathematics literacy, and science literacy every three years. All three subjects are assessed in each sitting although one is singled out as the main topic. Science was the focus in 2006. Reading will be the focus in 2009 and math in 2012.

PISA and TIMSS are governed differently due to their parent organizations' different histories. The IEA originated in 1958. Researchers from a small band of nations, including the United States, believed that much could be learned about education by studying schooling around the world. The IEA General Assembly, consisting of researchers and governmental representatives, exercises authority over the tests.[3] In 2007, sixty-six nations participated in TIMSS.

The OECD is a quasi-governmental organization based in Paris, France, dedicated to promoting economic development. The OECD includes representatives from the national governments of thirty nations, mostly European countries but also the United States and other industrialized nations. PISA was first administered in 2000. As a test created by governments rather than researchers, PISA is governed by a board representing national ministries and departments of education. Counting non-OECD partner nations taking the test, a total of fifty-seven countries participated in the 2006 PISA.

How Do PISA and TIMSS Differ?

Table 1-1 shows several contrasts between PISA and TIMSS. PISA uses an age-based sample, targeting all 15-year-olds enrolled in school. This age was selected because for many nations it represents the age at which compulsory schooling is about to end, hence the claim that the test measures the "yield" of educational systems. TIMSS assesses students in fourth and eighth grades. The two sampling designs produce different kinds of variation among the students selected. PISA's 15-year-olds are enrolled in several grades and vary in the amount of schooling they have experienced. TIMSS students are all in fourth or eighth grade but vary in age.

The two tests have different philosophies and scopes. TIMSS assesses how well students have learned school mathematics and science. The test's content comprises common topics that schools around the world teach in the two subjects. PISA, on the other hand, assesses the ability to apply learning in real-world situations. The content assesses mathematical and science "literacy," not mathematics and science. These are broader domains encompassing skills and knowledge learned both inside and outside of school. When the OECD states that PISA measures the yield of educational systems, the term "educational systems" is meant expansively. It includes families, peers, communities, and popular culture, not just schools.

The philosophical underpinnings of PISA are socio-constructivist, that is, a combination of constructivist and situated learning theory.[4] The focus on what students can do when presented with novel problems, not on whether they can reproduce the skills and knowledge that schools have taught them, mirrors the constructivist emphasis on building knowledge. In addition, competencies are seen as developed socially, that is, students learn from the totality of their interactions with social settings—not only with school curricula—and these competencies include both cognitive and non-cognitive outcomes.[5]

As a consequence of targeting learning that may be independent of the school curriculum, the authors of PISA 2006 caution, "If a country's scale scores in reading, scientific or mathematical literacy are significantly higher than those in another country, it cannot automatically be inferred that the schools or particular parts of the education system in the first country are more effective than those in the second. However, one can legitimately conclude that the cumulative impact of learning experiences in the first country, starting in early childhood and up to the age of 15 and embracing experiences both in school and at home, have resulted in higher outcomes in the literacy domains that PISA measures."[6]

This casts doubt on whether PISA can appropriately serve as an international benchmark. Additional doubts are raised by examining data on some of the non-cognitive outcomes in PISA's 2006 assessment of science literacy.

Attitudes, Values, and Beliefs in PISA

PISA collects extensive information on students' attitudes, values, and beliefs. As explained in the PISA 2006 report, "In PISA, attitudes are seen as a key component of an individual's science competency and include an individual's values, motivational orientations, and sense of self-efficacy."[7] National data on several affective characteristics are tabled along with test scores, giving readers an idea of whether, for example, science performance is statistically related to students' self-concepts or awareness of environmental issues.

The data on attitudes, values, and beliefs are examined below. First, let's step back and review how correlations are used in interpreting international test score data.

Understanding Correlations

Correlational relationships must be interpreted with caution. Beginning students of statistics learn that correlation is not causation. Repeated observations of roosters crowing before sunrise, even if millions of roosters are observed crowing over a long period of time, do not mean that roosters influence the behavior of the sun. The direction of causality—or that a causal connection exists at all—cannot be determined from the persistent coincidence of two phenomena. Correlations can be used to generate hypotheses about causality, but evidence of other possible influences, such as the rotation of the earth, must be considered before drawing any conclusions.

Analysts typically calculate two types of correlations with international assessment data: between-nation and within-nation. (Apologies for the grammatically incorrect terminology, but even when more than two nations are compared, the relationship is dubbed "between" and not "among.") Between-nation correlations answer such questions as whether national test scores are related to national averages on other variables. For example, on the question of whether enjoying science is related to learning science, between-nation correlations tell us whether countries in which students say they enjoy science score higher or lower than countries in which students do not enjoy science as much.

Within-country correlations are based on data from students in the same country. On the enjoyment question, they address whether American students who enjoy science more score higher in science knowledge than American students who enjoy it less, Swedish students who enjoy science more score higher than Swedish students who enjoy it less, and so on across several countries. What does this mean for trying to determine the "real" relationship between two variables? It is not that one form of correlation is necessarily more valid than the other. They answer different questions based on data aggregated at different levels.

PISA reports are chock full of policy recommendations.[8] Some are based on between-nation correlations, some on within-nation correlations, and others on multivariate regression analyses, a more sophisticated form of analysis that examines several variables at once and isolates the effect of one variable while holding the others constant. Recommendations are given on policies that range from school finance to testing and accountability, parental choice, universal pre-school, and whether students should be grouped by ability for instruction. Some of the statistical analyses are more sophisticated than others, but none of the techniques escapes the fact that PISA data are cross-sectional and, like millions of data points on crowing roosters and rising suns, cannot be used to show causality.[9]

None of the techniques escapes the fact that PISA data are cross-sectional and ... cannot be used to show causality.

Self-Efficacy and PISA

Chapter 3 in the 2006 report focuses on student engagement. The term can mean many things. In PISA, student engagement consists of: support for scientific inquiry, self-beliefs as learners, interest in science, and responsibility toward resources and environments. In all, fifteen correlation coefficients are presented modeling the association of these constructs with scientific literacy.

One of the variables under "self-beliefs as learners," self-efficacy, refers to self-confidence in overcoming challenges and mastering particular learning tasks. It is a close cousin to the much maligned notion of self-esteem or self-concept, which refers to how learners feel about themselves.[10] On PISA's self-efficacy item, students are asked whether they can perform several tasks easily, with a bit of effort, after a struggle, or not at all. Examples of tasks are: explaining why earthquakes occur more frequently in some areas than in others, recognizing the science question underlying a newspaper report on a health issue, and describing the role of antibiotics in the treatment of disease. Students are not actually asked to complete any of these tasks—only if they think they can complete them.

Before proceeding it is important to point out that self-efficacy is not always a virtue. If an unsupervised two-year-old crawls into an automobile, it is much safer for everyone if the toddler lacks the confidence to drive the car. As it pertains to education, self-efficacy has been shown to sometimes exhibit a negative relationship

Students are not actually asked to complete any of these tasks—only if they think they can complete them.

with learning. Classroom studies indicate that struggling students who are willing to ask for help—exhibiting the opposite of self-efficacy—are more likely to perceive their own weaknesses and get the assistance that they need to overcome learning difficulties. Sometimes knowing that one does not know is a precondition for learning.[11]

The difficulty of untangling cause and effect with cross-sectional data is especially relevant to this topic. In the literature examining the relationship, researchers are divided on whether self-efficacy boosts learning, learning boosts self-efficacy, or the two boost each other.[12] In many studies, the two constructs are negatively related, with high-efficacy students scoring low on knowledge.[13]

Researchers have long puzzled over a curious relationship between national test scores in mathematics and student reports of self-confidence. A 2006 Brown Center analysis of TIMSS data documented that average national math performance is inversely related to national indices of what we called "the happiness factor," including students' confidence in their math skills. Within countries, the relationship between confidence and test scores is positive, but between countries, the relationship is negative. The higher a nation's happiness factor, the lower its TIMSS score. American eighth graders believe they are good at mathematics. Singaporean eighth graders are not as confident in their math abilities. But on the TIMSS mathematics test, Singaporean eight graders score light years ahead of eighth graders in the United States.[14]

What Do PISA Authors Conclude About Self-Efficacy?

The 2006 PISA data show a positive within-nation correlation between self-efficacy and science achievement (the report calls this "performance"). In presenting the data, the PISA authors recognize the tendency to confound correlation and causality. They exercise due caution by calling the relationship between self-efficacy and performance an "association." But as revealed in the following statement, when it comes to policy recommendations, PISA authors try to have their cake and eat it too. "PISA cannot show to what extent lack of self-efficacy is a cause or an effect of weakness in scientific literacy, but this strong association shows that building students' confidence in their ability to tackle scientific problems is an important part of improving science performance."[15] The first clause in the sentence confesses that the PISA data cannot show cause and effect. The second clause asserts a cause and effect anyway, that building confidence improves science performance.

On page 137 of the report, a box containing a discussion of the cause-effect conundrum goes even further. Box 3.4 is entitled: "Do students' beliefs about their abilities simply mirror their performance?" The opening paragraph foreshadows the answer to that question: "One issue that arises when asking students about their own abilities, especially in terms of whether they can perform scientific tasks, is whether this adds anything of importance to what is known about their abilities from the assessment. In fact, both prior research and the PISA results give strong reasons for assuming that confidence helps to drive learning success, rather than simply reflecting it."[16]

Let's recall Governor Perdue's reason for wanting to benchmark his state's test scores to PISA, to help identify the characteristics of educational systems that best prepare students for success. Readers are told that PISA results show that "confidence helps to drive learning success" and that "building students'

Correlations of Attitudinal Variables and Student Performance on PISA 2006
(Rank-ordered by between-nation coefficient)

Table 1-2

Variable	Between-Nation (Brown Center)	Pooled Nations (OECD)
Students' awareness of environmental issues	0.66	0.43
Self-efficacy in science	0.23	0.33
Students' responsibility for sustainable development	-0.24	0.18
Students' level of concern for environmental issues	-0.48	0.01
General value of science	-0.49	0.22
Support for scientific enquiry	-0.52	0.25
Students' optimism regarding environmental issues	-0.52	-0.17
Personal value of science	-0.72	0.12
Self-concept in science	-0.73	0.15
Enjoyment of science	-0.76	0.19
Instrumental motivation to learn science	-0.77	0.09
Students' science-related activities	-0.79	0.04
Interest in scientific topics	-0.80	-0.06
General interest in science	-0.82	0.13
Future motivation to learn science	-0.83	0.08

positive values

Note: Between-nation correlations are Pearson-product moment correlations of 1) national means on science performance and 2) national means on attitudinal variables, computed by Brown Center. Correlations based on pooled OECD nations are reported by OECD, Table A10.3, page 373, Vol. 1, PISA 2006.

[PISA's] message is that policymakers should implement a science curriculum that includes confidence-building.

confidence is an important part of improving science performance." The message is that policymakers should implement a science curriculum that includes confidence-building.

What is the evidence? How strong is the relationship between student achievement and self-efficacy?

What the Data Show

To better understand the data on self-efficacy, let us first examine how all of the variables dealing with attitudes, values, and beliefs are treated—the affective characteristics assessed by PISA—in measuring their relationship with science achievement.

The PISA 2006 authors present a scorecard of within-nation correlation coefficients (see Table A10.3 on page 374 in Volume 1 of the 2006 report). Almost all are positive, which is consistent with previous research. PISA also reports correlation coefficients from a pooled sample that includes all students, inversely weighted by national sample size, which serves as the cross-national indicator of association. This technique is different from a traditional correlation of national means, but similar enough for the PISA authors to comment on specific countries based on the results. The accuracy of the calculation and the validity of the methodological approach are not in question here. The point is that this method produces correlations that are almost all positive, diverging from previous research.

To explore this matter, we calculated between-nation correlation coefficients in the traditional manner. The results are consistent with previous research—and quite different from the results reported in the PISA report. Table 1-2 compares our results, shown in the first column, with PISA results, shown in the second column. PISA reports correlation coefficients separately for OECD and partner nations. Here the OECD nations' correlations are displayed to simplify the comparison (they are not dramatically different from the partner nations). We computed one statistic for all nations possessing a full panel of data.

The shaded cells designate positive coefficients. Calculated by the traditional between-nation method, only two of the variables are positively correlated with achievement. Thirteen are negative, and most produce coefficients less than -0.50, among them personal value of science (-0.72), self-concept in science (-0.73), enjoyment of science (-0.76), and future

Calculated by the traditional between-nation method, only two of the variables are positively correlated with achievement. Thirteen are negative.

motivation to learn science (-0.83). The higher a nation registers on these characteristics, the lower a nation scores on science achievement.

