Advances in Evidence-Based Education

Volume 3

Performance Feedback: Using Data to Improve Educator Performance

Edited by
Ronnie Detrich
Randy Keyworth
Jack States

The Wing Institute

Advances in Evidence-Based Education

Volume 3

Performance Feedback: Using Data to Improve Educator Performance

Advances in Evidence-Based Education

Volume 3

Performance Feedback: Using Data to Improve Educator Performance

Edited by:
Ronnie Detrich
Randy Keyworth
Jack States

The Wing Institute
Oakland, CA

Published by The Wing Institute
2102 Dennison Street, Suite B
Oakland, CA 94606
510.534.9464
www.winginstitute.org

Volume 3 published Fall 2013.

ISSN 1941-1693
ISBN 978-0-9800177-2-4

Printed in the United States of America.
Printed on acid-free paper

Dedication

These proceedings from the 6th annual summit are dedicated to the memory of Ernie Wing, who is also the namesake of the Wing Institute. Ernie championed evidence-based education as an educator and child advocate. As an educator, he founded Spectrum Center, which has been a beacon for evidence-based practice and state-of-the art educational services since 1975. As an advocate, Ernie served hundreds of families with the most challenging special education needs, gaining the admiration and respect of both parents and school districts. Through his efforts, Ernie set the standards for professionalism, integrity, effectiveness, and caring as he helped thousands of children gain access to effective educational services. He was a good man and a good friend, and is missed.

Acknowledgments

We are indebted to the people who made this book possible. We wish especially to recognize the work of Jin An in managing the process of producing this publication. Without her patience and organization skills, we would be still in the conceptual phase of the book. We are grateful for the outstanding work of our copy editor, Susan Lang. Her ability to keep the writing clear and concise was essential in completing the book. We would like particularly to recognize the work of the three speakers from the 2011 summit: Dr. Mary Beth Celio, Dr. Aubrey Daniels, and Dr. Amanda VanDerHeyden. Their three presentations along with a contribution from the Wing Institute's Randy Keyworth form the basis of this book.

Contents

Introduction

Proceedings From the Wing Institute's Sixth Annual Summit on Evidence-Based Education

Performance Feedback: Using Data to Improve Educator Performance

JACK STATES
RANDY KEYWORTH
RONNIE DETRICH
The Wing Institute

Over the past 6 years, the Wing Institute has examined the state of education with the goal of building an evidence-based education model that can make a marked difference in meeting society's need for academically proficient young people. Much has been written about the disappointing performance of the American education system, most evident in the flat test scores of the past 30 years. Unfortunately, the most recent attempt to alter this picture, No Child Left Behind (NCLB), has hardly made a dent in changing student performance, as measured by high-stakes testing and graduation rates. Accountability and national standards that are at the core of NCLB, though initially praised across the political spectrum, have recently been called into question because of the lack of data to support students are making substantive progress.

Along with NCLB, other structural fixes have proliferated at the state and local levels: class size reduction, charter schools, increased spending, school sizes (large and small), stricter teacher credential standards, and voucher systems. When taken to scale, all of these interventions championed as methods to improve student performance have consistently disappointed school reformers (Yeh, 2007).

Against this backdrop of failure upon failure in school reform, the Wing Institute has presented a series of "summits." Each annual 2-day summit is designed to look at a specific pressing issue, provide the latest research on the topic, analyze the practices that are working, and explain what might be done to increase the likelihood of success when schools select and implement reform. Nationally recognized speakers are selected to present at a working session as

well as stimulate and lead the day's discussion. Participants are able to interact with a diverse group of stakeholders in our education system: researchers, university faculty, school administrators, national and local policy makers, service providers, and parents.

PERFORMANCE FEEDBACK: USING DATA TO IMPROVE EDUCATOR PERFORMANCE

This book is compiled from the proceedings of the sixth summit entitled "Performance Feedback: Using Data to Improve Educator Performance." The 2011 summit topic was selected to help answer the following question: What basic practice has the potential for the greatest impact on changing the behavior of students, teachers, and school administrative personnel?

The following chapters are developed from the presentations at the Wing Institute's 2011 summit. Research consistently finds feedback to be a powerful method for changing performance (Fuchs & Fuchs, 1986; Hattie, 2009; Walberg, 1999). Whether in the form of monitoring student progress, coaching teachers on how to implement practices, or requesting input needed to continuously improve organizations, feedback remains at the core of potent and successful change. The promise that accountability, as measured by high-stakes tests required in NCLB, is enough to boost student performance is too good to be true. In reality, outcome data are not especially effective in improving any category of performance without feedback on the practices that lead to the outcome, such as high-stakes tests. Successful use of data to change student or teacher performance requires systematic feedback on important outcomes—mainly, student achievement—as well as on teacher instructional skills and school administrator support promoting the acquisition of learning.

The chapters developed for this book expand on the 2011 proceedings to provide readers an in-depth examination of each topic. The chapter authors offer their wide range of experience and knowledge, from education to corporate organizational development, to help educators design and implement performance feedback systems.

Randy Keyworth, senior fellow at the Wing Institute, contributed the first chapter: *Feedback at the System Level: Benchmarking U.S. Education Performance*. Keyworth examines sources of data on the performance of the American education system. This information provides a historical context for judging the effectiveness of education and current practices to achieve meaningful results. He lays out the argument for establishing benchmarks as yardsticks of achievement to guide reform efforts. The highlighted benchmarks consist of high-stakes testing outcomes and graduation rates; process measures to gauge the performance of all educators including teachers and principals;

and system measures including equal access to resources and effective teacher preparation practices.

In the second chapter, *Feedback in Education: On Whom and for What*, Dr. Aubrey Daniels offers a perspective on education reform derived from his extensive experience in business. He provides examples from his work to improve performance, paying particular attention to the role of performance feedback.

Daniels reinforces the critical need for education stakeholders to define why schools exist. Only then can objective measures be established to tell us if current practices are achieving what we want of them. He emphasizes the need for implementing effective feedback mechanisms to ensure the system as well as individuals are performing as expected. He defines performance feedback and clarifies the purpose of feedback in education.

In the third chapter, *Seeking the Magic Metric: Using Evidence to Identify and Track School System Quality*, Dr. Mary Beth Celio reviews the impact that accountability and budget cuts have placed on embattled educators. For many educators, NCLB was viewed as a quick fix to solve the many problems of an ailing system through setting goals and offering incentives and consequences. Its failure to meet the very high expectations of so many people has resulted in calls for its repeal. Despite the failure of NCLB to meet expectations, Celio believes the law has had positive effect through focusing attention on standardizing the tracking and use of outcome data.

Celio examines the purpose of different types of data available to educators that can drive performance. Celio underscores the need for choosing indicators that enable educators to intervene early when there is still time to change performance before failure occurs. The chapter stresses the importance of finding indicators that are not only meaningful but also easy for users to read. The design and use of key indicator reporting is presented in a way that is practical, enabling educators to develop reports adapted to meet the needs of their own unique settings.

In the fourth chapter, *Are We Making the Differences That Matter in Education?*, Dr. Amanda VanDerHeyden criticizes the education system for paying too little attention to defining what it means when schools are successful. She argues that this leads us to adopt solutions that on the surface appear to address deficits, but are often not correlated with improved student achievement.

VanDerHeyden discusses the necessity for stakeholders to focus more on outcomes and not just process. She examines popular interventions with superficially broad appeal that have proved ineffective in achieving results, such as class size reduction and poorly designed reading curriculum. She addresses reasons why effective practices often fail to achieve the desired outcome, and the fact that many practices are implemented improperly with

little attention paid to treatment integrity. VanDerHeyden notes the many costs and profound negative impact on schools as a consequence of selecting ineffective practices —leading to a downward spiral of disenchantment with reform efforts. She details a model for responding effectively to the needs of students, Response to Intervention (RtI), highlighting the components that rely heavily on feedback as a reliable, cost-effective solution for overcoming deficits inherent in the current education system.

As a whole, these chapters provide an important look at one of the most powerful tools available to educators—performance feedback. It effectively links school reform initiatives to improved student performance.

REFERENCES

Fuchs, L. S. & Fuchs, D. (1986). Effects of systematic formative Evaluation: A meta-analysis. *Exceptional Children*, *53*(3), 199–208.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* New York: Routledge.

Walberg H. J. (1999). Productive teaching. In H. C. Waxman & H. J. Walberg (Eds.) *New directions for teaching, practice, and research* (pp. 75–104). Berkeley, CA: McCutchen Publishing.

Yeh, S. S. (2007). The cost-effectiveness of five policies for improving student achievement, *American Journal of Evaluation*, *28*(4), 416–436.

Chapter 1

Feedback at the System Level: Benchmarking U.S. Education Performance

RANDY KEYWORTH
JACK STATES
RONNIE DETRICH
The Wing Institute

ABSTRACT: *This chapter examines the performance of the U.S. K–12 education system over time, in comparison to other nations, and at different levels of organizational structure: states, school districts, and schools. It uses macro-level, aggregate data to benchmark outcomes in four critical categories of performance: participation, quality, equity, and efficiency. It also reviews previous and current attempts at system-level feedback and accountability. The resulting picture portrays an education system that has never had adequate performance outcome data to guide its decisions. Recent efforts by No Child Left Behind (NCLB) to establish Adequate Yearly Progress (AYP) resulted in metrics that have no uniform standards and tremendous incentives for states to interpret data as positively as possible. Review of existing macro-level data from enrollment, graduation rates, standardized tests, demographics, and resource allocation databases describe a system failing in all four critical performance areas. The United States ranks below at least 20 nations in enrollment of eligible school-age children students and in high school graduation rates (participation). Slightly more than one third of students are proficient in reading and mathematics, while only 75% graduate from high school (quality). There is a wide disparity in student performance, quality of resources, and funding between students of color and socio-economic backgrounds (equity). And there seems to be little link between total resources spent and performance outcomes (efficiency).*

The value of a nation's education system is measured by how well it serves *all* of its children, not just those fortunate enough to attend a model school or live in a high-performing school district. While there are numerous examples of such exemplary schools and school systems in our country, this chapter portrays an education system that has been failing a significant majority of its students for decades. Part of this failure has been a

lack of empirical, meaningful, and ongoing performance feedback at the system level. This chapter examines our nation's history of evaluating educational progress, identifies critical performance outcomes for an effective education system, and provides a snapshot of how our nation is performing against those benchmarks.

Education is increasingly referred to as the civil rights issue of our generation. Few social institutions have more impact on the health and well-being of a nation's citizenry. As a culture, we hold the assumption, codified in federal and state laws, that all students have access to an equal and effective education. Yet we have not systematically evaluated our progress toward this goal. The recent focus on high-stakes testing is a start, but only one piece of the puzzle. While academic test scores represent one critical performance outcome, an effective egalitarian education system must do much more. A broader set of outcomes was proposed by the Organisation for Economic Co-operation and Development (OECD), which has studied education systems across nations since the early 1960s. It concluded, "Governments need to create education systems that are accessible to everyone, not just a favoured few; that are globally competitive on quality; that provide people from all classes a fair chance to get the right kind of education to succeed; and to achieve all this at a price that the nation can afford…Put another way, this volume defines superior performance as high participation, high quality, high equity, and high efficiency" (Organisation for Economic Co-operation and Development [OECD], 2011).

This chapter analyzes our education system's performance in these four areas: (a) *participation* (how well our education system serves *all* school-age children); (b) *quality* (how well it meets identified educational goals); (c) *equity* (how fairly it allocates resources and achieves comparable outcomes across all categories of students); and (d) *efficiency* (how well it invests its resources in terms of achieving stated outcomes). There is now data that make it possible to benchmark our progress on these outcomes over time and against other educational systems.

A key element of benchmarking is the identification of critical performance indicators. Without such indicators, it is impossible to evaluate the effectiveness of our education system or make sound decisions about school reform initiatives. Indicators need to be *reliable* (repeated measures of the same event yield the same score), *valid* (they measure what we think they are measuring), and *socially relevant* (the outcomes reflect society's values). To make matters more challenging, both process measures and outcome measures are needed. Without process measures (treatment integrity) to tell us if education interventions are being implemented as designed, it is virtually impossible to draw conclusions. And, finally, outcome measures need to be collected on the performance of *all* aspects of the education system: students, staff, and organizations.

An effective education system requires meaningful and accurate feedback data for evaluating education performance at both a micro level (individual student and staff performance) and a macro level (system performance at different units of scale, i.e., school, district, state, and nation). This chapter examines our performance at the macro level. There exists a wealth of macro data on the education system's overall performance across time that sheds light on our performance in the areas of participation, quality, equity, and efficiency. We just haven't been using it to drive policy decisions.

METRICS FOR BENCHMARKING EDUCATIONAL PERFORMANCE OUTCOMES AT A MACRO LEVEL

Benchmarking education performance at the macro level has its limitations. The scale of measurement is large. The indicators often reflect the aggregation of data from multitudinous units of performance. Drawing conclusions about specific causal relations becomes very challenging as the data may consist of performance averages, include the cumulative impact of numerous interventions, and reflect snapshots in time (e.g., annual data). And given the scale of the analysis, changes in performance often move slowly, not unlike a large ocean-going vessel changing course. However, macro indicators represent critical performance outcomes, as ultimately an education system must be measured by the overall sum of its parts. It does matter how all the students are performing, especially when equity is important. And while there are limitations to which conclusions can be drawn, this chapter presents an overwhelming preponderance of evidence showing an education system in crisis. By virtually every macro indicator, we are failing the goals of high participation, high quality, high equity, and high efficiency.