Using the pooled method, OECD reports thirteen variables with positive relationships. PISA reports that many of these coefficients are statistically significant, but this may be because the pooled method, by using students rather than nations as the unit of analysis, increases the number of observations from a few dozen to hundreds of thousands. Statistical significance is much easier to find with large samples. Even if they reach statistical significance, all but two of the PISA correlations are trivial in a real-world sense.

As just noted, those two variables exhibit the only positive correlations produced using the traditional between-nation approach. They are for environmental awareness and self-efficacy. The relationship between performance and environmental awareness is strongly positive (r = 0.66). One can square a correlation coefficient to produce a statistic, r^2, indicating the percentage of variance explained by the association. The r^2 of 0.44 means that environmental awareness explains about 44 percent of the variation among nations in science scores. Since these are correlations and the causal relationship is unknown, it can also be said that achievement in science explains about 44 percent of the variance among nations in students' environmental awareness. PISA's environmental questions are examined below.

The coefficient for self-efficacy is positive but weak (0.23). Only about 5 percent of the variance between nations in performance can be explained, if the relationship is in fact causal as the PISA authors argue, by differences in national means on the self-efficacy index.

PISA's questions on the environment raise concerns of a different type.

PISA and Political Ideology

PISA asks students questions about the environment on the following topics: awareness of environmental issues, level of concern for environmental issues, optimism regarding environmental issues, and level of responsibility for sustainable development. These topics separate beliefs from knowledge and introduce political ideology into PISA. Let's look at two items.

The item on environmental awareness asks students to report how aware they are of the following issues:

- *The release of greenhouse gases in the atmosphere*
- *The use of genetically modified organisms (GMO)*
- *Acid rain*
- *Nuclear waste*
- *The consequences of clearing forests for other land use*

Response levels range from "I am familiar with this and I would be able to explain this well" to "I have never heard of this." As noted above, the authors report that awareness of environmental issues correlates highly with science performance. A one-level increase on the four-level awareness index is associated with a 44 point increase in performance on the scientific literacy test (about one-half standard deviation). Remember that the between-nation correlation coefficient computed by the Brown Center also shows a strong positive relationship (r=0.66). Not surprising, of course, just as students who know more about economics have a greater awareness of economic issues, so too PISA test takers who know more about

An obvious danger in asking students about public problems is that political ideology may creep into the test.

the environment have a greater awareness of environmental issues.

An obvious danger in asking students about public problems is that political ideology may creep into the test. The danger is acute in PISA because the attitudinal questions elicit students' beliefs about issues, not their knowledge of issues. From Galileo to the Scopes trial, history teaches us that beliefs untethered from knowledge can make for some rather unscientific decisionmaking. At the heart of the scientific method is a willingness to test beliefs in the form of hypotheses and either confirm or reject them based on evidence.

Ideology creeps into how issues are posed here. Asking about the consequences of clearing forests casts the question in a different political light than asking about the benefits (which can include protecting the biodiversity of forest floors, enhancing small stream flows, and serving as a fire management tool with old forests).[17] Asking about nuclear waste is a different prompt than asking students whether they are aware of the comparative environmental effects of nuclear energy and energy from carbon-based sources.

The section of the questionnaire measuring responsibility for sustainable development goes even further down this road:

> *To what extent do students link societies' actions with these environmental issues and feel responsibility for these issues? To gain a sense of students' responsibility for sustainable development, students were asked whether or not they agreed with seven possible sustainable development policies. Students who responded that they either agreed or strongly agreed were classified as expressing a sense of responsibility for sustainable development.*[18]

PISA asks students if they agree or disagree with the following statements:

- *Industries should be required to prove that they safely dispose of dangerous waste materials.*
- *I am in favor of having laws that protect the habitats of endangered species.*
- *It is important to carry out regular checks on the emissions from cars as a condition of their use.*
- *To reduce waste, the use of plastic packaging should be kept to a minimum.*
- *Electricity should be produced from renewable sources as much as possible, even if this increases the cost.*
- *It disturbs me when energy is wasted through the unnecessary use of electrical appliances.*
- *I am in favor of having laws that regulate factory emissions even if this would increase the price of products.*

The four-point scale consists of strongly agree, agree, disagree, and strongly disagree. PISA assumes through its scoring rubric that students agreeing or strongly agreeing with these seven policies possess "a sense of responsibility for sustainable development." Students who do not agree or strongly disagree with these policies are considered lacking such a sense of responsibility.

One of the policy statements is innocuous—being disturbed by the waste of energy—but responses to the others are reflective of political judgment. There are arguments on both sides of public policies and responsible citizens can, armed with facts and reason, come down on one side or the other. A good citizen brings not only sentiments to bear on policy decisions, but also an understanding of a policy's economic impact, the priority the policy should be

It is difficult to see how declaring support or opposition to a policy without knowing the details is a sign of responsible citizenship. It may in fact be exactly the opposite.

granted for governmental action, whether the public or private sector should address the problem, the level of government best suited for implementing the policy, ethical questions about the policy, and other considerations.

The policy statements in PISA embrace a superficial view of responsibility. None of the prompts asks students whether they are willing to take personal responsibility for sustainability. They ask whether someone else should—industries, car owners, factories, and society as a whole. Then there is this: how would a scientifically literate student respond to these statements? Many might search for a neutral response option, believing that these policies are too complex to summarize an intelligent position on them using a four-level scale. Thoughtful students might look for a response option indicating that more information is needed before rendering a judgment.

Consider the policy statement, "Industries should be required to prove that they safely dispose of dangerous waste materials." What does it mean to be required to prove? What kind of proof? Is the standard of proof reasonable? Are penalties involved if industries do not meet such a standard? What is safe disposal? How is safe disposal defined? How are dangerous waste materials defined?

It is difficult to see how declaring support or opposition to a policy without knowing the details is a sign of responsible citizenship. It may in fact be exactly the opposite.

Summary and Conclusion

The United States participates in two international assessments, PISA and TIMSS. The two tests differ in governance, content, philosophy, and target populations. Influential groups are urging states to benchmark their tests to PISA in the hope that policymakers can learn about the practices and policies of effective educational systems. PISA measures the reading, mathematical, and science literacy of 15-year-olds, domains that extend beyond the reading, mathematics, and science that students learn in school. Educational systems are defined broadly as well. They include schools, teachers, and curricula, of course, but also parents, families, peers, communities, and popular culture.

PISA considers students' attitudes, beliefs, and values part of literacy. Questionnaires are administered to students and some attitudinal items are interspersed in the portion of the test assessing knowledge. In the analysis above, the chapter of the 2006 PISA report on science that discusses student beliefs, attitudes, and values was examined. The analysis uncovered several problems that state policymakers should consider before benchmarking their assessments to PISA.

The Problem of Governance

PISA needs nongovernmental participation built into its governance as a check on policy recommendations that merely reflect conventional wisdom or mirror the status quo of ministries. As it now stands, the OECD takes policy positions, collects PISA data, then analyzes and interprets results pertaining to many of the policy positions already taken. This is an obvious conflict of interest. In the United States, much discussion takes place about how to ensure the independence of the two data collecting bodies—the National Assessment Governing Board and the National Center for Education Statistics—from other wings of the U.S. Department of Education, in particular, from the department's policymaking units.[19] In addition to structural safeguards, an ethos prevails

State policymakers should consider [several problems] before benchmarking their assessments to PISA.

of leaving the interpretation of data to others. For the most part, the United States collects educational data, releases it with a minimum of interpretation, and makes data files available to analysts for the purpose of more finely grained interpretations that lead to policy recommendations. Does it make sense for the United States to place such constraints on our own national test-givers but give credence to policy recommendations coming from officials in Paris, France, who are not so constrained?

The Problem of Alignment

PISA is currently not designed for evaluating the effectiveness of schools. The content of PISA reflects more than is taught by school curricula and assesses more than the activities of school systems. As stated by PISA, the test measures "the cumulative impact of learning experiences…starting in early childhood and up to the age of 15 and embracing experiences both in school and at home." Moreover, PISA tests most American students at the beginning of their sophomore year. PISA makes assumptions about these students' "educational systems" based on data gathered from their schools. It is specious to make policy recommendations based on correlations of the characteristics of the students' current schools, which they have only attended one year, with the learning that students have acquired over a lifetime.

The Problem of Selective Treatment of Data

In the section of the executive summary discussing attitudes and science performance, the PISA 2006 report shows scatter plots of national performance and scores for only two variables—self-efficacy and environmental awareness. As shown above, self-efficacy is neutral and environmental awareness is positively associated with scientific performance in these correlations. Scatter plots of the thirteen variables that are not shown evidence a negative association with performance. This is a selective choice of variables that biases the discussion of attitudes and performance in a positive direction.

The executive summary states, "Within each country, students who reported that they enjoy learning science were more likely to have higher levels of science performance. While this does not show a clear causal link, it appears that students with greater interest and enjoyment of science are more willing to invest the effort to do well."[20] A busy policymaker might not notice that scatter plots for interest and enjoyment of science had not been provided. They are not in the full report either. Had these two scatter plots been included, they would have shown sharply descending slopes, reflecting negative between-nation correlations for these two traits and performance (r = -0.76 for enjoyment and -0.82 for interest). This is a selective choice of correlational analyses to include in the report that biases the discussion, again, toward concluding that a positive association exists.

The Problem of Policy Recommendations Going Beyond Data

PISA's policy recommendations go beyond the PISA data. In the review above, the data in one chapter were examined and critiqued. What should policymakers think about students' attitudes, beliefs, and values? Should science programs include activities to build positive ratings on affective characteristics?

The evidence provided by PISA is too weak to guide policy on this matter. Within-nation correlation coefficients show that students who have higher levels of the affective characteristics surveyed by PISA

This is a science test. Stick to the science.

also know more science. The data are cross-sectional and cannot untangle whether students who possess particular attitudes, beliefs, and values are more likely to learn science or whether students who have already learned a lot of science are more likely to possess such characteristics. On thirteen of the fifteen characteristics, between-nation correlations show that nations with higher levels on the affective characteristics score lower on knowledge of science. The relationship is negative.

The point is not that policymakers should dream up ways of making science boring and dreary. No one is going to argue that. The point is that the data are ambiguous on the question of whether science curricula should be concerned with boosting attitudinal indices. Nations that launch bold new programs to increase student enjoyment of science, for example, may see no benefit from their efforts. Whether changing students' attitudes, beliefs, and values will help or hinder science learning cannot be determined from PISA data.

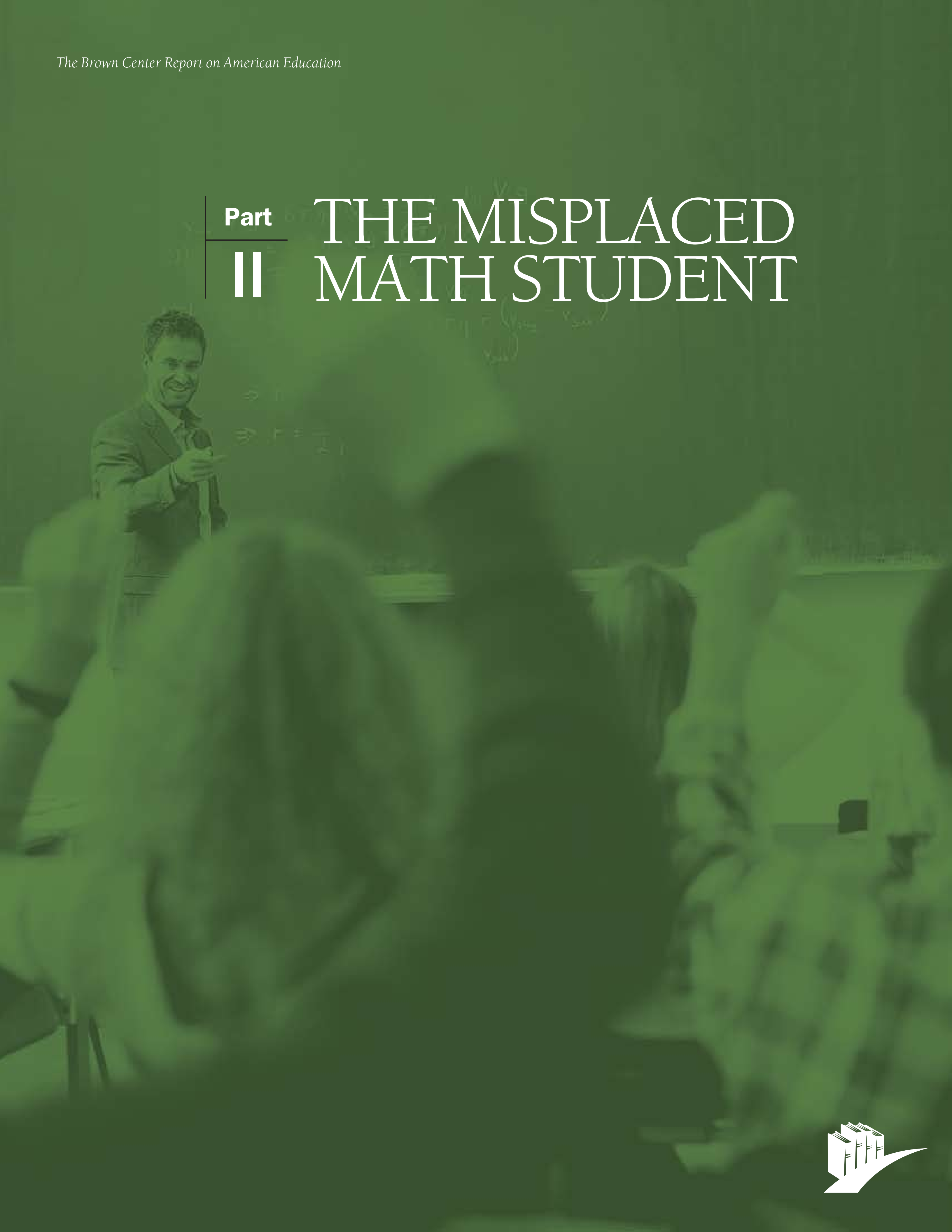
The Problem of Ideological Bias

The OECD routinely scrubs PISA items for gender and cultural bias. It is imperative that PISA institute procedures that scrub items for ideological bias as well. Under any circumstances, extreme caution must be taken when inquiring about students' political beliefs. Even more so in this case. The OECD represents governments and takes positions on political issues, including environmental policies. It has even created a curriculum for schools to use in teaching about environmental sustainability. For the OECD to ask students about their political views on sustainability in the context of an international assessment is inappropriate. The error is compounded by establishing a coding scale that interprets particular responses as evidence of student responsibility.

The solution is not for PISA to substitute another point of view for the one promulgated by the test. Nor should PISA attempt "balance" by asking several items slanted towards several different political ideologies. The solution is to avoid asking such questions altogether. This is a science test. Stick to the science. Scientific knowledge can be measured without eliciting the political pursuits to which students would put science to use. It would be informative, for instance, to measure what students know about the science of greenhouse gases or acid rain, then to investigate whether awareness of these issues is grounded in scientific knowledge or misinformation. Such an approach would offer a glimpse into whether students can apply science to the analysis of public problems.

The PISA assessment seeks to be relevant. It asks questions that put students in the position of applying science to public problems. That is a legitimate endeavor. But it also leads to an inescapable trap. If PISA asks a non-controversial question, almost all students will answer the same way and nothing will be learned. If PISA asks a controversial question, then it probably will be delving into topics that have no place on a student assessment.
A science test can avoid this trap by restricting itself to the science behind public policies. Finding out whether students know enough science to understand the many complex issues they will face as citizens is a worthy objective for the PISA assessment.

Part II
THE MISPLACED MATH STUDENT



ALGEBRA IN EIGHTH GRADE WAS ONCE RESERVED FOR THE mathematically gifted student. In 1990, very few eighth graders, about one out of six, were enrolled in an algebra course. As the decade unfolded, leaders began urging schools to increase that number. President Clinton lamented, “Around the world, middle students are learning algebra and geometry. Here at home, just a quarter of all students take algebra before high school.”[21] The administration made enrolling all children in an algebra course by eighth grade a national goal. In a handbook offering advice to middle school students on how to plan for college, U.S. Secretary of Education Richard Riley urged, “Take algebra beginning in the eighth grade and build from there.”[22] Robert Moses ratcheted up the significance of the issue by labeling algebra “The New Civil Right,” thereby highlighting the social consequences of so many poor and minority students taking remedial and general math courses instead of algebra.[23]

The campaign was incredibly successful. Several urban school districts declared a goal of algebra for all eighth graders. In 1996, the District of Columbia led the nation with 53 percent of eighth graders enrolled in algebra. From 1990 to 2000, national enrollment in algebra courses soared from 16 percent to 24 percent of all eighth graders. The surge continued into the next decade. Eighth-grade enrollment in algebra hit 31 percent nationally in 2007, a near doubling of the 1990 proportion. Today more U.S. eighth graders take algebra than any other math course.[24] In July 2008, the State of California decided to adopt an algebra test as its eighth-grade assessment of student proficiency. The policy in effect mandates that all eighth graders will be enrolled in algebra by 2011.

At first glance, this appears to be good news. Transcript studies indicate that 83 percent of students who take geometry in ninth grade, most of whom completed algebra in eighth grade, complete calculus or another advanced math course during high school.[25] Research also suggests that students who take algebra earlier rather than later subsequently have higher math skills.[26] These findings, however, are clouded by selection effects—by the presence of unmeasured factors influencing who takes

Are we enrolling eighth graders who know very little mathematics in higher-level math classes?

Statewide Enrollment in Advanced Math Classes, 2007 (with 8th-grade NAEP math score)

Table 2-1

Jurisdiction	8th-Grade NAEP Score	Total Advanced Enrollment
National	281	38%
Massachusetts	298	45%
Minnesota	292	35%
North Dakota	292	21%
Vermont	291	26%
Kansas	290	39%
New Jersey	289	40%
South Dakota	288	30%
Virginia	288	42%
New Hampshire	288	30%
Montana	287	24%
Wyoming	287	32%
Maine	286	29%
Colorado	286	44%
Pennsylvania	286	42%
Texas	286	28%
Maryland	286	52%
Wisconsin	286	30%
Iowa	285	27%
DoDEA	285	40%
Indiana	285	33%
Washington	285	31%
Ohio	285	35%
North Carolina	284	33%
Oregon	284	39%
Nebraska	284	35%
Idaho	284	37%
Delaware	283	36%
Connecticut	282	39%
South Carolina	282	41%
Utah	281	58%
Missouri	281	33%
Illinois	280	33%
New York	280	21%
Kentucky	279	34%
Florida	277	42%
Michigan	277	38%
Arizona	276	32%
Rhode Island	275	41%
Georgia	275	49%
Oklahoma	275	27%
Tennessee	274	31%
Arkansas	274	33%
Louisiana	272	24%
Nevada	271	34%
California	270	59%
West Virginia	270	33%
Hawaii	269	28%
New Mexico	268	34%
Alabama	266	30%
Mississippi	265	21%
District of Columbia	248	51%

Source: Author's calculations from 8th-grade math state main NAEP, NAEP data explorer http://nces.ed.gov/nationsreportcard/nde/

algebra early and who takes it late. Schools routinely assign incoming eighth graders to math courses based on how much math students already know. Moreover, it is no surprise that excellent math students want to take the most challenging math courses available to them and that low-achieving students avoid these courses as long as possible. Whether algebra for eighth graders is a good idea, especially for those who have not learned basic arithmetic, cannot be concluded from existing evidence. Studies that test for causality, such as experiments with random assignment of students to treatment and control groups, have not been conducted.

The push for universal eighth-grade algebra is based on an argument for equity, not on empirical evidence. General or remedial math courses tend to be curricular dead-ends, leading to more courses with the same title (for example, General Math 9, General Math 10) and no real progression in mathematical content. By completing algebra in eighth grade—and then completing a sequence of geometry as freshmen, advanced algebra as sophomores, and trigonometry, math analysis, or pre-calculus as juniors—students are able to take calculus in the senior year of high school. Waiting until ninth grade to take algebra makes taking calculus in high school more difficult. From this point of view, expanding eighth-grade algebra to include all students opens up opportunities for advancement to students who previously had not been afforded them, in particular, students of color and from poor families. Democratizing eighth-grade algebra promotes social justice.

Two Curious Patterns in NAEP Data

One catch. Course-taking is a means to an end, not an end in itself. Students take math courses to learn mathematics. Will policies mandating algebra for all eighth graders mean that the nation's students learn more math? Not necessarily. Although cross-sectional state test data cannot answer such a question, they can answer a different question: do states that enroll more students in advanced math courses score higher than states enrolling fewer students in advanced courses?

Table 2-1 shows the 2007 eighth-grade National Assessment of Educational Progress (NAEP) scores for states and jurisdictions and the percentage of eighth graders enrolled in advanced math classes (Algebra I, Geometry, and Algebra II). Massachusetts scores at the top (298) and has 45 percent of eighth graders enrolled in advanced math, more than the national average of 38 percent. But several high-scoring states enroll fewer students in advanced classes. North Dakota and Vermont, for example, are ranked third and fourth in math achievement but enroll a relatively low percentage of eighth graders in advanced math (21 percent and 26 percent, respectively). On the other end of the spectrum, the District of Columbia scores last on NAEP but continues to be one of the leaders in the percentage of students taking advanced math.[27] The Pearson correlation coefficient, a measure of the statistical relationship between two variables, for NAEP score and advanced math enrollment is -0.09, indicating no correlation.

Another intriguing pattern in eighth-grade NAEP scores emerges from examining the scores of eighth graders taking advanced math courses. The national average in eighth-grade math has been rising steadily, increasing by 8 points from 2000 to 2007, from 273 to 281 (see Figure 2-1). But one group stands out for not participating in the score increase—eighth graders in advanced classes. Their NAEP scores have declined from 299 in 2000 to 295 in 2007, a loss of

Universal eighth-grade algebra is based on an argument for equity, not on empirical evidence.

4 scale score points. The typical eighth grader knows more math today than in 2000. But the typical eighth grader in an advanced math course knows less. How can that happen?

As a cross-sectional measure of student achievement, NAEP provides snapshots of math achievement at one point in time. The data cannot prove or disprove causality. But NAEP data do provide rich descriptive information on what is going on in schools. Access to eighth-grade algebra has expanded dramatically. Almost nothing is known about the students who are taking these courses. Are we enrolling eighth graders who know very little mathematics in higher-level math classes?

Methods

We tackled this question by examining data on students in advanced math courses, their schools, and their teachers. The data analyzed below are from the 2005 NAEP restricted-use files, providing student-level information on a nationally representative sample of 160,000 eighth graders. Unlike the data used in most NAEP studies, these files require licensing for use and allow investigators to drill down to individual student characteristics. The 2005 data are the most recent available for this type of analysis. Advanced math courses are typically the courses that good math students take in the transition from middle to high school mathematics—in previous eras, during the first few years of high school. "Basic" refers to courses taken before students enroll in formal algebra, including pre-algebra, naturally, but also general math.

One important limitation to the data. Course-taking on the eighth-grade NAEP is reported by students. They are asked to check off the math course in which they are currently enrolled. Many students may not know the actual title of their math course, may exaggerate the level of the course, or may for some other reason not

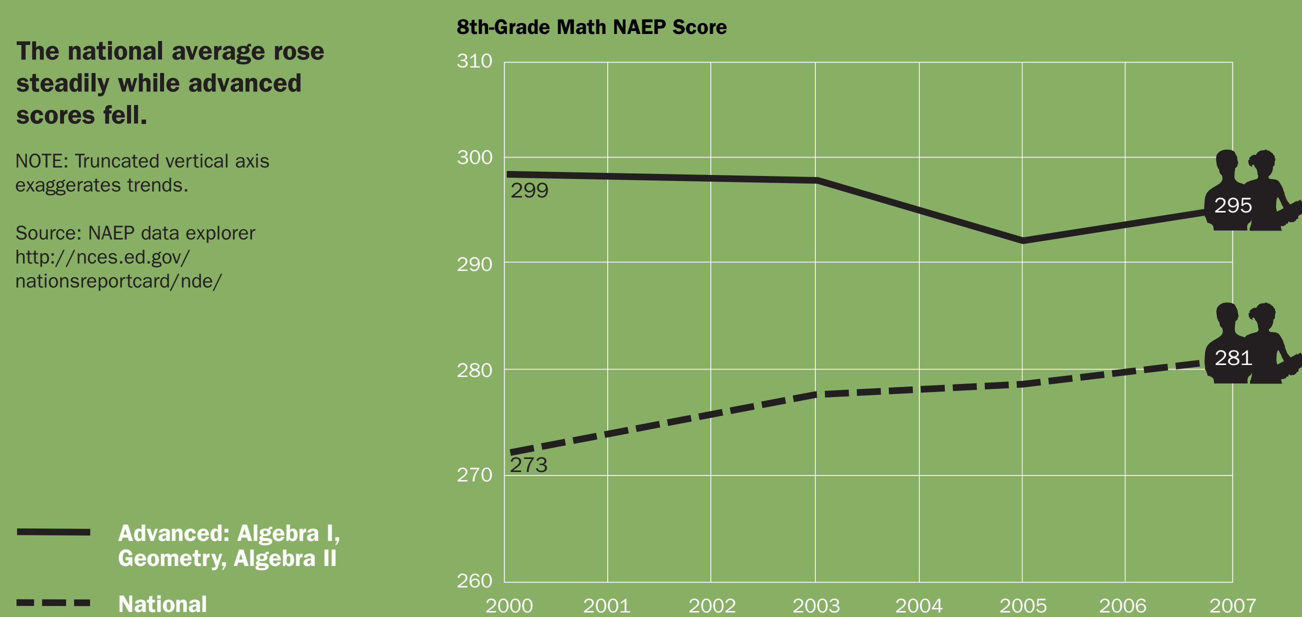
Eighth-grade NAEP scores: National average for students in advanced math (2000–2007)

Fig 2-1

The national average rose steadily while advanced scores fell.