There is an increasing amount of macro-level data being generated that can be used to benchmark an education system's critical performance outcomes. As the balance of this chapter will reference these resources, it is worth taking some time to discuss their relative strengths and weaknesses.

Standardized tests

Education stakeholders are in a constant debate about what constitutes a quality education and how best to measure student outcomes. In particular, there is significant disagreement about the use and value of high-stakes standardized tests. Both their validity (what they measure) and reliability (how well they measure) are often disputed when applied at the micro level (evaluating individual students or teachers). Despite these questions, standardized tests

provide extremely valuable student performance measures at the macro level over time. They may not measure every desired education achievement, but they can assess one of the system's most important outcomes: what students have learned in selected content areas (e.g., reading, math). And while some of the standardized tests used in different states and localities may merit criticism from validity and reliability perspectives, there are national and international standardized tests that meet the highest standards of reliability, validity, and social relevance. Data from these tests provide a clear and unambiguous picture of how well the U.S. education system is educating students on selected measures. These national and international tests include the following:

National Assessment of Educational Progress (NAEP): NAEP has often been called the gold standard for standardized academic testing because of its constant rigorous scrutiny (National Center for Education Statistics [NCES], 2010a). Established in 1964, with the first tests administered in 1969, NAEP provides a continuing assessment of what American students know and can do in math, reading, science, writing, the arts, civics, economics, geography, and U.S. history. NAEP is administered by the National Center for Education Statistics (NCES), a division of the Institute of Education Sciences in the U.S. Department of Education. Panels of technical experts within NCES and other organizations continually scrutinize tests for reliability and validity, keeping them similar from year to year and documenting changes. It is one of the only common metrics for all states, providing a picture of student academic progress over time.

Program for International Student Assessment (PISA): PISA is a carefully constructed and well-documented test instrument for measuring student academic performance across nations (Organisation for Economic Co-operation and Development [OECD], 2006). Coordinated by the Organisation for Economic Co-operation and Development, this international study has been conducted every 3 years since 2000. It measures the performance of 15-year-old students in 64 countries (34 member nations and 30 participating nations) in reading, mathematics, and science. In addition to reporting on test scores, PISA collects data on a large number of education system characteristics and identifies statistical correlations between results and selected variables.

Graduation rates

Few performance indicators have more significant social relevance than high school graduation rates. Research data from 2005–07 show that high school dropouts have a 50% higher unemployment rate than high school graduates (U.S. Bureau of Labor Statistics, 2013), earn 50% less income (U.S. Census Bureau, 2011), are 44% more likely to be in less than very good health (Egerter et al., 2009), and 530% more likely to be incarcerated (Sum, Khatiwada,

McLaughlin, & Palma, 2009). Yet, there has been a significant lack of valid and reliable data collected, analyzed, and reported at any level of the education system (school, district, state, national). Historically, some states failed to produce any graduation rate data whatsoever (Hall, 2005). Those that did often failed to account for students who left school prior to the 12th grade, dramatically skewing the data (Hall, 2005). This changed in 2011 with the new federal guidelines establishing a "four-year adjusted cohort graduation rate" (U.S. Department of Education, 2008). Prior to that, there have been other models that attempted to capture reliable historical data. The following have been used to obtain historical performance data in this metric:

Four-Year Adjusted Cohort Graduation Rate (Cohort Graduation Rate): The 4-year adjusted cohort graduation rate is the number of students who graduate in 4 years with a regular high school diploma divided by the number of students who entered high school 4 years earlier. It was adopted in 2008, when the U.S. Department of Education enacted regulations establishing a uniform and more accurate measure for calculating the rate at which students graduated from high school. Starting in the 2010–11 school year, the 4-year adjusted cohort graduation rate captures all students, including those who drop out in earlier grades. Above all, it is a metric that is uniform across all 50 states and can be used over time (U.S. Department of Education, 2008).

Average Freshman Graduation Rate (AFGR): Prior to implementation of the Adjusted Cohort Graduation Rate, the NCES developed a model for estimating graduation rates using enrollment data that accounted for students who were enrolled in the ninth grade but did not finish school. Based on a technical review and analysis, the AFGR was selected as the most accurate indicator from a number of alternative estimates that can be calculated using available cross-sectional data (Stillwell, Sable, & Plotts, 2011).

Cumulative Promotion Index (CPI): The CPI uses enrollment and diploma-count data from the U.S. Department of Education to approximate the probability that a student entering the ninth grade will complete high school on time with a regular diploma. It averages the percentage of students who successfully transition between grades (from 9 to 10, 10 to 11, and 11 to 12) to generate a graduation rate that is inclusive of all students. It is used by the Editorial Projects in Education (*Education Week*), Harvard Civil Rights Project, Urban Institute, and Education Commission of the States, among other groups. While it is not a true cohort, it is recognized as an accurate estimate (Hall, 2005).

Education system databases

A tidal wave of macro-level data on education system performance is being generated annually at all levels of the system: school, school district, state,

national, and international. These data are increasingly useful in benchmarking the performance of systems against each other and over time. A sample of these databases include:

The Condition of Education: Published annually by the NCES, *The Condition of Education* reports important developments and trends in education, including 49 indicators on the status and condition of education. The 2012 report examined data in three main areas: (a) participation in education; (b) elementary and secondary education and outcomes; and (c) postsecondary education and outcomes. It has been published annually since 1989, providing over 20 years of data with which to benchmark education performance at the system level in this country (Aud et al., 2012).

Digest of Educational Statistics: Published annually by the NCES, the *Digest of Educational Statistics* provides a compilation of statistical information covering the broad field of American education from pre-kindergarten through graduate school. The digest contains data on a wide variety of topics across all levels (students, staff, organization) relating to enrollment rates, educational attainment, student and family demographics, teacher characteristics, finances, and instruction. It has been published annually since 1962, providing over 50 years of data with which to benchmark education performance at the system level in this country (Snyder & Dillow, 2012b).

Education at a Glance: Produced annually by the OECD Centre for Educational Research and Innovation, *Education at a Glance* has become a leading international compendium of comparable national statistics measuring the state of education worldwide. The report analyses the education systems of the 34 OECD member countries, as well as those of 30 participating countries. It looks at who participates in education, the level and type of resources committed, how education systems operate, and the results achieved. The last includes indicators on a wide range of outcomes, from comparisons of student performance in key subject areas to the impact of education on adults' earnings and chances of employment. It has been published since 1998, providing data with which to benchmark the performance of the United States against other nations (Organisation for Economic Co-operation and Development [OECD], 2012a).

Return on Investment Analyses

Benchmarking requires more than comparative performance data. It also requires analysis of a system's use of resources in relation to what works and what doesn't. In a time of diminishing resources it becomes more critical than ever to identify interventions that produce the best results the most efficiently. Simply spending more money on education will not necessarily produce better outcomes. There are extremely well-funded school districts that are failing, and

less well-funded districts that are succeeding. The question becomes: Which interventions secure the best outcomes with the most cost-effective use of resources? To answer this question, a number of education systems are utilizing return on investment (ROI) analyses. A measure of how efficiently resources are producing results, ROI is a formula in which the benefit of an investment is divided by its costs. It has long been used in the world of business but historically has been resisted in the field of education. This is changing. ROI analyses are increasingly showing up in both education research and operations. In addition to a growing number of well-designed studies on this issue (reviewed later in this chapter), public education systems are beginning to track ROI as part of their ongoing school reporting measures. For example, Florida has developed an online individual school report card that documents performance and ROI for each school and school district in the state.

A HISTORY OF "FLYING BLIND"

Prior to No Child Left Behind (NCLB), the only education performance data that the federal government required individual states to report were data on student dropout rates. Any other production of education performance outcome data was left up to each state. The result was a hodgepodge of 50 different accountability systems that had one thing in common: They seldom met the standards of reliability, validity, and social relevance. State achievement tests varied significantly in terms of rigor, frequency, grade levels assessed, subject matter tested, and cut scores (the selected score that separates test takers into various categories, such as a passing score and a failing score). Test formats often changed, preventing comparison of one year with another. Formulas for calculating graduation rates (when they were reported at all) were as creative as they were inaccurate. This absence of reliable and valid feedback contributed to reform efforts that relied on opinion, philosophy, ideology, and fads. This lack of data contributed to the failure of our nation's education system to improve over the past 40 years. The road to implementing reliable and valid performance outcome metrics has been rocky.

The first serious attempt to implement reliable and valid metrics on a national scale occurred just over a decade ago. In 2001, NCLB attempted to enforce accountability standards through the concept of Adequate Yearly Progress (AYP), which among other provisions required each state to adopt and report high-stakes academic testing scores and high school graduation rates. The flaw in the plan was the absence of uniform standards. States were allowed to select their own tests, develop their own standards and proficiency cut scores, establish their own annual targets, and define their own formulas for graduation

rates. Most of the pre-NCLB flaws remained in place. Except now there were significant consequences for failing to meet AYP targets (schools faced increasing sanctions leading up to a massive overhaul of site leadership and staff), which gave states enormous incentives to report data in as flattering a way as possible. As will be seen, this variability showed up all levels: overall AYP calculations, standardized tests, and graduation rates.

AYP variability

One study demonstrated the inconsistency of the AYP metric across different states (Cronin, Dahlin, Xiang, & McCahon, 2009). It took actual performance data from 36 randomly selected schools (18 elementary schools and 18 middle schools) located around the country and applied the AYP standards from 28 different states to see how individual schools would fare in different states. The results for elementary schools are reflected in Figure 1:

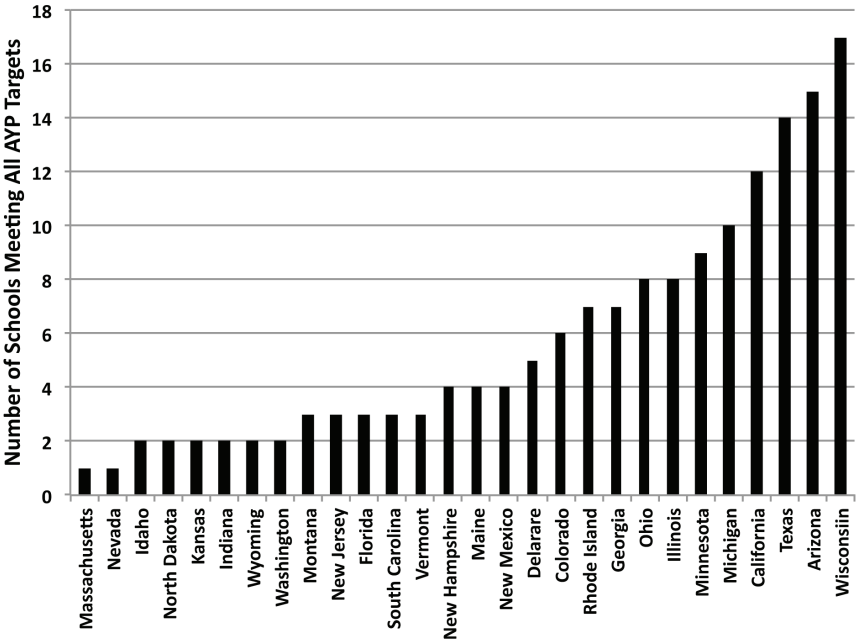

Figure 1. How individual schools fared using AYP criteria of 28 different states.

Adapted from *The Accountability Illusion* (p. 21), by J. Cronin, M. Dahlin, Y. Xiang, and D. McCahon, 2009, Washington, DC: Thomas B. Fordham Institute. In the public domain.

The data show great inconsistency between the AYP standards of different states, with some states having significantly more stringent requirements than others. When the study applied the AYP criteria adopted by Massachusetts or Nevada, only 1 of the 18 elementary schools met those states' targets. On the more lenient end of the continuum was Wisconsin, where 17 of these same 18 elementary schools met the AYP criteria. The remaining 25 states would have certified between 3 and 15 schools. This level of inconsistency clearly raises questions about the reliability, validity, and social relevancy of the AYP metric in the absence of uniform standards. Yet, AYP has been a cornerstone accountability measure for one of the most significant school reform initiatives in history.

Further scrutiny of AYP variability takes us to an analysis of two of its major components: state standardized testing and graduation rate data. Fortunately, there are established metrics for a benchmark analysis of each.

High-stakes academic testing variability

One of NCLB's fundamental goals is that *all* children will be "proficient" in reading and math by 2014.

As with AYP, states have very different standards for establishing cut scores and identifying which students meet "proficiency" in a given subject area. NCES generates ongoing research that compares NAEP proficiency standards against those of individual states. Figure 2 compares the percentage of students who met proficiency for fourth-grade reading according to various state tests versus the percentage of the same students who met proficiency according to NAEP results. The states are ranked from largest to smallest gap between state and NAEP standards.