NOTE: Truncated vertical axis exaggerates trends.

Source: NAEP data explorer http://nces.ed.gov/nationsreportcard/nde/



report the course accurately. Transcripts would provide more accurate information. Although they have been collected on high school students for other national surveys, transcripts generally are not available for eighth graders. Because of this, the NAEP data are the most authoritative in existence for tracking national trends in course-taking in eighth grade. Whatever flaws arise from student self-reports, there is no reason to believe that the reliability of the reports has changed significantly over time, allowing for reasonably accurate estimates of changes in course enrollments.

What Math Courses Are Eighth Graders Taking?

As shown in Figure 2-2, between 2000 and 2005 eighth graders shifted toward tougher courses. The percentage of students taking advanced courses shot up while basic math courses experienced enrollment declines. Enrollment in advanced courses increased by about 10 percentage points, from 26.7 percent to 36.6 percent, and in basic courses fell by about 16 percentage points, from 66.6 percent to 50.8 percent. It appears that many students who would have taken lower-level math courses were taking algebra, geometry, or advanced algebra in 2005. The campaign for algebra by eighth grade clearly succeeded in boosting the number of American youngsters enrolled in tougher mathematics courses.

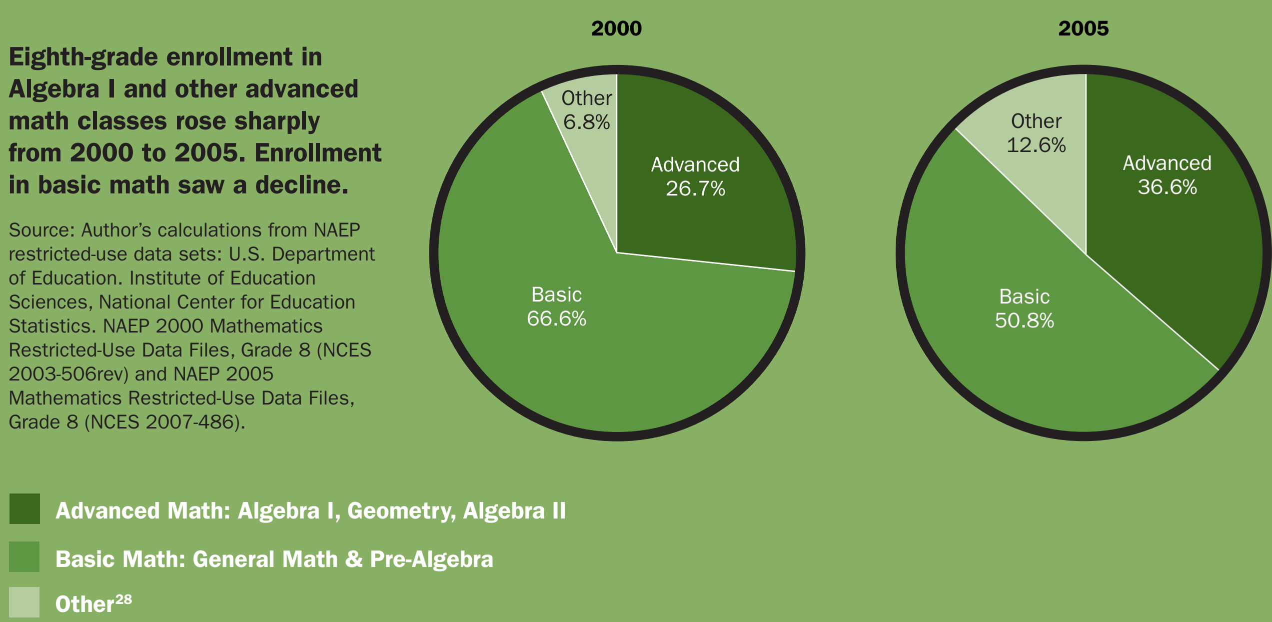
Are all of these new students in advanced courses actually good at math? Unfortunately, the answer is no. In fact, many are very poor math students, at least as measured by their performance on the NAEP math test. Let's consider students at the 10th percentile and below—the bottom 10 percent of students nationally on the NAEP test—as low-achieving or struggling math students. How did their course-taking change from 2000 to 2005? In 2000, only 8.0 percent of

Course-taking in eighth-grade math, 2000 and 2005

Fig **2-2**

Eighth-grade enrollment in Algebra I and other advanced math classes rose sharply from 2000 to 2005. Enrollment in basic math saw a decline.

Source: Author's calculations from NAEP restricted-use data sets: U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. NAEP 2000 Mathematics Restricted-Use Data Files, Grade 8 (NCES 2003-506rev) and NAEP 2005 Mathematics Restricted-Use Data Files, Grade 8 (NCES 2007-486).



low-achieving students enrolled in advanced math courses (see Table 2-2). Almost nine times as many, 73.7 percent, took general math or pre-algebra. In 2005, the percentage of low achievers enrolled in advanced math classes had ballooned to 28.6 percent. The percentage enrolled in basic courses fell to 46.3 percent. The ratio had fallen to less than two to one.

How Has the Composition of Advanced Classes Changed?

High achievers—students scoring at the 90th percentile or above—made up 27.0 percent of the advanced classes in 2000. In 2005, the percentage dropped to 20.0 percent. Low achievers more than doubled as a proportion of advanced classes, increasing from 3.0 percent in 2000 to 7.8 percent in 2005. Although appearing to be trivial, this small percentage adds up to approximately 120,000 students nationwide, a number that

Table 2-2

Math courses taken by low achievers (10th percentile and below students), 2000 and 2005
Percentage of low achievers enrolled in various math classes.

		2000		2005	
Advanced	**Algebra I**	4.8		17.4	
	Geometry	2.1	**8.0**	5.0	**28.6**
	Algebra II	1.1		6.2	
Basic	**General math**	50.7	**73.7**	27.1	**46.3**
	Pre-algebra	23.0		19.2	
Other		18.3	**18.3**	25.0	**25.0**

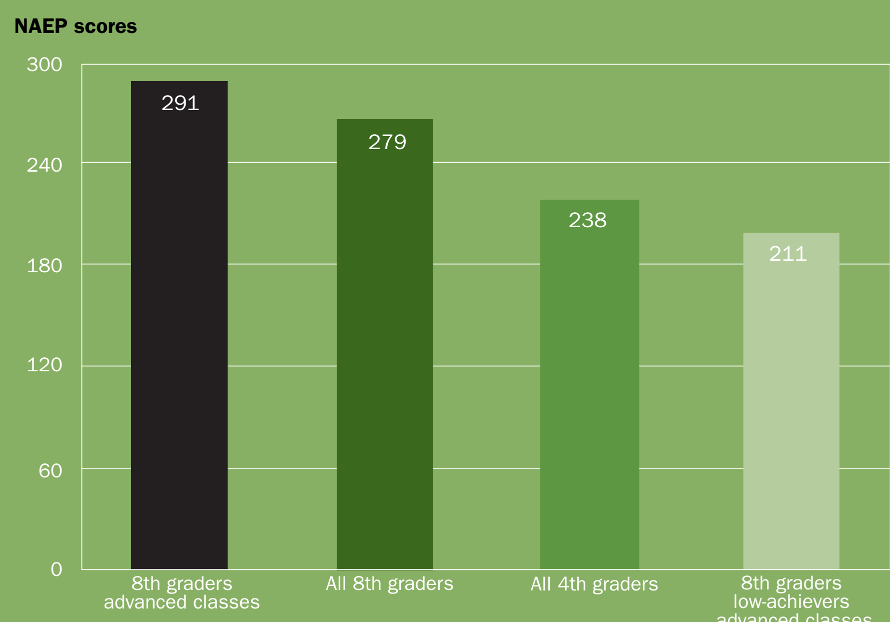
Source: Author's calculations from NAEP restricted-use data sets: U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. NAEP 2000 Mathematics Restricted-Use Data Files, Grade 8 (NCES 2003-506rev) and NAEP 2005 Mathematics Restricted-Use Data Files, Grade 8 (NCES 2007-486).

Fig 2-3

NAEP scores of different student groups, 2005

Low-performing eighth graders in advanced classes score even below the average fourth-grade student.

Source: Author's calculations from NAEP restricted-use data set: U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. NAEP 2005 Mathematics Restricted-Use Data Files, Grade 8 (NCES 2007-486) and NAEP data explorer http://nces.ed.gov/nationsreportcard/nde/



Sample NAEP item (working with percentages)
Grade 8 Item Block 2005-8M3 No. 17:

Table **2-3**

There were 90 employees in a company last year. This year the number of employees increased by 10 percent. How many employees are in the company this year?

A) 9
B) 81
C) 91
D) 99 ✔
E) 100

	Overall	Advanced Classes	Misplaced 10th
Percent answering correctly	**36.5**	**48.7**	**9.8**

Source: NAEP question tool http://nces.ed.gov/nationsreportcard/itmrls/startsearch.asp and author's calculations from NAEP restricted-use data set: U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. NAEP 2005 Mathematics Restricted-Use Data Files, Grade 8 (NCES 2007-486).

Sample NAEP item (rounding decimals)
Grade 8 Item Block 2005-8M4 No. 9:

Table **2-4**

Alba needed to know about how much the sum of 19.6, 23.8, and 38.4 is. She correctly rounded each of these numbers to the nearest whole number. What three numbers did she use?

A) 19, 23, 38
B) 19, 24, 38
C) 20, 24, 38 ✔
D) 20, 24, 39

	Overall	Advanced Classes	Misplaced 10th
Percent answering correctly	**85.2**	**87.9**	**37.1**

Source: NAEP question tool http://nces.ed.gov/nationsreportcard/itmrls/startsearch.asp and author's calculations from NAEP restricted-use data set: U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. NAEP 2005 Mathematics Restricted-Use Data Files, Grade 8 (NCES 2007-486).

is growing and a phenomenon that, until now, has been viewed as an accomplishment, not a cause for worry.[29]

The scope of this development is also significant when viewed from the perspective of a classroom teacher. About one out of every thirteen eighth graders in an advanced math class knows very little mathematics. An algebra teacher with a class of twenty-six kids can expect to have two students performing several years below grade level. The vast majority of students taking the class are functioning above grade level, but the number of struggling math students in advanced classes is increasing at the same time the proportion of high-achieving students in those classes is declining.

How Far Behind Are the Misplaced Students?

The average NAEP score for eighth graders in advanced math classes is 291 (see Figure 2-3). The national average for all eighth graders is 279. On the same NAEP scale, the national average for fourth graders is 238. The misplaced eighth graders score an average of 211, which is 27 scale score points below the national average for fourth grade. Analysts consider 11 NAEP scale score points as approximately equivalent to one year of learning, which means that these misplaced students know about as much math as a typical second grader. Advanced students score about one year above grade level. The misplaced students function about seven grade levels below peers enrolled in the same courses.

Examining a few sample NAEP items illustrates the misplaced students' gaps in knowledge. The first item is quite challenging for eighth graders (see Table 2-3). It asks students to calculate the result of a particular percentage increase, an arithmetic skill that, as shown in previous Brown Center Reports, eludes most eighth graders. Indeed, in 2005 only 36.5 percent of eighth graders answered the question correctly. Students in advanced courses did somewhat better, with 48.7 percent arriving at the correct solution. The misplaced students in advanced courses performed abysmally, with only 9.8 percent getting this item right.

The second item is much easier (see Table 2-4). Students are asked to round decimals to the nearest whole number. Rounding requires number sense, especially in terms of understanding the relative value of numbers on a number line. Most eighth graders have no trouble with this item—85.2 percent of all eighth graders got it right in 2005, 87.9 percent of the students enrolled in advanced classes. But only 37.1 percent of misplaced students could answer the item correctly. Failing to round simple decimals accurately to the nearest whole number signals a serious lack of understanding of the number system. Taking the same math courses as peers who easily grasp such concepts makes such deficiencies even more glaring.

Those two NAEP items are in public release, meaning that they are no longer used on NAEP tests and can be made public. We gathered data on three additional items involving fractions. Math educators consider knowledge of fractions essential to preparation for algebra.[30] Although the three items are not in public release and cannot be disclosed, performance on them can be reported.

Table 2-5 compares the performance of the misplaced 10th percentile students who are enrolled in advanced courses with students in advanced courses and all eighth graders in the nation. On the easiest of the three items, item A, the misplaced students scored far below their peers in advanced classes. Less than half get an item right that their classmates find relatively easy. On the more difficult items, items B and C, fewer than one in ten misplaced students answer these items correctly. They score even lower than the 20 percent rate attained by simply guessing on a multiple choice item with five possible answers.

Fractions are taught in elementary school, not in an algebra course. Sadly, facility with fractions is a skill that the misplaced students do not know, need to know, and are unlikely to be taught in the math course in which they are enrolled.

Characteristics of the Misplaced Students

Who are these 120,000 misplaced students? We examined information contained in the NAEP surveys on the students' families, schools, and teachers. What we found is troubling. These students tend to be some of the nation's most vulnerable children. We already know that they struggle at mathematics, scoring among the bottom 10 percent of all eighth graders in the country. They also possess characteristics that make recovery from a lost year of math instruction unlikely.

Tables 2-6 and 2-7 describe the misplaced students and compare them to students in advanced math classes and the typical American eighth grader. All of the differences highlighted in the following discussion, unless otherwise noted, are statistically significant ($p < .05$).

Table 2-5

Performance on sample NAEP items involving fractions (percentage answering correctly)

	Overall	Advanced Classes	Misplaced 10th
Item A	72.6	78.4	42.3
Item B	45.1	57.2	3.9
Item C	47.2	58.4	6.6

Source: Author's calculations from NAEP restricted-use data set: U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. NAEP 2005 Mathematics Restricted-Use Data Files, Grade 8 (NCES 2007-486).

What Are the Background Characteristics of the Misplaced Students?

Table 2-6 displays demographic data. Misplaced students are more likely to come from poor families—69.8 percent qualify for the federal free or reduced-price lunch program, a proxy for family income. This is more than double the percentage for students in advanced classes (30.4 percent) and nearly twice that of the national average (36.1 percent). Misplaced students are overwhelmingly black and Hispanic, about 77.0 percent versus 32.3 percent of all eighth graders in the nation. Only 20.3 percent

These students tend to be some of the nation's most vulnerable children.

Demographic characteristics: misplaced students and comparison groups, 2005
Percentage of students by characteristic

Table **2-6**

	Misplaced 10th	Advanced Classes	National Average
Eligible free lunch	69.8	30.4	36.1
White	18.5	60.9	61.1
Black	38.4	14.2	16.1
Hispanic	38.6	17.1	16.2
Mother college grad	20.3	44.8	36.9

Source: Author's calculations from NAEP restricted-use data set: U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. NAEP 2005 Mathematics Restricted-Use Data Files, Grade 8 (NCES 2007-486).

School characteristics: misplaced students and comparison groups, 2005

Table **2-7**

	Misplaced 10th	Advanced Classes	National Average
Urban	50.9%	33.4%	31.3%
Suburban	35.4%	46.4%	43.1%
Rural	13.7%	20.2%	25.6%
School enrollment	1012	844	794
Private school	2.3%	10.5%	8.8%
>50% eligible lunch	67.6%	30.4%	31.6%
8th-grade math untracked	34.8%	22.8%	26.9%

Source: Author's calculations from NAEP restricted-use data set: U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. NAEP 2005 Mathematics Restricted-Use Data Files, Grade 8 (NCES 2007-486).

report that their mothers graduated from college. The argument that advanced math courses are a civil right apparently has had an impact on schools, boosting the enrollment of black, Hispanic, and poor children in advanced courses. Unfortunately, the children in the current study are unprepared for algebra. And they come from homes in which, probably lacking the resources to afford tutors and other remedial materials, support may be tenuous when academic troubles occur.

What Kind of Schools Do the Misplaced Students Attend?

Table 2-7 shows the characteristics of these students' schools. About half of the misplaced students attend urban schools (50.9 percent), and they are less likely to attend suburban or rural schools than the average eighth grader. Their schools tend to be large, enrolling about 27 percent more students than the typical school housing an eighth grade (1,012 students versus 794). Almost all of the misplaced students are attending public schools, with only 2.3 percent going to private schools. The schools serve vast numbers of students in poverty. Two-thirds of the schools (67.6 percent) are high-poverty schools, defined as schools in which more than half of the students qualify for free or reduced-price lunch. Only about one-third of schools in the country fit this definition.

Schools attended by misplaced students also are more likely to shun the assignment of students to eighth-grade math classes based on mathematics ability (also known as tracking). The advanced math classes attended by misplaced students attempt to serve a wider range of mathematics abilities than the typical eighth-grade advanced math class, with 34.8 percent of schools reporting that math is untracked compared to 22.8 percent.

In sum, the profile sketched here—academically diverse classes in large, urban public schools attended predominantly by students from poverty—resembles the kind of setting that, being under great stress, many federal and state programs attempt to assist with extra financial aid. Unfortunately, it is also the kind of setting where students who are enrolled in the wrong course may fall through the cracks and flounder academically.

What Are the Teachers of Misplaced Students Like?

What do we know about the teachers of misplaced students? Teacher characteristics are displayed in Table 2-8. Compared to teachers of the typical eighth grader, the teachers of misplaced students are more likely to have taught for less than five years (30.3 percent versus 22.5 percent), less likely to hold a regular or advanced teaching certificate (74.7 percent versus 82.5 percent) and less likely to have majored in mathematics as an undergraduate (20.1 percent versus 26.2 percent). Granted, these factors are only crude indicators of teacher quality, but they are recognized by many experts as important. Less experience, fewer formal credentials, and weaker mathematics training are characteristics associated with lower-, not higher-quality teaching staffs. These unprepared students are arriving in algebra classes that are staffed by underprepared teachers.

In less than two decades, policies designed to push eighth graders into algebra classes have succeeded in doubling the percentage of students enrolled in advanced mathematics. The data assembled here document a stark consequence of such policies: large numbers of students taking courses for which they are unprepared in settings that are not particularly conducive to learning.

Teacher characteristics: misplaced students and comparison groups, 2005
Percentage of students by characteristic

Table **2-8**

	Misplaced 10th	Advanced Classes	National Average
Less than 5 years experience	30.3	21.3	22.5
Regular or advanced teaching certificate	74.7	83.7	82.5
Undergraduate major: mathematics	20.1	28.6	26.2

Source: Author's calculations from NAEP restricted-use data set: U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. NAEP 2005 Mathematics Restricted-Use Data Files, Grade 8 (NCES 2007-486).

Discussion and Policy Recommendations

One hundred twenty thousand eighth graders are sitting in advanced math classes even though they score in the bottom 10 percent of students nationwide on the NAEP math test. They know about as much math as the typical second grader. They do not know basic arithmetic and cannot correctly answer NAEP items using fractions, decimals, or percents. These students are disproportionately black and Hispanic. They hail from poor households with parents whose own education is below the national average. The schools that these children attend are large, urban public schools with predominantly low socioeconomic status populations. Their algebra classes are populated by students with mathematical abilities spanning several years. Their math teachers are less experienced, less credentialed, and less well prepared in mathematics training than the typical teacher of advanced math students in eighth grade.

No element of this story is educationally sound. It arose from good intentions: to democratize advanced math courses by assigning students to Algebra I, Geometry, and Algebra II who were once locked out of such courses. But this is false democratization. No social benefit is produced by placing students in classes for which they are unprepared. Indeed, it is difficult to imagine any educational benefit accruing to these students. They do not possess the family or school resources to overcome problems arising from taking inappropriate courses.

Let us not forget the hundreds of thousands of well-prepared students—who

No social benefit is produced by placing students in classes for which they are unprepared.

are also predominantly black, Hispanic, or poor—sitting in the same classrooms as the misplaced students and equally deserving of a good education. Well-prepared students need a real algebra class, not a fake one teaching elementary school mathematics. Any teacher who stops to teach misplaced students fractions shortchanges the well-prepared students who sit in that algebra class. William Sanders, whose studies on effective teachers in Tennessee are widely cited in the literature, declared high-achieving minority students the children "whom the system serves worst." In particular, Sanders attributed the decline in test scores of high-achieving black students to "their higher likelihood of being in a succession of classrooms where the instruction is geared to lower achievers."[31]

The chances of algebra classrooms existing with instruction geared to low achievers is probably much greater than the current study documents. In some schools, courses purported to be covering algebra have been revealed to be fraudulent—covering a watered down form of basic math. The entire class of students may be misplaced and receiving counterfeit algebra to make up for it.[32] Moreover, using the 10th percentile as the upper boundary of defining the misplaced student yields a conservative estimate of the real dimensions of the phenomenon. Are students at the 20th, 30th, or 40th percentiles on NAEP adequately prepared for algebra? They, too, function significantly below grade level in mathematics, and by including them in the pool of misplaced math students, the numbers skyrocket.

There will be advocates, despite the data presented here, who will continue to argue for placing low-performing eighth graders in algebra classes. They believe that a more rigorous course is always preferable to a less rigorous one. Many do not believe that students must learn basic mathematics in order to successfully tackle higher-level mathematics.[33] They will argue that keeping remedial math students out of algebra in eighth grade denies these students the opportunities that good math students take for granted. What they will not say is this: the burden of realizing such an idealistic view of mathematics learning falls on the classroom teacher. Teachers are expected to make up for students' skill deficiencies. If students enter algebra classes without the preparation to succeed, then algebra teachers must find a way to fix the problem.

Algebra teachers already feel the strain of such unrealistic expectations. The National Opinion Research Center (NORC) surveyed a nationally representative sample of Algebra I teachers in 2007. The teachers described their students' preparation for algebra as weak, especially in working with rational numbers and word problems. The teachers named poor work habits as a prominent barrier to learning. When asked how they would change the emphasis on mathematical topics in the elementary grades to improve preparation, the teachers' most common answer was to focus more on the mastery of basic mathematical concepts and skills. More than half felt that mixed-ability classes were a moderate (28 percent) or serious (23 percent) problem. When given ten response options to describe the "single most challenging aspect of teaching Algebra I students successfully," the most frequent response—by a landslide, chosen by 58 percent—was "working with unmotivated students." The second most frequent response, selected by 14 percent of middle school teachers and 9 percent of high school teachers, was "making mathematics accessible and comprehensible to all my students."[34]

A simple calculation illustrates the predicament. Recall that the misplaced students described above are eighth graders who function at approximately the second-grade level in math. In other words, after eight years of schooling they have learned about one-third of what the average student has learned. In eighth grade they are now expected to learn, in a single year, the six years of math that they have not yet learned along with a full year of algebra. No one—no teacher, no researcher, no governor, no school board member, no philanthropist—knows how to teach in one year what has not been learned in six and then how to teach algebra on top of that. Algebra teachers are being asked to do the impossible. The greatest teachers in the world do not know how to teach algebra to students who do not know basic arithmetic.

Elements of a Realistic Algebra Policy.

1. Get the goal right. Focus on learning, not completing a course. California is a good example. At least it puts the emphasis in the right place, by mandating a test of algebra. But why eighth grade? The mathematics on the current California High School Exit Exam is pitched below the level of the test proposed for eighth grade. Needless to say, requiring more out of eighth graders than twelfth graders is bizarre. Require that students pass a comprehensive test of algebra before graduating from high school, a requirement that about half of current American high school graduates (and more in California) would not fulfill. As economists Richard Murnane and Frank Levy have documented, research exists showing that knowledge of algebra is now essential for entry into occupations earning middle class wages. No evidence exists that it matters whether algebra is learned in eighth grade or later, and some students may need more than a year to learn the subject.[35]

2. Teach and assess prerequisite skills. The recent report of the National Mathematics Advisory Panel identified facility with whole numbers and fractions as key to preparation for algebra. Proficiency on these fundamental mathematical topics needs to be acquired before entry to algebra. Indeed, in a 2008 study of students in San Diego, Zau and Betts found that fourth-grade math scores were as good at predicting success on the California high school exit exam as ninth-grade scores.[36] This finding suggests that elementary mathematics is essential and failure to learn it has long-term consequences.