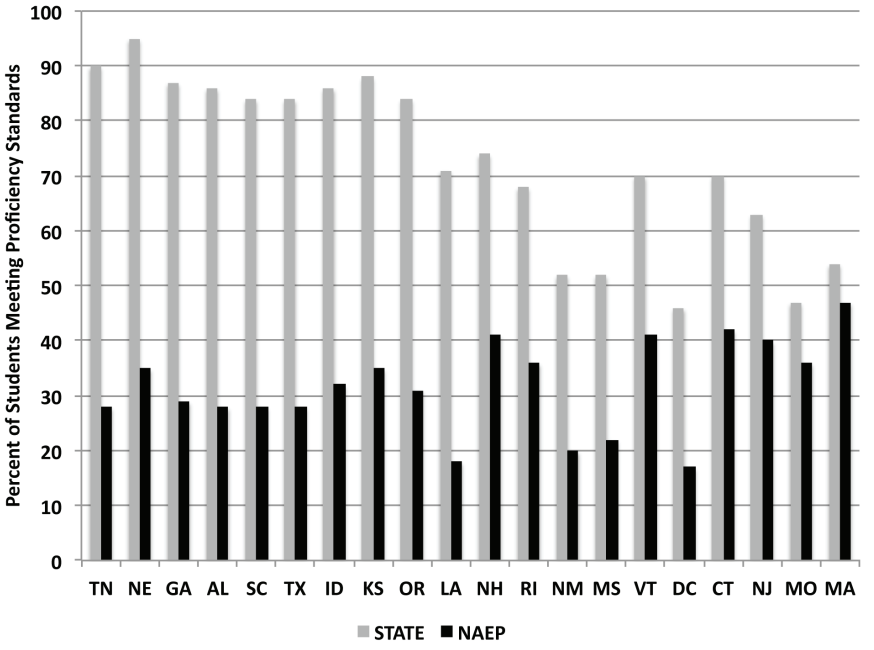

Figure 2. Percent of students meeting proficiency standards for state test versus NAEP test (reading fourth grade, 2009).

Data are drawn from National Center for Education Statistics (NCES) (2011f).

The data show a dramatic disconnect between the proficiency standards of states and those of NAEP. For example, Tennessee reported that 90% of its fourth-grade students were at or above reading proficiency. NAEP data for Tennessee reported only 28%. Similar gaps occurred across the other states analyzed. Only in Massachusetts were state test results comparable to NAEP's (54% to 47%). Comparable gaps existed across grades and in mathematics as well as reading. The following table analyzed the average proficiency outcomes for all states and the District of Columbia compared to their NAEP proficiency rates for the same students.

Table 1.
Average fourth- and eighth-grade reading and math scores (state testing versus NAEP), 2009

	Reading		Mathematics	
	Grade 4	**Grade 8**	**Grade 4**	**Grade 8**
% students reported as proficient using state proficiency standards	74	72	73	66
% students reported as proficient using NAEP proficiency standards	32	31	39	33
% difference in proficiency	42	41	34	33

Data are drawn from National Center for Education Statistics (NCES) (2011f).

On average, in 2009, states reported twice as many students proficient in reading and math than did NAEP. As with AYP, these data also demonstrate significant inconsistency between states. But given the integrity of the NAEP testing process, the data also suggest that many states established tests or cut scores that artificially inflated student achievement. Relying on such inaccurate data makes it very difficult for states to draw the right conclusions about progress in their education systems.

NCES examined state proficiency standards in the context of NAEP's three achievement levels, or benchmarks, for student performance: "*Advanced"* represents superior performance, "*proficient*" represents solid academic performance, and "*basic*" denotes partial mastery of prerequisite knowledge and skills fundamental for proficient work at each grade. (NAEP also reports data on students who are *"below basic*"). "*Proficiency*" becomes a critical benchmark because it is the level at which students have met the standards for a subject area. It is also the benchmark by which NCLB holds school districts accountable (National Center for Education Statistics [NCES], 2011c).

NCES concluded that most state proficiency standards were not just below NAEP's proficiency levels, but were actually at or below NAEP's definition for

basic performance. A sample of the findings include:

- In fourth-grade reading, 35 of the 50 states included in the analysis set standards for *proficiency* (as measured on the NAEP scale) that were lower than the scale score for *basic* performance on NAEP. The remaining 15 states' *proficiency* standards were in NAEP's *basic* range. This meant that most states identified students as proficient readers when they were actually below partial mastery of reading skills.
- In fourth-grade mathematics, 7 of the 50 states included in the analysis set standards for *proficiency* (as measured on the NAEP scale) that were lower than the scale score for *basic* performance on NAEP, 42 were in NAEP's *basic* range, and 1 in NAEP's *proficient* range. Again, state standards were much lower than NAEP standards.

(Bandeira de Mello, 2011)

Graduation rate variability

The other AYP pillar of evaluation—graduation rates—has long been considered an important metric for measuring education progress. Yet, until recently there has been no established uniform standard for calculating this metric. Unfortunately, as with testing, NCLB provided states with serious incentives to report high graduation rates and maximum flexibility on how they calculated the rates. The results showed a clear pattern of misusing standards and data to overstate graduation rates. For example, very few states included students who dropped out prior to the 12th grade. North Carolina used a calculation based on the percentage of graduates who got their diplomas in 4 years or less, ignoring the number of students who dropped out. New Mexico reported only the percentage of 12th graders who graduated, ignoring students who dropped out in the 9th, 10th, and 11th grades. Alaska's graduation rate was based on the number of students who graduated divided by the number of students enrolled on the last day of school (Hall, 2005).

These standards produced data that were not an accurate reflection of actual graduation rates. How inaccurate was this representation? Figure 3 displays the five states with the greatest discrepancy between state-reported data and data from two more accurate graduation-rate models (CPI and AFGR).

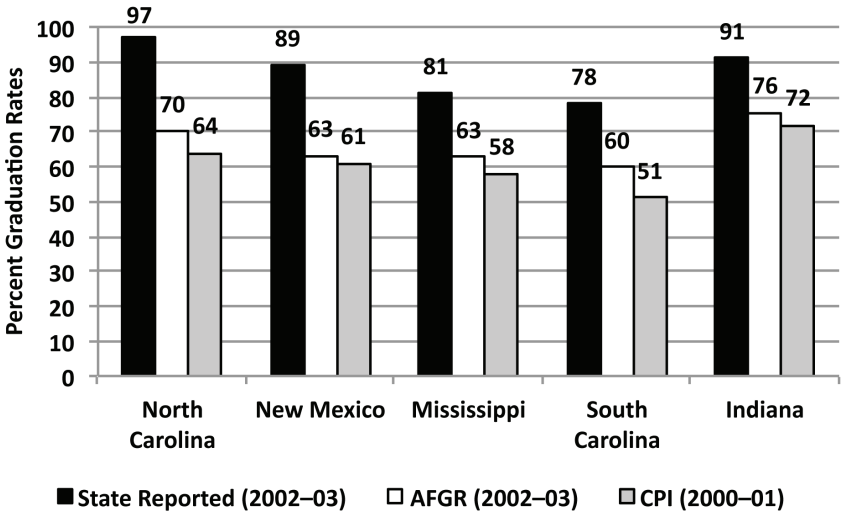

Figure 3. Comparison of state-reported, AFGR, and CPI graduation rates.

CPI = the Cumulative Promotion Index calculation was for the previous year (2000–01), but is still relevant because graduation rates do not change much from year to year. AFGR = the Average Freshman Graduation Rate is generated by the NCES. It calculates the number of regular diplomas issued in a given year divided by the average enrollment base for the freshman class 4 years earlier. Data are drawn from Hall (2005, p. 5) and Synder and Dillow (2012a).

The difference is dramatic. In school year 2002–03 North Carolina reported that 97% of its students graduated, whereas the more accurate calculations placed the number at between 64% and 70%. New Mexico reported a graduation rate of 89%, compared to more accurate figures in the low 60% range. The pattern was repeated for most states. For many states, these discrepancies were even greater when data were disaggregated by race. North Carolina reported a graduation rate for African-American students of 95% versus CPI's calculation of 54%; a graduation rate for Latino students of 94% versus CPI's calculation of 58%; and a graduation rate for Native American students of 96% versus CPI's calculation of 34% (Hall, 2005).

If you don't know where you are going…

The balance of this chapter documents our nation's significant investment in education and its failure to produce desired outcomes in virtually all the identified benchmark categories. Certainly, there are many reasons for this failure. However, it is hard to imagine one more important than our history of "flying blind." Given the absence of systemic education performance metrics that are based on consistent measures and standards, it is no wonder that education reform has foundered for almost 40 years. We simply never had meaningful performance outcome data to know how we were doing or what was working.

BENCHMARKING PERFORMANCE OUTCOMES IN EDUCATION

Despite the challenges of implementing reliable, valid, and socially relevant performance metrics at state and local levels, the performance of our overall national education system can be benchmarked over time and against those of other industrialized nations. The first question is that of social relevancy. What are the crucial performance outcomes for an education system? In benchmarking the education performance data from its participating countries, OECD identified the following four critical outcomes for a high-performing education system:

High participation: Almost all the system's students are in high school at the appropriate age and complete the requisite course work for a diploma/ degree.

High quality: The system's average student performance is high using well-established national standards as well as international standards. The education system continually makes significant progress in improving student performance.

High equity: The education system delivers high-quality learning consistently to all students in all schools so that every student benefits from excellent academic opportunities. Education resources are equally distributed across schools regardless of students' ethnicity or socio-economic status.

High efficiency: Academic achievement is high relative to per-pupil spending (return on investment).

The following analyses use available macro data to benchmark the performance outcomes of the U.S. education system in the areas of participation, quality, equity, and efficiency.

High participation benchmark

Children need to attend school if they have any hope of benefiting from school. They also need to complete high school (referred to as upper secondary school by most nations). This section examines the comparative performance data on the percentage of students participating in education at both later ages (15 to 19) and early ages (3 and 4). It also reviews the data on those who complete high school.

Participation rates

OECD tracks data on the percentage of children enrolled in education by age group. Two particular age groups are highlighted in this analysis: (a) 15- to 19-year-olds and (b) 3- and 4-year olds. Tracking students 15 to 19 years of age is another way of assessing an education system's success in serving students through completion of high school. Table 3 displays data on 30 OECD member and participating nations for 2010. The United States ranked 24th in the percentage of 15- to 19-year-olds enrolled in school.

Table 2.
OECD enrollment rates for ages 15 to 19 in 2010

Rank	Nation	% Enrolled	Rank	Nation	% Enrolled
1	Ireland	95.7	16	South Korea	85.9
2	Belgium	93.3	17	Slovak Republic	85.3
3	Poland	92.7	18	Switzerland	85.1
4	Slovenia	91.8	19	Denmark	85.0
5	Hungary	91.7	20	Spain	84.3
6	Netherlands	90.7	21	France	84.2
7	Czech Republic	90.2	22	Greece	83.4
8	Germany	89.5	23	Italy	83.3
8	Iceland	87.8	**24**	**United States**	**81.7**
10	Saudi Arabia	87.1	25	Australia	81.4
11	Finland	86.8	26	Canada	80.8
12	Estonia	86.5	27	New Zealand	79.1
13	Portugal	86.4	28	Austria	78.4
14	Sweden	86.4	29	United Kingdom	77.4
15	Norway	86.3	30	Luxembourg	76.7

Note: Canada's data are from 2009. Luxembourg's data are underestimated because many resident students go to school in neighboring countries.

Adapted from *Education at a Glance 2012: OECD Indicators* (p. 330), by the Organisation for Economic Co-operation and Development, 2012, Paris: Organisation for Economic Co-operation and Develoment. Copyright 2012 by Organisation for Economic Co-operation and Development.

The other participation age range analyzed is that of ages 3 and 4. According to OECD, "Early childhood education is associated with better performance later on in school. Fifteen-year-old pupils who attended pre-primary education perform better on PISA than those who did not, even after accounting for their socio-economic backgrounds" (OECD, 2012a). Full enrollment in education (defined by OECD as enrollment rate exceeding 90%) begins between the ages of 3 and 4 in half of OECD countries. Table 4 displays data for 30 OECD member and participating nations in 2010. The United States ranked 26th in the percentage of 3- and 4-year-olds enrolled in school that year.

Table 3.
OECD enrollment rates for ages 3 and 4 in 2010

Rank	Nation	% Enrolled	Rank	Nation	% Enrolled
1	France	100.0	16	Israel	82.6
2	Spain	99.0	17	Hungary	82.2
3	Belgium	98.9	18	South Korea	80.2
4	Iceland	95.8	19	Portugal	79.5
5	Norway	95.7	20	Austria	75.1
6	Italy	94.8	21	Czech Republic	72.5
7	Germany	92.4	22	Russian Federation	71.1
8	Denmark	92.3	23	Mexico	69.4
9	Sweden	92.0	24	Ireland	66.9
10	New Zealand	90.5	25	Slovak Republic	66.5
11	United Kingdom	90.0	**26**	**United States**	**59.9**
12	Estonia	89.2	27	Chile	56.5
13	Japan	86.1	28	Argentina	55.3
14	Luxembourg	84.5	29	Poland	52.5
15	Slovenia	83.7	30	Finland	51.7

Note: Canada's data are from 2009. Luxembourg's data are underestimated because many resident students go to school in neighboring countries.

Adapted from *Education at a Glance 2012: OECD Indicators* (p. 330), by the Organisation for Economic Co-operation and Development, 2012, Paris: Organisation for Economic Co-operation and Development. Copyright 2012 by Organisation for Economic Co-operation and Development.

Graduation rates

One of the best sources of data for measuring student participation in the U.S. education system is the percentage of students who graduate from high school each year. The Average Freshman Graduation Rate (AFGR) data paint a grim picture in this benchmark category. In the 2008–09 school year, almost 25% of all freshman students (one in four) starting high school in 2004–05 nationwide failed to complete high school graduation requirements. This translated to 1 million students failing to earn a diploma in 2009. In addition to documenting extremely poor performance, the data show very little improvement over the last 18 years (Figure 4).

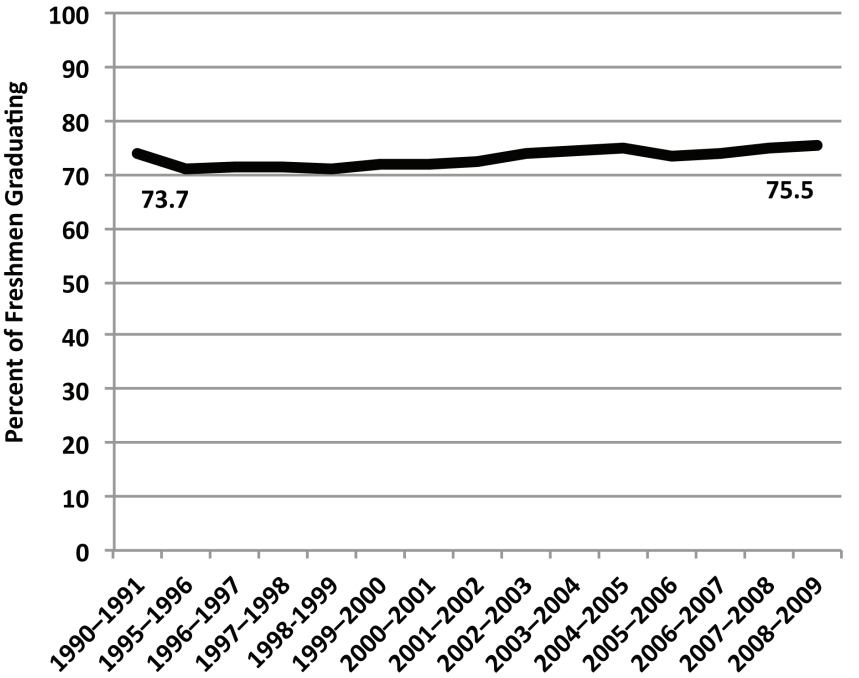

Figure 4. Percent of freshman graduating from public high schools in all states and the District of Columbia, 1991–2009.

Data are drawn from Snyder and Dillow (2012a).

As will be seen repeatedly in this chapter, performance data vary dramatically from state to state. Graduation rates are a prime example. In 2008–09, AFGR in individual states ranged from 56.3% in Nevada and 62% in Mississippi to 89.6% in Vermont and 90.7% in Wisconsin (Stillwell et al., 2011).

At the international level, OECD provides data that allow for the comparison of upper secondary (equivalent to high school in the United States) graduation rates across nations. Table 4 displays data for 26 of OECD's 34 member nations in 2010. The United States ranked 22nd in graduation rate.

Table 4.
OECD upper secondary (high school) graduation rates, 2010

Rank	Nation	% Grad. Rate	Rank	Nation	% Grad. Rate
1	Portugal	104.0	14	Slovak Republic	85.6
2	Japan	95.6	15	Hungary	85.5
3	Greece	94.1	16	Poland	83.5
4	South Korea	93.9	17	Chile	83.3
5	Slovenia	93.8	18	Italy	83.2
6	Ireland	93.7	19	Canada	80.5
7	Finland	93.3	20	Spain	80.4
8	Israel	91.8	21	Czech Republic	79.2
8	United Kingdom	91.6	**22**	**United States**	**76.8**
10	Iceland	87.8	23	Sweden	74.8
11	Norway	87.2	24	Luxembourg	69.7
12	Germany	86.5	25	Turkey	54.2
13	Denmark	86.2	26	Mexico	47.0

Note: 1. Portugal's 104% graduation rate is an exceptional and temporary situation following the implementation of the "New Opportunities" initiative in that country. Many individuals went back to school and have now graduated from this program. 2. Canada's data are from 2009.

Adapted from *Education at a Glance 2012: OECD Indicators* (p. 53), by the Organisation for Economic Co-operation and Development, 2012, Paris: Organisation for Economic Co-operation and Development. Copyright 2012 by Organisation for Economic Co-operation and Development.

The graduation rates of participating nations increased by an average of 8 percentage points since 1995 (OECD, 2012a). During that same period, the United States' graduation rate increased by only 4.5 percentage points. (Snyder & Dillow, 2012a)

High participation benchmark review

The macro data make it clear that the U.S. education system scores very poorly in the high participation benchmark. The preponderance of evidence is overwhelming:

- When compared with other developed nations, in 2010 the United States ranked 24th in enrollment of 15- to 19-year-olds, 26th in enrollment of 3- and 4-year-olds, and 22nd in high school graduation rate.
- Twenty-five percent of U.S. students do not graduate from high school.
- The variability in graduation rates among states is dramatic, ranging from 56.3% to 90.7%.
- There has been little or no progress in this metric for as far back as reliable data go.

High quality benchmark

As discussed previously, standardized testing represents one of best, and perhaps only, empirical quality indicators for measuring student academic performance at the macro level. The NAEP and PISA tests provide data across a range of subjects and ages. This benchmark analysis focuses specifically on reading and mathematics.

Student performance data (NAEP)

The richest set of student achievement data come from NAEP, which provides data on subject matter achievement in two ways: scale scores (long-term trend assessment) and achievement levels (main NAEP assessment). The long-term trend assessment makes available test data in mathematics and reading going back to 1970, with test scores by age (9, 13, and 17). The main NAEP assessment reports test results on 12 different subject areas going back to 1992, with student data by grade (4, 8, and 12).

Scale scores provide a numeric summary of what students know and can do in a particular subject and are presented for groups of students. NAEP scale scores for reading and math range from 0 to 500. Figures 5 and 6 display NAEP scale scores from 1971 through 2008 for reading and 1978 through 2008 for mathematics. They show a remarkable lack of student progress in reading and mathematics over the last 40 years. This "flat line" performance occurred despite numerous and significant school reform initiatives (*A Nation at Risk*, Goals 2000, NCLB).

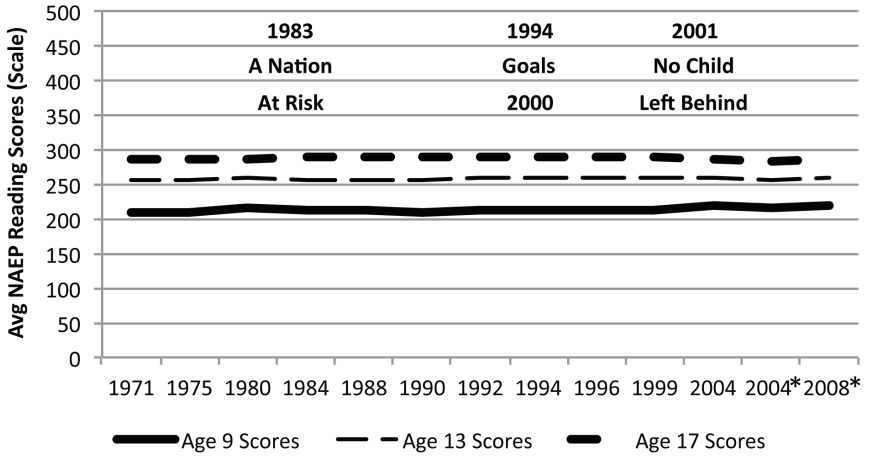

Figure 5. NAEP reading scores, long-term trend assessment, 1971–2008.

Data are drawn from National Center for Education Statistics (NCES) (2011a). *Test formats were changed in 2004. Both old and new test formats were reported for that year. Year 2008 used the new format.

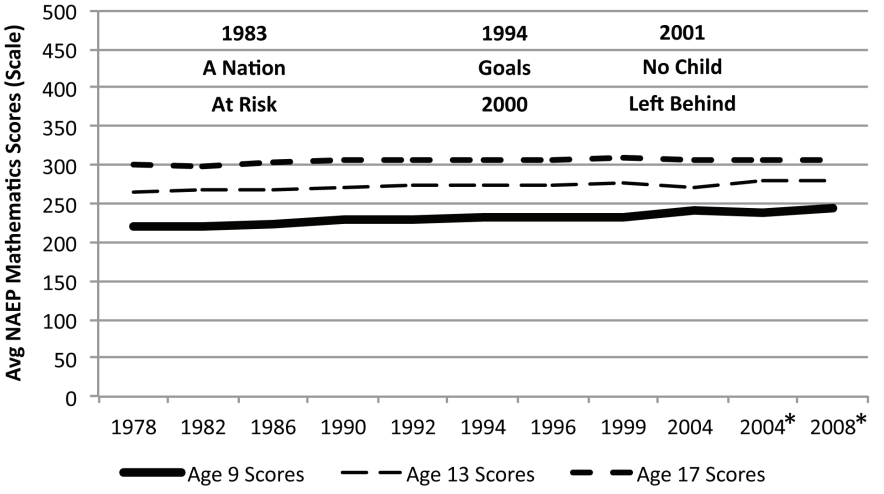

Figure 6. NAEP mathematics scores, long-term trend assessment, 1978–2008.

Data are drawn from National Center for Education Statistics (NCES) (2011a). *Test formats were changed in 2004. Both old and new test formats were reported for that year. Year 2008 used the new format.

Flat scale scores would be acceptable, and even desirable, if the scores reflected high levels of proficiency in the subject matter, but this was not the case. NAEP achievement data can be analyzed to identify the percentage of students at a given grade level who were at or above proficiency. Again, "proficiency" means that students at this level have demonstrated competency over challenging subject matter for their grade level. "Below proficiency" means that students have only partial mastery. Figure 7 shows the percentage of fourth-grade children who could read at or above proficiency level from 1992 through 2011.

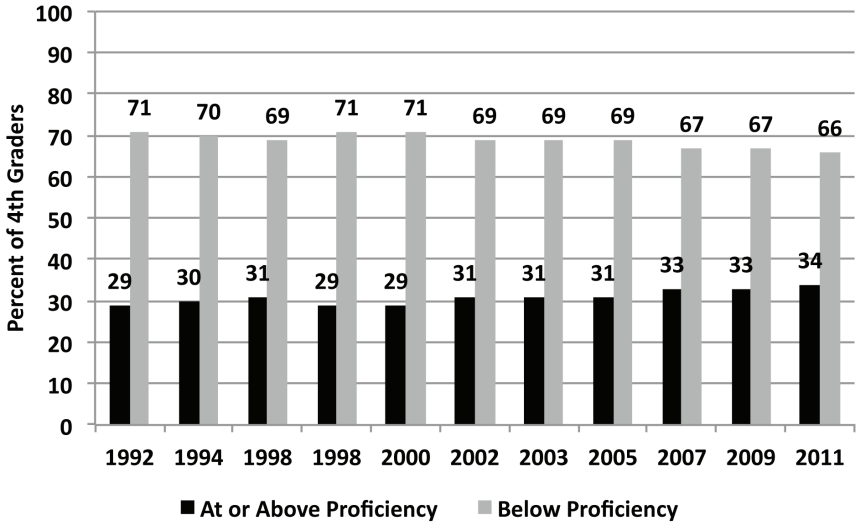

Figure 7. Percent of fourth graders reading at or above proficiency, 1992–2011.

Adapted from *The Nation's Report Card: Reading 2011* (p. 10), by the National Center for Education Statistics, 2011, Washington, DC: U.S. Department of Education. In the public domain.

In 2011, only one third of fourth-grade students read at or above proficiency level, which represents only a 5 percentage point improvement since 1992. This is particularly problematic as research tells us that children who fall significantly behind in reading at an early age have a very small chance of making up the difference (OECD, 2012a). Fourth-grade reading proficiency

data varied significantly across states, with New Mexico and Mississippi having the lowest percentages of proficient readers at 20% and 22%, respectively, in 2011. The state with the greatest percentage of proficient readers was Massachusetts, with 51% (National Center for Education Statistics [NCES], 2011e).

The data did not improve significantly when it came to the percentage of 12th-grade students who read at or above proficiency (Figure 8).

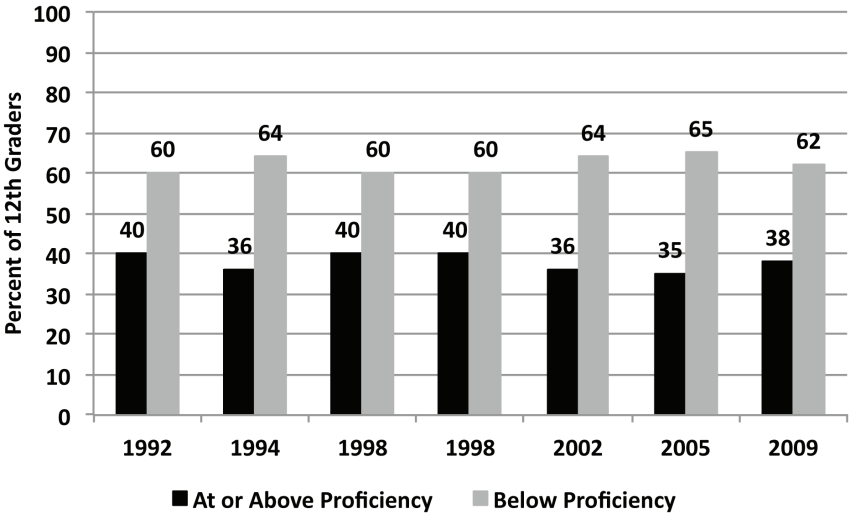

Figure 8. Percent of 12th graders reading at or above proficiency, 1992–2009.