3. Early intervention. Preparing students for algebra is the culmination of many, many years of teaching and learning and the product of hard work by students, teachers, and families. Mandating algebra in eighth grade is the equivalent of mandating, by policy, that all buildings immediately erect a fiftieth floor—regardless of their current height. Use diagnostic assessments of whole number and fraction arithmetic in the elementary grades to identify students who are struggling at math. Build student accountability into the system by requiring summer school for students who need more time to learn the building blocks of mathematics.

4. Collect data, conduct research. Many advocates of algebra for all eighth graders express the belief that lofty public policy goals can be attained through sheer will power, a "mandate it and it will be accomplished" ideal. Governor Schwarzenegger, for example, in a letter to the California State Board of Education, compared mandatory algebra in eighth grade to President Kennedy's pledge that

Americans would reach the moon. The analogy is specious. True, in 1960 the man on the moon goal was ambitious, but a body of science indicated that going to the moon was possible. President Kennedy did not say we would put a man on Pluto. Not even Venus. He said the moon because the principles of physics, decades of experiments with rocketry, and the early successes of Russia and the United States in space proved that it could be done. No such science supports algebra for all eighth graders.[37]

Algebra for eighth graders is an ideal policy for randomized experiments. The mandate could be introduced in some schools and districts but not others and student outcomes compared. Just as charter schools use lotteries to decide who can attend when the number of applicants exceed available seats, lotteries could be employed to assign students to eighth-grade algebra classes. By controlling for unobservable characteristics that influence math learning, studies with random assignment can offer a reasonable estimate of the true effects of a particular class on student outcomes. Summer boot camps that attempt different strategies for remediation could be started and carefully evaluated, again with randomized studies, and the effective programs then should be funded for dissemination.

Conclusion

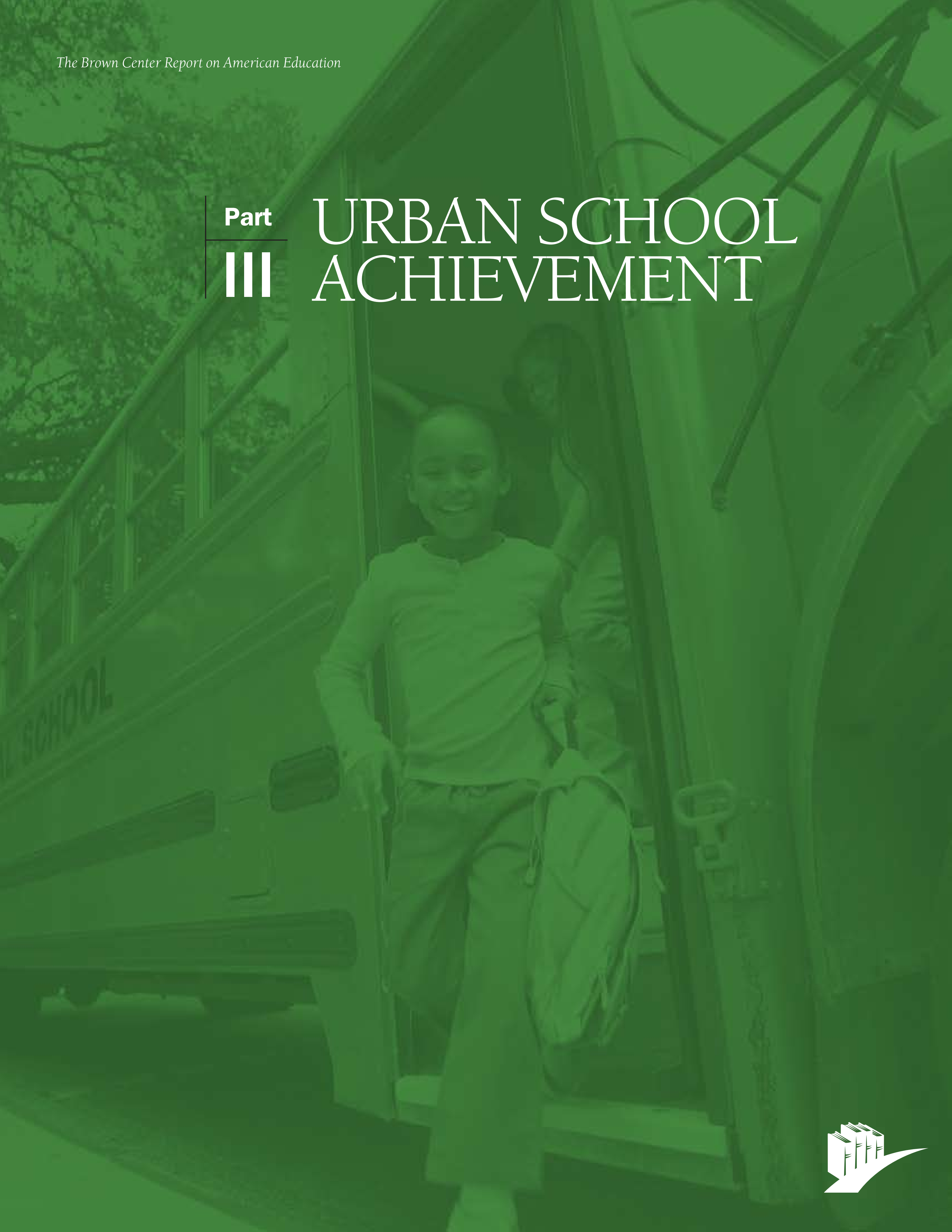
One hundred twenty thousand students are misplaced in their eighth-grade math classes. They have not been prepared to learn the mathematics that they are expected to learn. This unfortunate situation arose from good intentions and the worthy objective of raising expectations for all American students. Two groups of students pay a price. The misplaced eighth graders waste a year of mathematics, lost in a curriculum of advanced math when they have not yet learned elementary arithmetic. They should be taught whole number and fraction arithmetic so that they can then move on to successfully learn advanced mathematics.

Their classmates also lose—students who are good at math and ready for algebra. These well-prepared but ill-served students also tend to be black and Hispanic and to come from low socioeconomic backgrounds. Teachers report that classes of students with widely diverse mathematics preparation impede effective teaching, that too many students arrive in algebra classes unmotivated to learn, and that they wish that elementary schools gave greater emphasis to basic skills and concepts in math. When algebra teachers have to depart from the curriculum to teach arithmetic, the students who already know arithmetic and are ready for algebra are the losers.

This study is not a call to lower expectations. Nor is it a call for cynicism. But we must establish the right goals and pursue sound strategies for achieving them. The goal must not be for students to take an algebra course by eighth grade; it must be for more students to learn algebra. The strategy must not be to designate an arbitrary grade—unsupported by research or policy experience—in which all students are swept into an algebra course. Universal eighth-grade algebra is creating more problems than it solves, with 120,000 students not learning the mathematics that they need to know and hundreds of thousands of their classmates paying an educational price along with them.

True, in 1960 the man on the moon goal was ambitious, but a body of science indicated that going to the moon was possible.

Part III URBAN SCHOOL ACHIEVEMENT



RECENT REPORTS INDICATE THAT STUDENT ACHIEVEMENT IS rising in big city schools. The Trial Urban District Assessment (TUDA), which tracks achievement in about a dozen districts as part of the National Assessment of Educational Progress (NAEP), reported significantly higher scores in 2007 than in 2003.

Equity is at the forefront. Is it fair that children growing up in large cities are likely to receive an inferior education simply because of where they live?

In 2008, the Council of the Great City Schools analyzed trends in state tests that are used for accountability purposes. The study examined achievement in fifty urban school districts from 2003 to 2007, and like NAEP, documented large gains in mathematics and smaller, but still statistically significant, gains in reading.[38]

Is the reported progress a significant accomplishment? Urban schools are improving on NAEP and state tests, and that is cause for celebration. But perhaps schools in other neighborhoods are improving at even a greater rate—and urban children falling farther behind. Urban schools are a cause for concern not only because they score so poorly on achievement tests but also because they lag far behind their suburban and rural counterparts. With this in mind, we approach the question of urban school achievement from a different angle than the studies just mentioned. Equity is at the forefront. Is it fair that children growing up in large cities are likely to receive an inferior education simply because of where they live? A trend in which American schools markedly improve but big city schools improve by a smaller amount would not resolve the equity problem. The question takes on added significance because of the disproportionate number of African American and Hispanic youngsters attending urban schools.

Methods

The 2001 Brown Center Report analyzed how well students in the nation's largest cities performed on state tests given in 2000. Here we replicate that analysis using 2007 achievement data. We present test scores from the largest school district serving each of the nation's top fifty cities.[39] We do not measure performance on a fixed scale to determine whether these districts' achievement is high or low. Instead, we measure the performance of city school districts in relation to other districts in the same state and report the progress that city schools have made in catching up to their peer districts—or the lack of progress if they have fallen further behind. In the current environment of rising national test scores, this

City schools are indeed improving, and they are improving more than other school districts in their own states.

approach applies a more stringent standard than simply determining whether schools are improving.[40] Sure, city schools' scores are going up, but the scores of suburban and rural schools are going up as well. Are big city schools merely staying with the pack that runs ahead of them or are they making greater strides and narrowing the gap?

The approach has limitations. Students in different states take different tests, and the tests vary in quality and in what they assess. Scores are expressed on different scales. To make scores comparable, we computed a z-score for each city, an indicator of the distance—expressed in standard deviation units—between the city district's test score and its state's average score. State averages are fixed at 0.00. Positive scores are above average and negative scores are below average. This kind of relative measure means all test scores could be falling in a particular state and city schools would look good by going down less. We combined data on fourth-grade reading and eighth-grade math into a single composite score for each school district. These are two crucial grades, respectively, for reading and math achievement. Although unlikely, perhaps academic achievement in other grades or subjects has behaved differently.[41]

Analysis

Table 3-1 presents test scores for thirty-seven cities, along with key demographic data that are known to be correlated with achievement—percentage of students eligible for free and reduced lunch and the percentage of black and Hispanic students. The cities are ranked by change in z-scores from 2000 to 2007.

The results are good news. City schools are indeed improving, and they are improving more than other school districts in their own states. In 2007, ten big city districts scored at or above their state average—showing positive z-scores. That is a gain since 2000, when eight cities were at or above state averages. City districts still lag behind, but twenty-nine of the big city school districts narrowed the gap between their test scores and state averages. Eight did not.

Overall, the thirty-seven cities scored -1.26 in 2000 and -0.77 in 2007, registering a gain of 0.49 z-scores—or a little less than one-half a standard deviation on state tests. This is a significant and noticeable amount of improvement. All z-scores can be converted into a percentile score. Expressed that way, the cities' gains are equal to improving from about the 10th percentile to the 21st percentile. So the typical big city school district was outscored by 90 percent of districts in its home state in 2000 and by 79 percent of districts in 2007.

The good news must be kept in perspective. The 21st percentile is still significantly below average. Most big city school districts still trail far behind their suburban and rural peers. Five cities post z-scores of -2.00 or less—two standard deviations below state averages. They are Milwaukee, Indianapolis, Detroit, Philadelphia, and Baltimore. These cities' schools score as far below their state averages as an individual student at the 3rd percentile scores below the average student. A few of these cities are not getting any better. Detroit's scores have fallen since 2000, and Indianapolis and Baltimore are treading water.