Adapted from *The Nation's Report Card: Grade 12 Reading and Mathematics 2009 National and Pilot State Results* (p. 9), by the National Center for Education Statistics, 2010, Washington, DC: U.S. Department of Education. In the public domain.

Only 38% of 12th-grade students were reading at or above proficiency in 2009, a decrease in performance from 1992, when 40% were reading at that level. While 12th-grade achievement data historically has not been collected at the state level, 11 states volunteered to participate in a pilot program in which their test scores were reported separately (National Center for Education

Statistics [NCES], 2010b). Once again, individual states had widely differing performances. West Virginia (29%), Arkansas (30%), and Florida (32%) scored the lowest percentages of readers at or above proficiency in grade 12, and New Hampshire (44%) and Massachusetts (46%) scored the highest percentages (NCES, 2010b).

NAEP achievement levels in mathematics painted a very similar picture. While a significant improvement in test scores for fourth graders occurred between 2000 and 2007, there was little change subsequently, with performance leveling out at 39% to 40% proficiency (Figure 9).

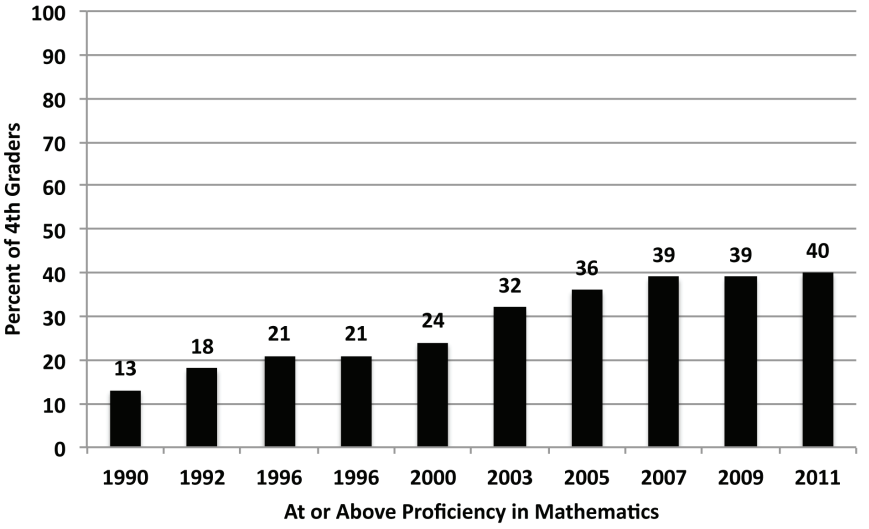

Figure 9. Percent of fourth graders at or above proficiency in mathematics, 1990–2011.

Adapted from *The Nation's Report Card: Mathematics 2011* (p. 11), by the National Center for Education Statistics, 2010, Washington, DC: U.S. Department of Education. In the public domain.

Mathematics achievement data for 12th-grade students is available only for 2005 and 2009, as a change in the mathematics framework for the assessment necessitated a new trend line at that grade level. Only 23% percent of 12th graders performed at or above proficiency in 2005, and 26% in 2009 (NCES, 2010b). As with reading, the only individual state data came from the 11 state pilot programs in 2009. The performance of individual states varied widely, just as they did in reading achievement. West Virginia (13%) and Arkansas (15%) had the lowest percentage of 12th-grade students at or above proficiency in mathematics, while New Hampshire (32%) and Massachusetts (36%) had the highest (NCES, 2010b).

Student performance data (PISA)

PISA test results are a second source of student performance outcome test data. In 2009, they showed the United States trailing 13 nations in reading, 16 nations in science, and 24 nations in mathematics (Table 5). PISA now has test scores over 10 years that highlight changes in performance. The United States' reading test scores actually dropped by 5 points between the 2000 and 2009 PISA tests while its science scores improved by 5 points between 2003 and 2009, but neither change was statistically significant. Its science scores increased by 13 points between 2006 and 2009, which was considered statistically significant (Organisation for Economic Co-operation and Development [OECD], 2010a).

Table 5.
2009 PISA reading, science, and mathematics scores

Rank	Reading		Science		Mathematics	
1	South Korea	539	Finland	554	South Korea	546
2	Finland	536	Japan	539	Finland	541
3	Canada	524	South Korea	538	Switzerland	534
4	New Zealand	521	New Zealand	532	Japan	529
5	Japan	520	Canada	529	Canada	527
6	Australia	515	Estonia	528	Netherlands	526
7	Netherlands	508	Australia	527	New Zealand	519
8	Belgium	506	Netherlands	522	Belgium	515
9	Norway	503	Germany	520	Australia	514
10	Estonia	501	Switzerland	517	Germany	513
11	Switzerland	501	United Kingdom	514	Estonia	512
12	Poland	500	Solvenia	512	Iceland	507
13	Iceland	500	Poland	508	Denmark	503
14	**United States**	**500**	Ireland	508	Slovenia	501
15			Belgium	507	Norway	498
16			Hungary	503	France	497
17			**United States**	**502**	Slovak Republic	497
18					Austria	496
19					Poland	495
20					Sweden	494
21					Czech Republic	493
22					United Kingdom	492
23					Hungary	490
24					Luxembourg	489
25					**United States**	**487**

Adapted from *PISA 2009 Results: What Students Know and Can Do–Student Performance in Reading, Mathematics and Science (Volume I)* (p. 15), by the Organisation for Economic Co-operation and Development, 2010, Paris: Organisation for Economic Co-operation and Development. Copyright 2010 by Organisation for Economic Co-operation and Development.

High quality benchmark review

While there is much debate about the components of a quality education system, the acquisition of core reading and mathematics skills is perhaps the most important. If students are not gaining proficiency in critical academic skills, nothing else much matters. Although standardized test scores have their detractors, they do measure this critical benchmark. The preponderance of evidence is overwhelming:

- NAEP scores for both reading and mathematics have not improved since the inception of the tests in the late 1960s, despite significant education reform efforts.
- NAEP achievement scores identifying proficiency levels in reading and mathematics have also shown very little to no improvement since their inception in the early 1990s, with the exception of a jump of 16 points in fourth-grade math scores between 2000 and 2011 (from 24% at or above proficiency to 40%).
- The most recent NAEP reading achievement test scores showed that only 34% of 4th-grade students (2011) and 38% of 12th-grade students (2009) were at or above proficiency level in reading.
- The most recent NAEP mathematics achievement test scores revealed that only 40% of 4th-grade students (2011) and 26% of 12th-grade (2009) students were at or above proficiency levels in mathematics.
- The United States ranked 14th in reading and 25th in mathematics among OECD nations on PISA test scores in 2009.

The results from our two most reliable and valid tests—NAEP and PISA standardized tests—overwhelmingly support the premise that the United States is failing the high quality benchmark.

High equity benchmark

The high equity benchmark can be analyzed in two areas: (a) the relationship between learning outcomes and students' ethnicity/socio-economic background and (b) whether or not there is equal access to resources (quantity and quality) in all schools regardless of ethnicity/socio-economic conditions.

Equity in learning outcomes

ETHNICITY

Benchmark data suggest that in the U.S. education system there is significant inequality in learning outcomes for children of particular ethnicities (Black, Hispanic, American Indian) and from lower socio-economic backgrounds. This

conclusion is reinforced by multiple measures: graduation rates, dropout rates, and NAEP test scores.

The AFGR disaggregates high school graduation rates by various ethnicities. Figure 10 highlights the differences in graduation rates.

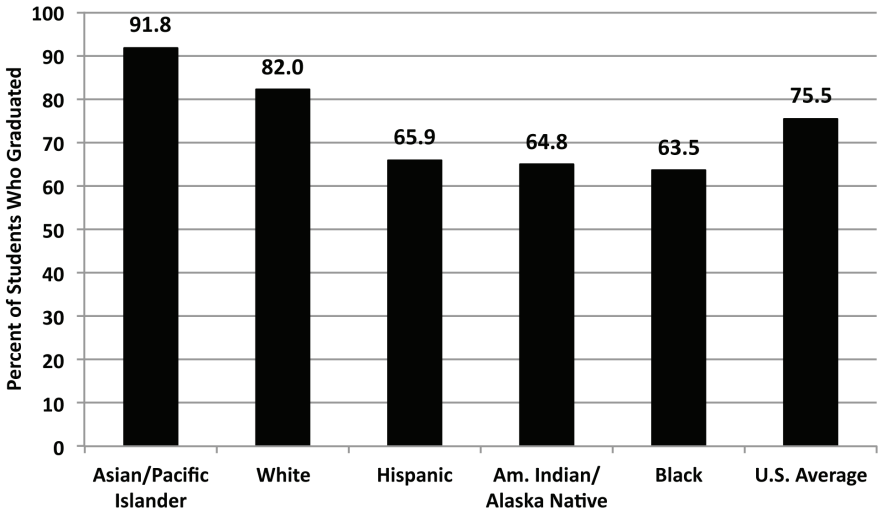

Figure 10. High school graduation rate by ethnicity, 2008–09. Data are drawn from Stillwell, Sable, and Plotts (2011, pp. 8–9).

Graduation rate data show a clear link between learning outcomes and ethnicity. The graduation rate for Black, Hispanic, and American Indian/Alaska Native students averaged 64.8% in 2008-09, 11 percentage points lower than the national average and 17 percentage points lower than the average for White students. As with other data, graduation rates varied significantly from state to state.

- Black students: On a national scale, only 63.5% of Black students graduated from high school. Some of the larger states reported even worse performances: California, 57.7%; New York, 58.1%; Ohio, 56.8%; and Florida, 59.8%.
- Hispanic students: Only 65.9% of Hispanic students graduated nationally. The District of Columbia graduated only 50.1% of its Hispanic students, and several states didn't do much better: Connecticut, 55.5%; Georgia, 56.6%; and New York, 57.4%. New Hampshire fared the worst, graduating only 41.6% of Hispanic students.

- American Indian/Alaska Native students: Students of this ethnicity had an overall graduation rate of 64.8%. Among the worst performing states were Wyoming, 45%; Mississippi, 49.3%; and Washington state, 51.3% (Stillwell et al., 2011).

Another metric that demonstrates inequality in learning outcomes is the number of students who drop out of school. Figure 11 displays the relative dropout rates by ethnicity for the 2008–09 school year. Black, Hispanic, and American Indian/Alaska Native students were more than twice as likely to drop out of school as White students. Once again, there was significant disparity in the outcome data among ethnicities.

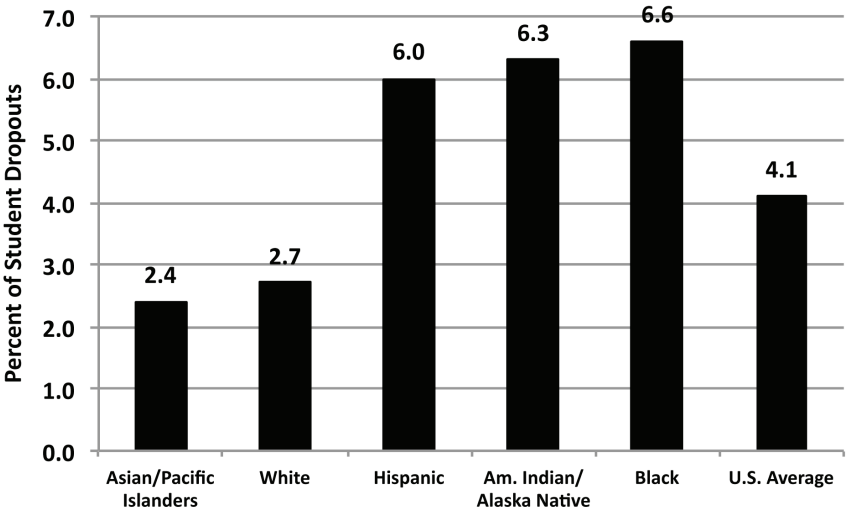

Figure 11. Dropout rates by ethnicity, 2008–09.
Data are drawn from Stillwell, Sable, and Plotts (2011, pp. 16–17).

NAEP achievement test data can also be disaggregated by ethnicity. Figures 12 and 13 track NAEP reading proficiency for 4th- and 12th-grade students by race/ethnicity over a 19-year period.

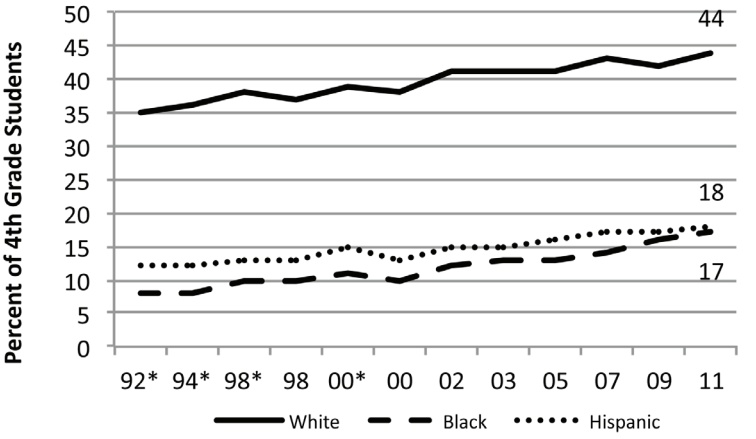

Figure 12. NAEP fourth-grade reading proficiency by race/ethnicity, 1992–2011.

Data are drawn from National Assessment of Educational Progress (NAEP) (2011b). *Accommodations were not permitted for this assessment.

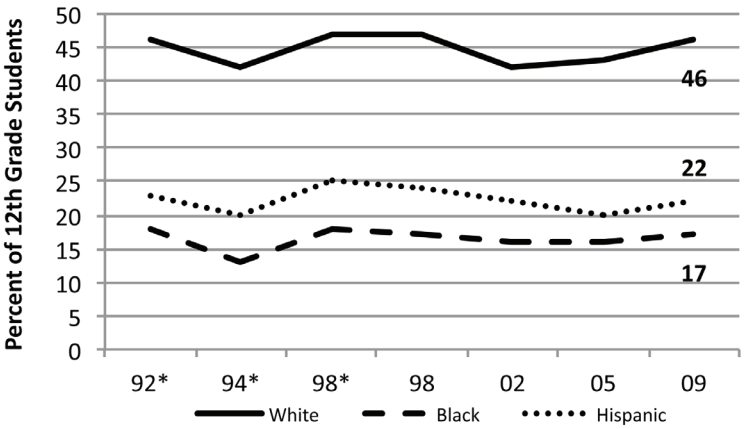

Figure 13. NAEP 12th-grade reading proficiency by race/ethnicity, 1992–2009.

Data are drawn from National Assessment of Educational Progress (NAEP) (2011b). *Accommodations were not permitted for this assessment.

The data show obvious gaps in reading proficiency among ethnicities over the years. For comparison purposes, the most recent observations refer to 2009 data as there were no 2011 statistics for 12th-grade reading.

- In 2009, there was a significant gap in fourth-grade reading proficiency between White (42%) and Black (16%) students (a difference of 26 percentage points), and between White (42%) and Hispanic (17%) students (25 percentage points).
- The gap did not narrow for 12th-grade students in 2009. The reading proficiency gap between White (46%) and Black (17%) students was 29 percentage points, and between White (46%) and Hispanic (22%) students 24 percentage points.
- The gaps have remained virtually constant over the testing years going back to 1992. In 1992, the fourth-grade reading proficiency gap between White (35%) and Black (8%) students was 27 percentage points, and between White (35%) and Hispanic (12%) students 23 percentage points. In 1992, the 12th-grade reading proficiency gap between White (46%) and Black (18%) students was 28 percentage points, and between White (46%) and Hispanic (23%) students 23 percentage points.

Figures 14 and 15 track NAEP mathematics proficiency over a 21-year period by race/ethnicity for fourth graders, and a 4-year period of time for 12th graders (the NAEP mathematics test was changed significantly in 2005 limiting comparisons to earlier test scores).

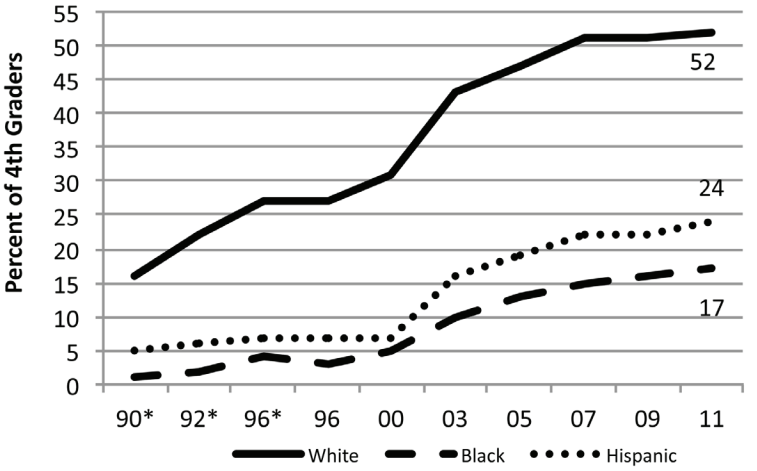

Figure 14. NAEP fourth-grade math proficiency by ethnicity, 1990–2011.

Data are drawn from National Assessment of Educational Progress (NAEP) (2011a). *Accommodations were not permitted for this assessment.

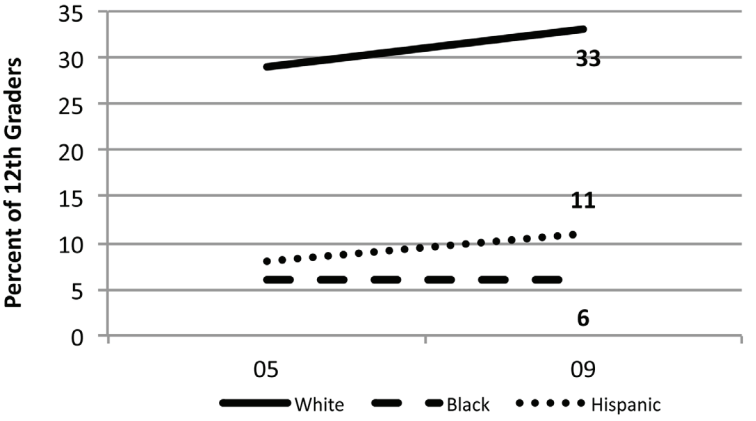

Figure 15. NAEP 12th-grade math proficiency by race/ethnicity, 1992–2009.

Data are drawn from National Assessment of Educational Progress (NAEP) (2011a). *Accommodations were not permitted for this assessment.

The 2009 mathematics proficiency data show performance gaps between ethnicities, and they're even larger than the gaps in reading proficiency data. For comparison purposes, 2009 data is used because, as in the case of 12th-grade reading, there were no 2011 statistics for 12th-grade math.

- There was a significant gap in fourth-grade math proficiency between White (51%) and Black (16%) students (a difference of 35 percentage points), and between White (51%) and Hispanic (22%) students (2 percentage points).
- The gap in 12th-grade math proficiency narrowed slightly between White (33%) and Black (6%) students (a difference of 27 percentage points), and between White (33%) and Hispanic (11%) students (22 percentage points).
- The gaps between ethnicities have widened over time. In 1990, the fourth-grade math proficiency gap between White (16%) and Black (1%) students was 15 percentage points, and between White (16%) and Hispanic (5%) students 11 percentage points. In 2005, the 12th-grade math proficiency gap between White (29%) and Black (6%) students was 23 percentage points, and between White (29%) and Hispanic (8%) students 21 percentage points.

All of the learning outcomes macro-level data (graduation rates, dropout rates, NAEP reading and math scores) overwhelmingly support the conclusion that in the United States there is significant inequity in learning outcomes based on ethnicity.

SOCIO-ECONOMIC STATUS

While graduation rate data are not disaggregated by socio-economic indicators, student dropout data are. The following benchmark indicators (Figures 16 and 17) examine dropout rates in the context of family income, dividing income levels into four quartiles (lowest to highest). These data show an alarming trend in inequity in learning outcomes based on socio-economic background.

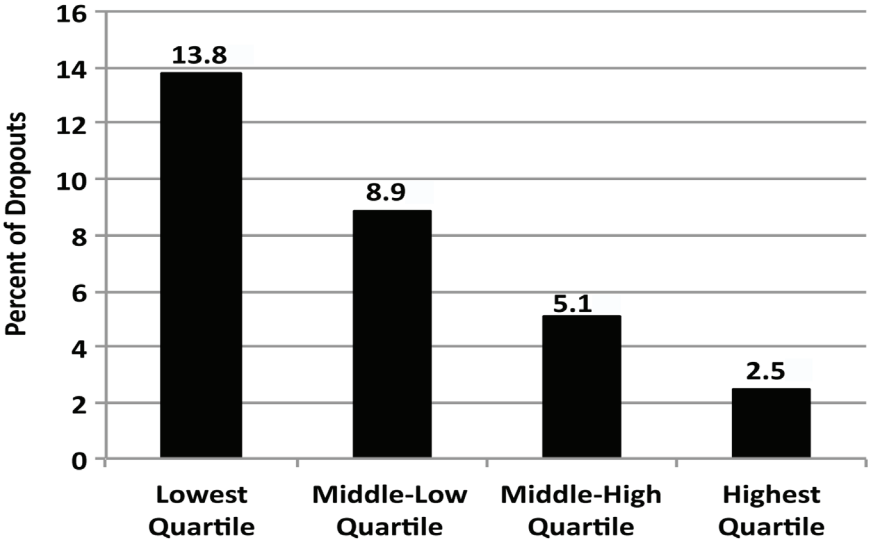

Figure 16. Dropout rate by family income (2010) (lowest quartile to highest quartile).

Data are drawn from Snyder and Dillow (2012b, p. 183).

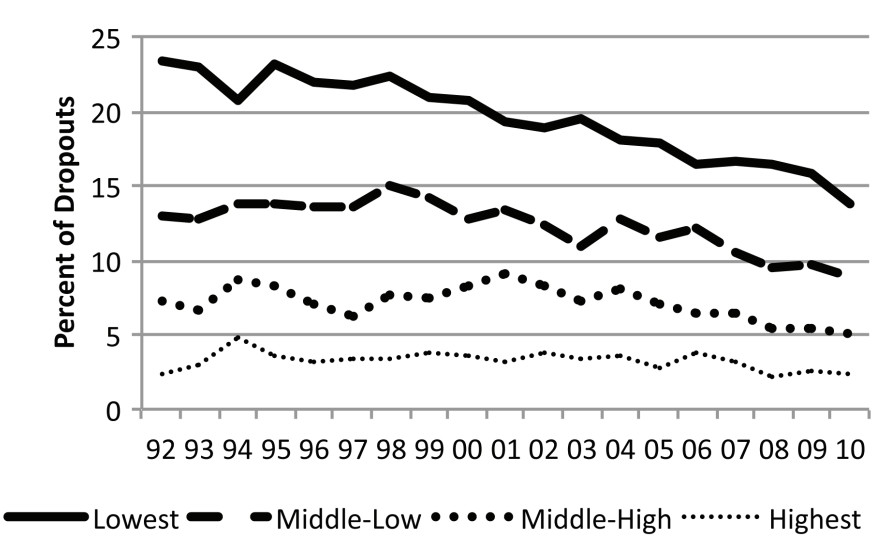

Figure 17. Dropout rate by family income, 1992–2010.

Data are drawn from Snyder and Dillow (2012b, p. 183).

The relationship could not be clearer. Children from families of the lowest socio-economic status are more likely to drop out of high school than children of families with higher incomes, and the relationship follows each quartile proportionately. Children from the lowest income quartile are five times more likely to drop out of school than children from families in the highest income quartile (Figure 16). The trend from 1992 to 2010 (Figure 17) does show a narrowing of the gap, but the linear relationship between family income and dropout rate remains the same.

NAEP achievement scores can also be disaggregated by socio-economic status. One way of identifying a student's socio-economic background is whether or not the student qualifies for the National School Lunch Program (also referred to as the free or reduced-price lunch program, or FRLP). Children may qualify for either free or reduced-price lunch based on family income; free lunch eligibility represents the lowest income families. Figures 18 and 19 chart the reading and mathematics proficiency of fourth-grade students in these three categories: eligible for free lunch, eligible for reduced-price lunch, and not eligible for either.

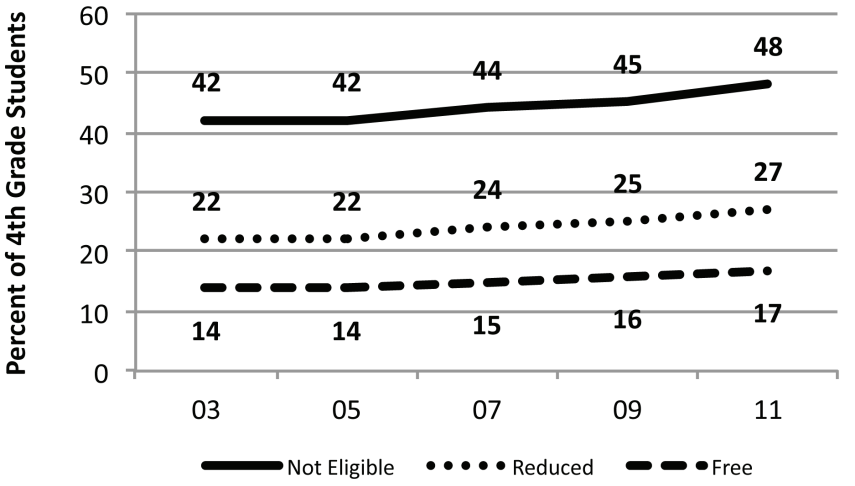

Figure 18. NAEP fourth-grade reading proficiency by eligibility for free and reduced-price lunch program, 2003–11.

Data are drawn from National Assessment of Educational Progress (NAEP) (2011b).

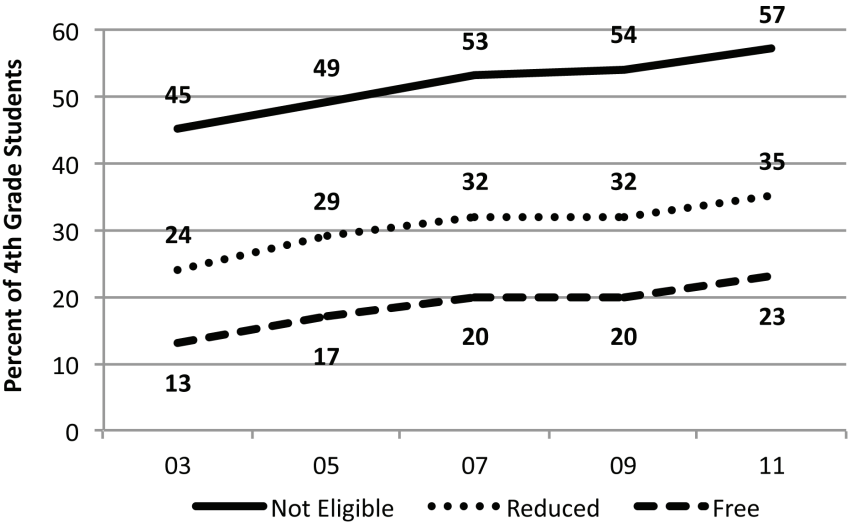

Figure 19. NAEP fourth-grade math proficiency by eligibility for free and reduced-price lunch program, 2003–11.