Two California cities, San Jose and San Diego, suffered test score declines in a state in which urban achievement is generally rising. It should be noted that urban poverty (as measured by students qualifying for free lunch) and black and Hispanic percentage of enrollment grew only slightly during the 2000–2007 period, although some big city districts experienced larger demographic

Achievement in big city school districts
(2000–2007)

Table **3-1**

City	State	School District	Change in Z Score	2007 Average Z Score	2007 Free Lunch	2007 Black + Hispanic	2000 Average Z Score	2000 Free Lunch	2000 Black + Hispanic
New Orleans	LA	New Orleans Public Schools	3.03	1.05	0.29	0.93	-1.98	0.70	0.92
Dallas	TX	Dallas Independent School District	1.89	-1.29	0.83	0.93	-3.18	0.65	0.86
Minneapolis	MN	Minneapolis Public Schools	1.59	-1.81	0.67	0.57	-3.40	0.54	0.45
Austin	TX	Austin Independent School District	1.42	-0.10	0.46	0.69	-1.52	0.42	0.59
Long Beach	CA	Long Beach Unified Public School District	1.42	0.54	0.69	0.68	-0.88	0.64	0.58
Miami	FL	Miami-Dade County Public Schools	1.40	-0.37	0.61	0.89	-1.77	0.53	0.84
San Antonio	TX	San Antonio Independent School District	1.27	-0.96	0.19	0.97	-2.23	0.80	0.94
New York City	NY	New York City Public Schools	1.09	-1.31	0.78	0.76	-2.40	0.58	0.74
Philadelphia	PA	School District of Philadelphia	1.06	-2.32	0.71	0.81	-3.38	0.42	0.75
Chicago	IL	Chicago Public Schools	0.98	-1.31	0.74	0.86	-2.29	0.56	0.86
Charlotte	NC	Charlotte-Mecklenburg Schools	0.96	0.27	0.46	0.58	-0.69	0.29	0.43
Milwaukee	WI	Milwaukee Public Schools	0.83	-3.57	0.72	0.78	-4.40	0.66	0.72
Virginia Beach	VA	Virginia Beach City Public Schools	0.65	0.91	0.26	0.33	0.26	0.17	0.26
Fresno	CA	Fresno Unified School District	0.62	-0.54	0.82	0.68	-1.16	0.62	0.53
El Paso	TX	El Paso Independent School District	0.49	-0.54	0.70	0.86	-1.03	0.59	0.80
Jacksonville	FL	Duval County Public Schools	0.41	0.00	0.42	0.51	-0.41	0.38	0.42
Houston	TX	Houston Independent School District	0.38	-0.64	0.80	0.88	-1.02	0.60	0.86
Colorado Springs	CO	Colorado Springs Public Schools	0.34	0.46	0.40	0.31	0.12	0.23	0.23
Fort Worth	TX	Fort Worth Independent Schools	0.34	-1.19	0.70	0.82	-1.53	0.53	0.70
Atlanta	GA	Atlanta Public Schools	0.29	-0.58	0.75	0.90	-0.87	0.74	0.92
Oakland	CA	Oakland Unified School District	0.28	-1.03	0.65	0.75	-1.31	0.60	0.73
Los Angeles	CA	Los Angeles Unified School District	0.27	-1.04	0.80	0.88	-1.31	0.73	0.82
San Francisco	CA	San Francisco Unified School District	0.25	0.50	0.01	0.35	0.25	0.65	0.38
Seattle	WA	Seattle Public Schools	0.15	0.39	0.39	0.34	0.24	0.19	0.31
Boston	MA	Boston Public Schools	0.15	-1.80	0.73	0.76	-1.95	0.46	0.73
Pittsburgh	PA	Pittsburgh Public Schools	0.12	-1.68	0.60	0.61	-1.80	0.42	0.56
Sacramento	CA	Sacramento Unified School District	0.10	-0.37	0.64	0.52	-0.47	0.59	0.43
Indianapolis	IN	Indianapolis Public Schools	0.08	-2.56	0.81	0.71	-2.64	0.64	0.59
Baltimore	MD	Baltimore City Public School System	0.04	-2.69	0.71	0.91	-2.73	0.65	0.85
Mesa	AZ	Mesa Unified School District	-0.08	0.92	0.53	0.38	1.00	0.19	0.19
Denver	CO	Denver Public Schools	-0.11	-1.81	0.65	0.76	-1.70	0.51	0.68
Phoenix	AZ	Paradise Valley Unified School District	-0.22	1.22	0.24	0.23	1.44	0.16	0.10
Tucson	AZ	Tucson Unified School District	-0.22	-0.14	0.56	0.60	0.08	0.34	0.47
San Diego	CA	San Diego Unified School District	-0.28	-0.53	0.60	0.58	-0.25	0.64	0.50
San Jose	CA	San Jose Unified School District	-0.48	-0.48	0.41	0.54	0.00	0.43	0.51
Las Vegas	NV	Clark County School District	-0.67	-1.06	0.46	0.51	-0.39	0.27	0.33
Detroit	MI	Detroit Public Schools	-1.12	-2.54	0.75	0.98	-1.42	0.64	0.93
Mean			0.49	-0.77	0.57	0.68	-1.26	0.51	0.61

Source: Author's calculations from state achievement files and the U.S. Common Core of Data.

Note: Only cities with full panels of data included

shifts.[42] The decline of test scores in Las Vegas (-0.67) could be related to the changing demographic characteristics of the school district.

Let's look at some specific cities with big gains. New Orleans leads the big cities in improvement. In 2000, the city was scoring nearly two standard deviations below (-1.98) the Louisiana state average but in 2007 scored 1.05 z-scores above average. Of course we are all aware of the tragic dislocation of students that occurred from Hurricane Katrina. The city has embarked on an ambitious set of reforms, including launching a vast network of charter schools with a much smaller (and more socially advantaged) population of students than before Katrina. The improvement in New Orleans test scores cannot be attributed to any recent reform and may be due to several factors. In addition to the dramatically changed student clientele, the effects of federal, state, and local education efforts in New Orleans cannot be easily disentangled. The city has attracted an army of school reformers, garners a lot of media attention, and will be studied intensely in the coming years.[43]

Other cities making large gains are also known for reform efforts. Three Texas cities make the list of top ten gainers: Dallas, Austin, and San Antonio. Texas was one of the earliest states to adopt an accountability system rewarding or sanctioning schools based on test scores. That system grew under both Democratic and Republican administrations in the state. Miami-Dade (ranked 6th in gains) was a nominee this year for the Broad Prize for Urban Education. The prize was first awarded in 2002. Past winners include New York (8th), Houston (17th), and Boston (25th). Making gains in academic achievement is one qualification for the Broad Prize but not the only one. Other criteria include closing gaps among racial and ethnic groups and adopting policies and practices that the experts on Broad's selection panel deem effective. Three smaller urban districts not included in the current study have also won the prize: Garden Grove Unified School District in California, Norfolk Public Schools in Virginia, and this year's winner, Brownsville Independent School District in Texas.

Table 3-2 looks at urban achievement from a state perspective. The analysis looks at a much larger set of districts. The original sample was selected in 2001. The current study collected recent achievement data on those same districts. Within each state, we examined test scores for districts coded as either "large central city" or "midsize central city" in the U.S. Department of Education's Common Core of Data. That database contains information on all schools and districts in the United States. We examined districts with at least 40 percent of students qualifying for free and reduced lunch, narrowing the analysis to the nation's poorest urban districts. In addition to the big cities listed in Table 3-1, several smaller cities are included in the figures for Table 3-2. We calculated z-scores in the manner described above, grouped the districts by state, and compared the results to the earlier analysis.

In Table 3-2, the states are sorted by gain from 2000 to 2007. As with the previous table, the most interesting finding comes from overall patterns, not the ranking of any one state. The picture is quite positive. About two-thirds of the states show z-score gains. The seven states with losses exhibit declines that are less than the standard errors for the 2007 statistic—so their test score losses could be statistical noise. Of the gainers, Minnesota and Wisconsin have large gains but the gains were generated by a small number of districts, only two districts in Minnesota (Minneapolis and St. Paul) and one in Wisconsin (Milwaukee). The gains by Louisiana (+0.55), Virginia (+0.50), and California (+0.15) are more impressive. The gains were generated by several districts, and the average gain exceeds twice the standard error for the statistic. California's gain of 0.15 may be negligible in terms of "real world" significance, however, as changes of less than 0.20 standard deviations are generally regarded.

Discussion

This study calculated how city school districts are performing relative to other school districts in the same state. Trends in performance are presented for 2000–2007. Big city schools have made significant improvement. Smaller urban districts have made similar gains. Unfortunately, the data cannot pinpoint reasons for the improvement, and any discussion of causes is automatically speculative. Given the fact that the rise in urban achievement is a

Some evidence suggests that the increases are associated with accountability systems that reward or sanction schools based on gains among low achievers.

Achievement in poor urban school districts, 2000–2007
(by state)

Table
3-2

	N	CHANGE IN Z SCORE	2007 Average Z Score	2007 Standard Error	2007 Free Lunch	2007 Black + Hispanic	2000 Average Z Score	2000 Standard Error	2000 Free Lunch	2000 Black + Hispanic
Minnesota	2	1.53	-1.78	0.03	0.69	0.49	-3.31	0.08	0.52	0.36
Wisconsin	1	0.84	-3.57	—	0.72	0.78	-4.41	—	0.66	0.72
Lousiana	6	0.55	0.02	0.27	0.70	0.67	-0.53	0.37	0.55	0.62
Virginia	9	0.50	-0.74	0.14	0.53	0.70	-1.24	0.25	0.50	0.64
Illinois	4	0.34	-1.55	0.32	0.70	0.82	-1.89	0.17	0.51	0.76
Pennsylvania	8	0.27	-1.78	0.43	0.67	0.59	-2.05	0.44	0.48	0.50
Florida	6	0.17	-0.26	0.10	0.56	0.50	-0.43	0.30	0.45	0.43
California	60	0.15	-0.55	0.07	0.63	0.62	-0.70	0.07	0.61	0.49
Massachusetts	3	0.11	-2.27	0.25	0.79	0.81	-2.38	0.22	0.52	0.75
Maryland	1	0.04	-2.69	—	0.71	0.91	-2.73	—	0.65	0.85
Texas	43	0.03	-0.57	0.11	0.55	0.82	-0.60	0.13	0.59	0.74
Georgia	7	-0.03	-0.75	0.14	0.67	0.76	-0.72	0.11	0.56	0.69
Colorado	2	-0.06	-1.16	0.64	0.65	0.69	-1.10	0.60	0.46	0.60
New York	12	-0.06	-1.94	0.20	0.69	0.61	-1.88	0.20	0.53	0.49
Arizona	10	-0.07	-0.72	0.10	0.72	0.87	-0.65	0.20	0.52	0.73
Indiana	6	-0.23	-2.50	0.31	0.70	0.62	-2.27	0.38	0.53	0.56
Washington	2	-0.27	-1.25	0.52	0.59	0.53	-0.98	0.40	0.44	0.41
North Carolina	1	-0.83	-0.57	—	0.53	0.61	0.26	—	0.40	0.57

Source: Author's calculations from state achievement files and the U.S. Common Core of Data.

Note: Sample only includes districts with a full panel of data from both years

national phenomenon and not confined to a handful of jurisdictions pursuing a single common reform strategy, the underlying cause of the improvement may be national as well. The largest gains in NAEP scores since 2000 have occurred at the bottom of the achievement distribution, among low-achieving students. Some evidence suggests that the increases are associated with accountability systems that reward or sanction schools based on gains among low achievers. With big city school districts serving large numbers of low achievers, NCLB-style accountability may be what is driving the rise in urban achievement.[44]

What are other possible explanations? What kind of reform strategies are urban districts attempting? The 2001 report discussed three reforms embraced by urban school reformers: school choice, standards, and class size. Those strategies continue to be popular. The most popular reform of recent years is mayoral control.

At the root of all this is the ambiguity of policy effects—and the need for more reliable information.

Mayoral control rests on the idea that the quality of city schools, like that of any municipal service, should be the responsibility of mayors. Mayors can better integrate community and school-based services and raise more federal and state money. When power over education is shared by mayors and elected boards of education—each with their own constituents—policies may be adopted that are fragmented, even contradictory. When mayors appoint school boards, voters can hold mayors accountable by rewarding or punishing them at the polls. Chicago, Boston, New York City, and Washington, D.C., are among the large cities that have increased mayoral control of educational systems in recent years. Mayors in these cities have used their newly acquired powers primarily to recruit high profile superintendents (or chancellors) with ambitious plans for school reform.

Has mayoral control worked? As noted by Hess in a 2008 review of the literature, the evidence thus far is inconclusive.[45] An early study by Wong and Shen (2003) concluded that eight districts with mayoral control had registered achievement gains in elementary grades, but gains were not as large in later grades.[46] McGlynn (2008) analyzed data from a larger sample, forty-seven city school districts, and found no effect of mayoral control on student achievement. He did find that mayoral control is associated with an increase of approximately $774 in per pupil expenditures.[47] Mayors use their additional powers to increase educational spending, it appears, but whether the added dollars improve learning in the schools is not known.[48]

The current study is unable to resolve the question. The number of big city districts with mayoral control is small, and they can be found at both the top and bottom of the list in Table 3-1. New York, Philadelphia, and Chicago have made large gains in student achievement, but several cities with elected, independent boards have made larger gains. In Boston, Oakland, and Baltimore, where mayors wield control, only small gains have been registered. The city at the bottom of the list, Detroit, had mayoral control for most of the years analyzed in the study, returning to an elected board in 2006.