Data are drawn from National Assessment of Educational Progress (NAEP) (2011a).

The NAEP achievement data show a similar pattern of continuous, significant gaps in reading and mathematics proficiency based on student socio-economic status.

The most recent reading test scores (2011) reported a difference of 21 percentage points between fourth-grade students of the highest socio-economic status and those of the mid-level. The gap between fourth graders of the highest and the lowest socio-economic rankings was 31 percentage points.

Math achievement scores in 2011 showed even larger gaps: 22 percentage points between fourth-grade students of the highest socio-economic status and those of the mid-level, and 34 percentage points between fourth-grade students of the highest and lowest socio-economic rankings.

The gaps between fourth-grade students of different socio-economic statuses in both reading and math proficiency have worsened over time. In 2003, the gap in reading proficiency between students of highest and mid-level socio-economic rankings was 20 percentage points, and between highest and lowest 28 percentage points. In 2003, the gap in math proficiency between highest and mid-level was 21 percentage points, and between highest and lowest 32 percentage points.

The NAEP data demonstrate that a student's socio-economic status remains one of the most significant predictors of academic proficiency.

PISA data can also be analyzed across socio-economic status, using the same metric as the NAEP data: students who qualify for FRPL (free or reduced-price lunch program). Their analysis examined test scores by individual schools and the percentage of students in the school who qualified for FRPL. The lower the percentage of students qualifying for FRPL, the higher the socio-economic status of the student population. Figure 20 examines PISA test scores using this metric.

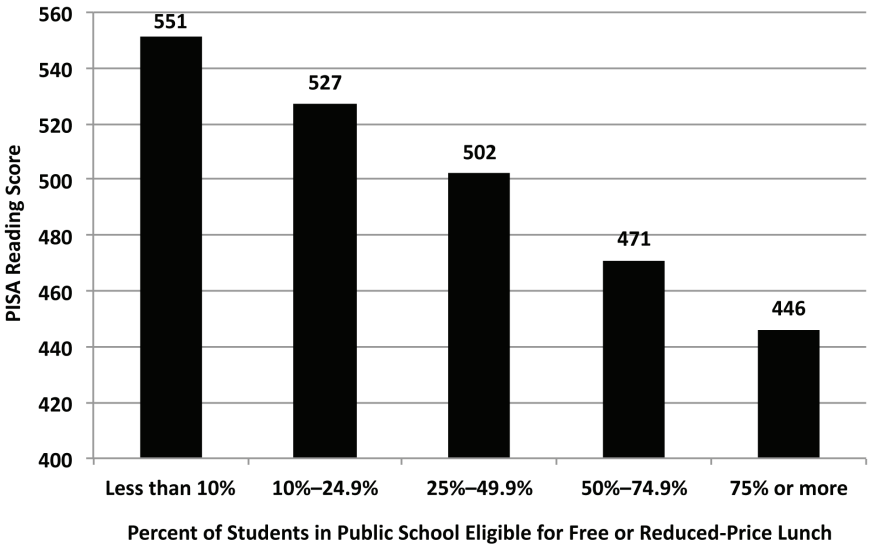

Figure 20. PISA reading scores by socio-economic status, 2009. Data are drawn from Fleischman, Hopstock, Pelczar, and Shelley (2010, p. 15).

The PISA data show the same link between student performance and socio-economic status. Schools with less than 10% of students qualifying for FRPL had average scores that were 105 points higher than schools with 75% or more FRPL students. As with previous data, the consistently linear nature of the relationship between student performance and socio-economic background is remarkable.

PISA conducted detailed analyses of member and participating countries, examining the degree to which student performance in reading was related to socio-economic background.

The first analysis looked at the amount of variance in reading test scores that were attributable to the socio-economic status of individual students. PISA found that 17% of the variance in individual student performance in the United States was attributable to socio-economic background. Using this metric, in 2009 the United States ranked 9th of 34 OECD countries in reading, and 22nd among 64 reporting countries (Organisation for Economic Co-operation and Development [OECD], 2010b).

The second PISA analysis looked at the variance in student reading performance that was attributable to socio-economic background differences between schools. Were these variances spread across schools or clustered

within schools? In other words, was there a disproportionate concentration of lower performing students, who were also of lower socio-economic status at the school level? PISA found that in the United States the between-school student performance variance explained by the socio-economic makeup of schools was nearly 80%. In 2009, the United States ranked 31st of 33 OECD countries and 61st out of 63 reporting countries (OECD, 2010b).

The between-school variance highlights one of the most challenging—and growing—demographic characteristics of public education in the United States: an increasing segregation of students by socio-economic conditions resulting in disadvantaged schools (high-poverty) and advantaged schools (low-poverty).

This trend of increasing segregation of students by socio-economic status is confirmed by National Center for Education Statistics (NCES). Figure 21 describes the percentage of public schools by poverty level based on FRPL data. High-poverty schools are defined as public schools in which more than 75% of the students are eligible for the FRPL program; mid-high poverty, in which 51% to 75% are eligible; mid-low poverty, in which 26% to 50% are eligible; and low-poverty, in which 25% or less are eligible.

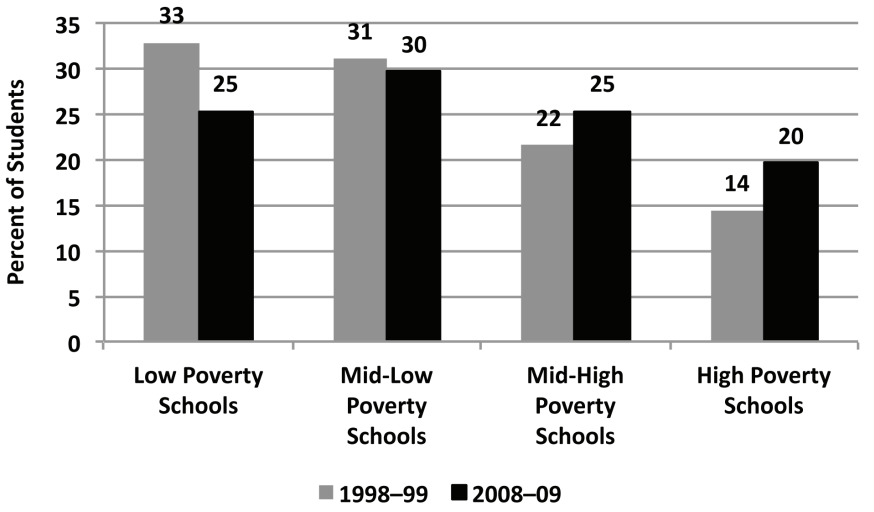

Figure 21. Percent of public schools by student poverty level, 1998–99 and 2008–09.

Data are drawn from Aud, Hussar, Kena, Bianco, Frohlich, Kemp, and Tahan (2011, p. 238).

The data reflect two alarming trends. First, in 2008–09, almost half of the schools were at the mid-high to high poverty levels, and one fifth at high-poverty, or 75% or greater FRPL participation. This spread illustrates a clustering of students from high-poverty families within a school. Second, this trend has increased over a 10-year period. In 1998–99, only 14% of schools were at the high-poverty classification. Over the next 10 years, this figure increased by more than a third, to 20%.

NCES data also show disproportionality in the ethnic makeup of low-poverty versus high-poverty schools. Figures 22 and 23 show the ethnic composition of students attending low-poverty schools versus high-poverty schools.

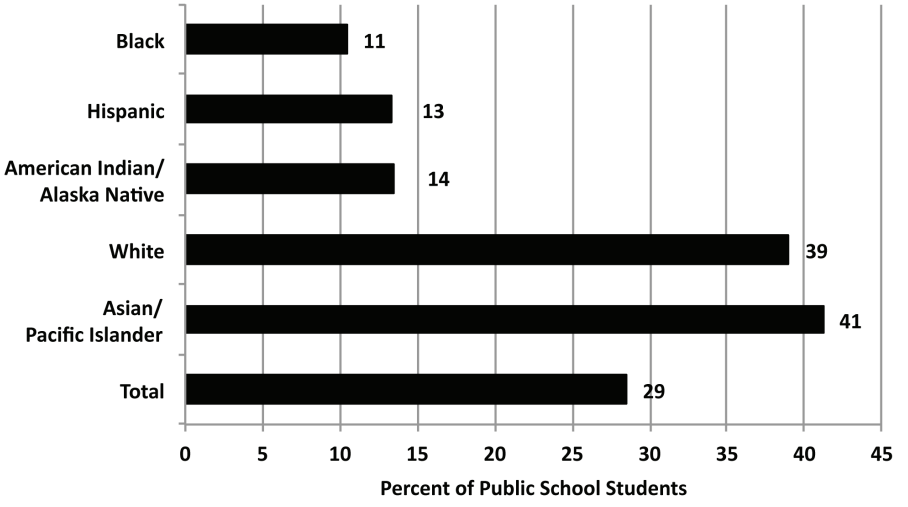

Figure 22. Percent of ethnic groups in student population in low-poverty schools in 2008–09.

Data are drawn from Aud, Hussar, Kena, Bianco, Frohlich, Kemp, and Tahan (2011, p. 240).

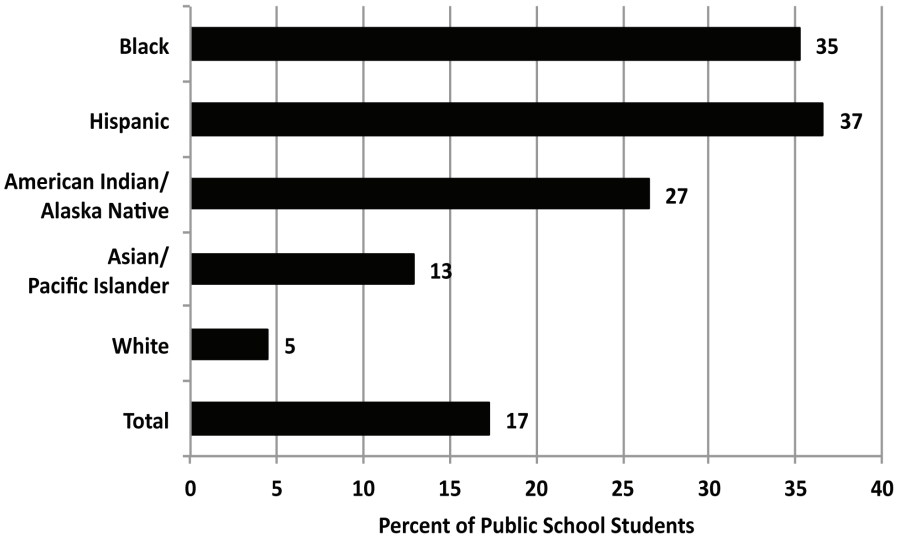

Figure 23. Percent of ethnic groups in student population in high-poverty schools in 2008–09.

Data are drawn from Aud, Hussar, Kena, Bianco, Frohlich, Kemp, and Tahan (2011, p. 240).

High-poverty schools have disproportionately high percentages of Hispanic (37%) and Black (35%) students and a disproportionately low percentage of White students (5%). This feature is reversed when examining low-poverty schools: The percentage of White students is disproportionately high (39%), and the percentages of Hispanic (13)%, and Black (11%) students disproportionately low. A separate study concluded the following: "One in thirty white students and less than a tenth of Asian students, but 40% of black and Latino students attend schools where 70-100% of the children are poor" (Orfield, 2009). The clustering of lower socio-economic non-White students in individual schools (and often school districts) has a significant impact in funding and resource equity, as will be seen in the next section.

Student performance metrics show a clear inequity in learning outcomes related to ethnicity and socio-economic status. As children who are Black, Hispanic, American Indian, or from lower socio-economic families have the same learning potential as all children, there must be discrepancies in learning opportunities and resources to produce these results. The relationship between ethnicity and poverty, and the clustering of high-poverty students in individual schools suggests that the issue of equal access to resources plays a role.

The following discussion begins to answer the question of how inequity in learning outcomes happened.

Equity in access to resources

Effective allocation of education resources is one of the most critical tools an education system has to address the educational needs of society. Two resources stand out as the most critical: funding and high-quality educators. These resources have to be allocated efficiently and—just as important—equitably to address the differing needs of a student population. It is no easy task.

Funding for our nation's K–12 education system is highly decentralized, complicated, and capricious. It is decentralized in that each individual state has the responsibility to establish its own state school finance system with rules, regulations, and policies that establish school-funding formulas. The system is complicated because funding of public schools is divided among federal (8.2%), state (48.3%), and local (43.5%) governments, and each funding source has its own rules and guidelines (Education Finance Statistics Center [EDFIN], 2009). And, as will be seen, the system is capricious because funding results in significant disparities across a wide range of units of analysis—in 2009–10, across 13,629 school districts and 98,817 public schools in 50 states (Snyder & Dillow, 2012b). Rather than dissect the Byzantine labyrinth of these formulas, this section will look at the resulting funding levels using the following benchmarks: (a) funding effort; (b) funding equity across states, school districts, and public schools; and (c) funding progressivity.