Politics does not disappear with mayoral control. Mayors have a strong interest in presenting their pet programs in the best light. That includes programs for improving schools. With education authority concentrated in a single office, mayors become both the architects of change and the scorekeepers for tallying whether change is effective. Not surprisingly, they invariably find that new programs are working. When test scores go up, mayors claim credit for the increase. When test scores go down or stay flat, they point to other causes or claim that reforms were not implemented properly. Having multiple authorities over local education does not completely mitigate the problem—it is a product of incumbency rather than concentrated power—but power sharing increases the likelihood that skeptics or critics of a mayor's program can serve as a check on misguided policies.

At the root of all this is the ambiguity of policy effects noted above—and the need for more reliable information. Whether urban voters hold one or a dozen elected figures responsible for the productivity of schools, the public needs to have a much better grasp of what is going on. Urban leaders need to know which policies work. For many cities, only in the last decade have students been tested at regular intervals so that the outputs of educational systems can be gauged. Massive databases have been constructed at the state, national, and even international levels. But limited progress has been made in collecting data on the inputs to local systems—on the policies and practices that influence learning. Survey data on policies are collected at the international and national level, but a concerted effort to collect policy information at the local level would help answer two questions that arise from the current study. Why are test scores in big city school districts going up? Why are they going up more in some cities than in others?

An annual national inventory of local policies and practices would put us on the right track. Mayoral control, like so many reforms popular today, is a superficial descriptor. What really matters is how different mayors exercise their control. What policies are adopted? When? How are they implemented? A federal Schools and Staffing Survey regularly collects information on teachers and other educators: characteristics such as age, salary, training, and credentials.[49] Researchers can tap a wealth of information on the characteristics of teachers and track change over time. But researchers do not have high-quality data on the policies and practices of local jurisdictions. We need a national survey of policies and practices to be administered on a regular basis.

Consider a policy as simple as textbook selection. The nation's 14,000+ school districts spend enormous sums of money on textbooks. We know that textbooks shape the curriculum (what students are taught) and instructional practice in classrooms (how students are taught). Yet no database exists that describes, subject by subject, the textbooks used by schools. Some estimates of the popularity of various texts can be derived from publishers' data, but publishers are hardly disinterested parties in evaluating textbook usage. Like so many educational materials and practices, how textbooks are used varies from jurisdiction to jurisdiction. That is also true of mayoral control, school choice, school finance reform, class size, professional development, and a host of other policy issues. Knowing the details—the variations in policy from place to place—is essential for drawing correlations between policy and outcomes.[50]

Big city schools are making substantial gains in student achievement. We should be happy about that. But much remains to be done. Unfortunately, we are left to speculation as to the policies that produce substantial gains in student learning. To learn from recent successes and to sustain gains into the future, better data are needed on policies and practices—those that contribute to both our accomplishments and our failures.

NOTES

1 "Top Education Policy Organizations Form Expert Advisory Group on International Benchmarking," Press Release (Washington: National Governors Association, September 8, 2008). See also National Governors Association, Council of Chief State School Officers, and Achieve, Inc., *Benchmarking for Success: Ensuring U.S. Students Receive a World-Class Education* (Washington: Achieve, Inc., December 2008).

2 Dougal Hutchison and Ian Schagen, "Comparisons between PISA and TIMSS—Are We the Man with Two Watches?" in *Lessons Learned: What International Assessments Tell Us about Math Achievement* edited by Tom Loveless (Washington: Brookings Institution, 2007).

3 Full disclosure. Tom Loveless represents the United States in the IEA General Assembly and is a member of the PISA advisory committee for the United States.

4 www.european-agency.org/site/national_pages/finland/general.html. The strong showing of Finland on PISA has been attributed to its socio-constructivist pedagogy and curriculum. See Irmeli Halinen, Director of the General Education Division, Finnish National Board of Education, "The Finnish Curriculum Development Process," an address in Helsinki, Finland, August 12, 2005.

5 For a debate of the merits of situated learning, see John R. Anderson, Lynne M. Reder, and Herbert H. Simon, "Situative Versus Cognitive Perspectives: Form Versus Substance," *Educational Researcher* 26, no. 1 (1997): 18–21 and James G. Greeno "On Claims that Answer the Wrong Questions," *Educational Researcher* 26, no.1 (1997): 5–17. The discussion was triggered by an earlier article, John R. Anderson, Lynne M. Reder, and Herbert H. Simon "Situated Learning and Education," *Educational Researcher* 25, no. 4 (1996): 5–11.

6 *PISA 2006: Science Competencies for Tomorrow's World, Vol. 1* (Paris: Organisation for Economic Co-operation and Development, 2007), p. 347.

7 See *PISA 2006, vol.1*, p. 164.

8 *PISA 2006: Scientific Competencies for Tomorrow's World,* Executive Summary (Paris: Organisation for Economic Co-operation and Development, 2007), p.9. "Key Features" lists policy orientation first.

9 For a discussion of problems with drawing causal inferences from international data, see Stephen Raudenbush and Ji-Soo Kim, "Statistical Issues in Analysis of International Achievement," in *Methodological Advances in Cross-National Surveys of Educational Achievement* edited by Andrew Porter and Adam Gamoran (Washington: National Academy Press, 2002), pp. 267–294.

10 PISA has a separate item measuring self-concept.

11 Sharon Nelson-LeGall, "Help-Seeking Behavior in Learning," *Review of Research in Education*, 12 (1985): 55–90.

12 The problem begins with self-evaluations of performance. Hansford and Hattie's review of 128 studies reports correlation coefficients ranging from -.77 to .96 on the relationship of self-measures to external measures of performance and achievement. Brian C. Hansford and John A. Hattie, "The Relationship Between Self and Achievement/Performance Measures," *Review of Educational Research* 52, no. 1, (Spring 1982): 123–142.

13 Mary Ann Scheirer and Robert E. Kraut, "Increasing Educational Achievement Via Self Concept Change," *Review of Educational Research* 49, no. 1 (Winter 1979): 131–150.

14 Tom Loveless, *The 2006 Brown Center Report on American Education: How Well Are Students Learning?* (Washington: The Brookings Institution, 2006) pp. 12–20.

15 See *PISA 2006, vol.1*, p. 165.

16 See *PISA 2006, vol.1*, p. 137.

17 Alston Chase, *In a Dark Wood* (New York: Houghton and Mifflin, 1995).

18 See *PISA 2006, vol.1*, p. 161.

19 Debra Viadero, "Debate Erupts on How to Pick Chief of US Schools Data," *Education Week*, June 16, 2008, p.15.

20 *PISA 2006: Scientific Competencies for Tomorrow's World*, Executive Summary (Paris: Organisation for Economic Co-operation and Development, 2007), p.28.

21 Remarks by President Clinton, Education Roundtable, Springbrook High School, Silver Spring, Md., March 16, 1998. Available at www.ed.gov/inits/Math/timsroun.html.

22 Quoted in Matthew Bowers, "Virginia and the U.S. Are Improving Slightly at Math, but We Lag Behind Our Economic Competitors in the Developed World," *The Virginian Pilot,* March 28, 1997, p. B3.

23 Robert Moses, "Algebra, the New Civil Right," in *The Algebra Initiative Colloquium, Vol. II,* edited by Carol Lacampagne and others (Department of Education, 1995), pp. 53–67.

24 Data available on the main NAEP data explorer: nces.ed.gov/nationsreportcard/nde/. See also Jay Matthews, "Adding Eighth Graders to the Equation," *The Washington Post,* March 12, 2007, p. B1.

25 Carolyn Shettle and others, *America's High School Graduates: Results from the 2005 NAEP High School Transcript Study* (Department of Education, 2007), p. 11. Other than calculus, advanced math is defined as pre-calculus or AP statistics.

26 Julia B. Smith, "Does an Extra Year Make Any Difference? The Impact of Early Algebra on Long-term Gains in Mathematics Attainment," *Educational Evaluation and Policy Analysis* 18 no. 2 (1996): 141–153.

27 Both California and D.C. schools serve a large number of students of low socioeconomic status. Nevertheless, the math scores for California and D.C. look dismal even if comparisons are made among similar groups of students in terms of race, ethnicity, and parental education.

28 Other includes all other courses, along with no responses and multiple responses, on the NAEP survey item.

29 In 2005, approximately 4.2 million students were enrolled in eighth grade. The estimate of 120,000 comes from [0.078 * 0.366 * 4,200,000 = 119,901.6]. All other figures in this paragraph are author's calculations from restricted-use NAEP data.

30 U.S. Department of Education, *Foundations for Success: The Final Report of the National Mathematics Advisory Panel* (Washington: 2008).

31 William Sanders, "Teachers, Teachers, Teachers," *Blueprint Magazine,* September 1, 1999.

32 William Schmidt and others, "Relationship of Tracking to Content Coverage and Achievement: A Study of Eighth Grade Mathematics," (Michigan State University, 2008).

33 Several years ago, a prominent education scholar gave a talk at Brookings in which she commented that her own son had never learned the multiplication tables but went on to graduate from an Ivy League college.

34 U.S. Department of Education, *Foundations for Success: The Final Report of the National Mathematics Advisory Panel* (Washington: 2008).

35 Richard J. Murnane and Frank Levy, *Teaching the New Basic Skills* (New York: The Free Press, 1996).

36 Andrew Zau and Julian Betts, *Predicting Success, Preventing Failure: An Investigation of the California High School Exit Exam* (San Francisco: Public Policy Institute of California, 2008).

37 Indeed, Kennedy mentioned the first astronaut in space, Alan Shepherd, whose suborbital flight had occurred about three weeks before the speech. President John F. Kennedy, "Special Message to the Congress on Urgent National Needs," May 25, 1961. Text available at: www.presentationhelper.co.uk/kennedy_man_on_the_moon_speech.htm.

38 National Center for Education Statistics, *The Nation's Report Card: 2007 Trial Urban District Assessment in Mathematics* (Department of Education, 2007). Also see Jason Snipes and others, *Beating the Odds VIII: An Analysis of Student Performance and Achievement Gaps on State Assessments: Results from the 2006–2007 School Year* (Washington: Council of the Great City Schools, 2008).

39 The top 50 U.S. cities in population are based on the 2000 census. Only cities with a full panel of data are included in the analysis.

40 It would be a less stringent standard if all scores were going down and the gap closing was because of city scores going down less.

41 We also dropped from the analysis charter schools that were treated as separate districts in the state databases.

42 Demographic statistics were obtained from the Common Core of Data "Build a Table" database. Data refer to the 2005–2006 year and are available for download at nces.ed.gov/ccd/bat/. We found a few cities with suspect data for free and reduced lunch. San Antonio, for example, changed from 80 to 19 percent students on free and reduced lunch between our two sample years.

43 For an example of popular press coverage of New Orleans, see Walter Isaacson, "The Greatest Education Lab," *TIME Magazine*, September 17, 2007, p. 47.

44 Tom Loveless, Steve Farkas, and Ann Duffett, *High-Achieving Students in the Era of NCLB* (Washington: Thomas B. Fordham Institute, 2008).

45 Frederick M. Hess, "Looking for Leadership: Assessing the Case for Mayoral Control of Urban School Systems," *American Journal of Education*, 144 no. 3 (2008): 219–245.

46 Kenneth K. Wong and Francis X. Shen, "Big City Mayors and School Governance Reform: The Case of School District Takeover," *Peabody Journal of Education* 78 no. 1 (January 2003): 5–32.

47 Adam J. McGlynn, "Assessing the Effect of Mayoral Control on Educational Inputs and Achievement," Paper presented at the Midwest Political Science Association, Chicago, April 3–6, 2008.

48 Also see Kenneth Wong and others, *The Education Mayor: Improving America's Schools* (Washington: Georgetown University Press, 2007).

49 See NCES' website on the Schools and Staffing Survey, nces.ed.gov/surveys/sass/.

50 A good start in this direction is exemplified by the surveys in Trish Williams, Michael Kirst, Edward Haertel and others, *Similar Students, Different Results: Why Do Some Schools Do Better? A large-scale survey of California elementary schools serving low-income students* (Mountain View, Calif.: EdSource, 2005).

Views expressed in this report are solely those of the author.

Designed by Lipman Hearne, Chicago/Washington

BROWN CENTER on
Education Policy
at BROOKINGS

Brookings
1775 Massachusetts Avenue, NW • Washington, D.C. 20036
Tel: 202-797-6000 • Fax: 202-797-6004
www.brookings.edu

The Brown Center on Education Policy
Tel: 202-797-6469 • Fax: 202-797-2480
www.brookings.edu/brown.aspx