FUNDING EFFORT

While research and experience suggest that spending more money on education by itself does not necessarily improve education outcomes (as will be seen later), the level of resources committed to education does matter and is an important benchmark. It is a reflection of the commitment a society makes to education, and, if used wisely, can result in higher student performance. The level of resources metric also allows a system to compare its efficiency and effectiveness with those of other education systems and with itself over time. Annual per-pupil spending (PPS) data can be benchmarked across various system levels (nation, state, school district), as a proportion of a nation or state's wealth (percentage of gross domestic product), and over time.

At the international level, OECD tracks per-pupil spending of its member nations. It calculates expenditures in U.S. dollar equivalencies for meaningful comparisons. OECD analyzes cumulative expenditures over the duration of a student's education, from age 6 to 15 (Table 6).

Table 6.
Cumulative expenditures by educational institutions per student ages 6 to 15 in 32 OECD member nations, 2009

Nation	Expenditures	Nation	Expenditures
Luxembourg	$176,013	Spain	$85,117
Switzerland	$122,797	Finland	$83,774
Norway	$120,349	Japan	$82,857
United States	**$116,268**	France	$81,121
Austria	$115,563	Germany	$75,259
Denmark	$109,017	Korea	$73,854
Iceland	$100,022	New Zealand	$70,090
Netherlands	$94,678	Portugal	$68,931
United Kingdom	$94,583	Estonia	$58,390
Belgium	$93,146	Czech Republic	$55,168
Slovenia	$91,883	Israel	$54,580
Sweden	$91,763	Poland	$52,038
Ireland	$90,743	Slovak Republic	$48,712
Canada	$89,966	Hungary	$46,292
Australia	$89,113	Chile	$29,456
Italy	$88,992	Mexico	$22,688

Adapted from *Education at a Glance 2012: OECD Indicators–Chapter B: Financial and Human Resources Invested in Education–Indicators* (Table B1.3b), by the Organisation for Economic Co-operation and Development, 2012, Paris: Organisation for Economic Co-operation and Development. Copyright 2012 by Organisation for Economic Co-operation and Development.

The data show that the United States spends significantly more on K–12 education than most other OECD countries. In 2009, it spent an average of 40% more than the nations with four of the five next largest economies: approximately 40% more than Japan, 54% more than Germany, 43% more than France, and 23% more than the United Kingdom (no data on education expenditures were reported for China). It also spent significantly more than nations with much higher performing education systems: approximately 29% more than Canada, 57% more than South Korea, 66% more than New Zealand, and 39% more than Finland.

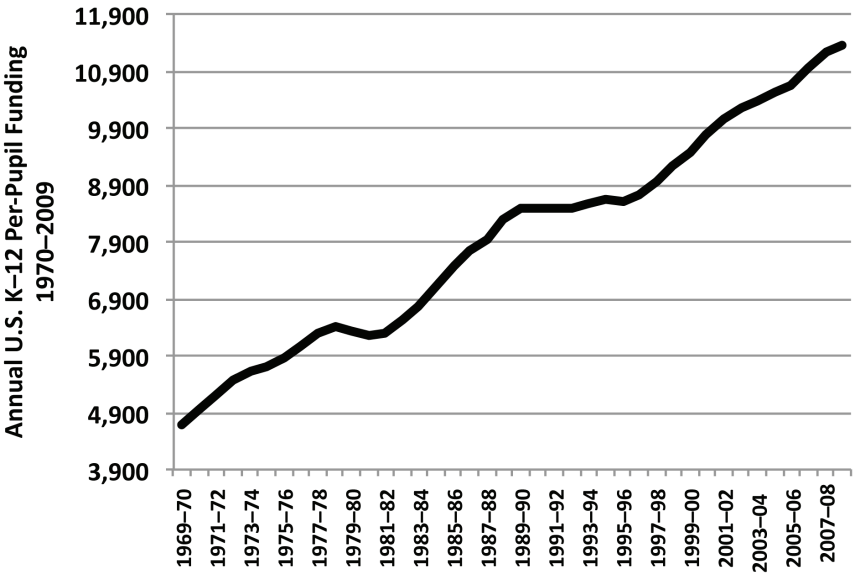

Figure 24. Expenditures per pupil in public elementary and secondary schools, 1970–2009. Adjusted for inflation (2009–10 dollars).

Data are drawn from Snyder and Dillow (2012b, p. 272).

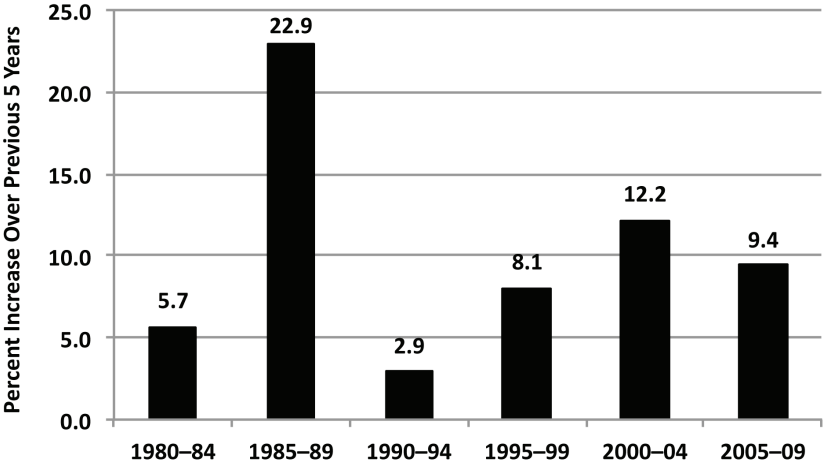

Figure 25. Percent of increase in per-pupil spending over the previous 5 year period (1980–2009). Adjusted for inflation (2009–10 dollars).

Data are drawn from Snyder and Dillow (2012b, p. 272).

When benchmarked against itself over time, expenditure data show that the United States has been steadily increasing the amount of money it spends on K–12 education. Figure 24 shows the trend over the past 40 years. Figure 25 shows the percentage of increase in 5-year increments.

The United States has seen periods of dramatic increases in "real spending" (increases in spending over annual inflation) for K–12 education. These have often been driven by various calls to action and reform initiatives. The 5 years following the publication of *A Nation at Risk*, 1985 through 1989 saw spending increase nearly 23% (Gardner et al., 1983). The 15-year period from 1995 through 2009 saw an increase of nearly 30%, much of it coinciding with the No Child Left Behind initiative. In terms of total spending, the United States has clearly committed an ever-increasing amount of financial resources to education.

Another benchmark metric for evaluating the funding effort is to look at the percentage of gross domestic product (GDP) spent on education. GDP is an accepted measure of a nation's wealth or standard of living. It represents the market value of the goods and services produced within a country in a given period. Table 7 examines the percentage of GDP spent on K–12 education by OECD nations in 2009.

Table 7.
Expenditure on educational institutions as a percent of GDP, 2009

Nation	% GDP Spent on Education	Nation	% GDP Spent on Education
Iceland	5.16	Israel	4.03
New Zealand	5.16	Mexico	3.99
Denmark	4.77	Portugal	3.98
Ireland	4.66	Slovenia	3.96
Korea	4.66	Austria	3.86
United Kingdom	4.48	Chile	3.65
Belgium	4.44	Poland	3.64
Switzerland	4.39	Canada	3.63
United States	**4.27**	Italy	3.39
Sweden	4.24	Luxembourg	3.33
Australia	4.22	Spain	3.32
Estonia	4.17	Germany	3.31
Norway	4.16	Slovak Republic	3.07
Netherlands	4.14	Japan	2.99
Finland	4.10	Hungary	2.95
France	4.07	Czech Republic	2.90

Adapted from *Education at a Glance 2012: OECD Indicators* (p. 244), by the Organisation for Economic Co-operation and Development, 2012, Paris: Organisation for Economic Co-operation and Development. Copyright 2012 by Organisation for Economic Co-operation and Development.

Although the United States was not ranked quite as high as it was in cumulative expenditures (Table 6), it still spent a higher percentage of GDP on education than many other OECD nations in 2009. However, the percentage of GDP spent by the United States is an accumulation of individual state spending, which presents a different picture.

The Education Law Center generates an annual report on school funding in which it takes the level of analysis to the state level, calculating the percentage of each state's GDP allocated to education. Its results for 2009 are summarized in Table 8.

Table 8.
Expenditure on education by state as a function of state GDP, 2009

State	% GDP Spent on Education	State	% GDP Spent on Education	State	% GDP Spent on Education
Vermont	5.7	Arkansas	4.1	Hawaii	3.5
New Jersey	5.0	Wisconsin	4.1	Maine	3.5
New York	4.9	Alaska	4.0	Utah	3.3
New Hampshire	4.5	Mississippi	3.9	Florida	3.3
Indiana	4.5	Montana	3.9	Nevada	3.2
West Virginia	4.4	Kentucky	3.9	Louisiana	3.2
Maryland	4.4	Iowa	3.9	Oklahoma	3.1
South Carolina	4.4	Alabama	3.8	Washington	3.1
Michigan	4.3	Texas	3.8	California	3.1
New Mexico	4.3	Massachusetts	3.7	Colorado	3.1
Ohio	4.2	Illinois	3.7	Oregon	3.1
Kansas	4.2	Idaho	3.6	Arizona	3.0
Pennsylvania	4.2	Nebraska	3.6	Tennessee	3.0
Wyoming	4.2	Minnesota	3.6	North Dakota	2.9
Rhode Island	4.1	Missouri	3.5	South Dakota	2.6
Georgia	4.1	Virginia	3.5	Delaware	2.5
Connecticut	4.1	North Carolina	3.5		

Adapted from *Is School Funding Fair? A National Report Card* (p. 22), by B. Baker, D. Sciarra, and D. Farrie, 2012, Newark, NJ: Education Law Center. Copyright 2012 by Education Law Center.

The analysis shows great disparity among states in terms of funding effort. Vermont (5.7% of GDP) spends twice as much of its GDP as either Delaware or South Dakota. The average of the top 10 states is 4.64%, which is 57% higher than the average of the bottom 10 at 2.95%. These results are discussed further in the next section.

Overall, from a total funding effort perspective, the data suggest that the United States demonstrates a high funding effort in the amount it spends on education. However, this effort is not equal across individual states.

FUNDING EQUITY ACROSS STATES, SCHOOL DISTRICTS, AND INDIVIDUAL SCHOOLS

This benchmark analyzes annual per-pupil spending equity for K–12 education at three different organizational levels: state, school district, and individual school. At the state level, the most recent data from NCES, for the 2008–09 school year, showed tremendous disparity in funding, ranging from states spending large amounts annually per pupil—for example, New York ($19,212) and New Jersey ($18,367)—to states spending considerably less—notably, Idaho ($8,601) and Utah ($8,446). The per-pupil spending for the District of Columbia ($26,753) was significantly higher than for any state (Snyder & Dillow, 2012b).

However, individual states and jurisdictions such as the District of Columbia have different economic conditions that can make comparisons difficult. An analysis by the Education Law Center calculated an adjusted per-pupil spending level that took into account factors beyond a state or jurisdiction's control, such as student poverty, regional wage variation, economies of scale, and population density. The results are shown in Table 9.

Table 9.
State K–12 annual per-pupil spending adjusted for regional factors, 2009

State	Adjusted PPS	State	Adjusted PPS	State	Adjusted PPS
Wyoming	$19,520	Kansas	$11,060	Colorado	$9,198
Alaska	$17,967	Wisconsin	$10,807	Missouri	$9,163
New York	$17,375	Iowa	$10,764	Oregon	$9,129
New Jersey	$16,817	Ohio	$10,625	Nevada	$9,094
Connecticut	$15,693	Virginia	$10,621	Alabama	$9,071
Vermont	$15,020	Nebraska	$10,404	Florida	$8,975
Dist. of Columbia	$14,596	Louisiana	$10,289	Kentucky	$8,930
Massachusetts	$14,091	New Mexico	$10,113	California	$8,897
Maryland	$13,505	West Virginia	$9,995	Texas	$8,862
Rhode Island	$13,047	Illinois	$9,841	Arkansas	$8,808
Delaware	$13,031	North Carolina	$9,754	South Dakota	$8,575
Pennsylvania	$12,976	Washington	$9,686	Mississippi	$7,930
Hawaii	$12,445	South Carolina	$9,657	Arizona	$7,899
New Hampshire	$12,206	Michigan	$9,611	Idaho	$7,509
Maine	$12,125	North Dakota	$9,542	Oklahoma	$7,449
Minnesota	$11,533	Georgia	$9,458	Utah	$7,379
Indiana	$11,065	Montana	$9,300	Tennessee	$7,306

Adapted from *Is School Funding Fair? A National Report Card* (p. 12), by B. Baker, D. Sciarra, and D. Farrie, 2012, Newark, NJ: Education Law Center. Copyright 2012 by Education Law Center.

The average annual per-pupil spending level in the United States in 2009 was $10,774. The range of disparity among states was remarkable, reflecting a significant lack of funding equity. Wyoming spent $12,214 more per year per pupil than Tennessee (a difference of 167%). The five top-spending states averaged $17,474 per pupil spending, or 133% more than the $7,508 average of the bottom five states. As state funding formulas are primarily responsible for how much money is spent, this inequity is of enormous significance.

The funding disparity continues at the individual school district level, even within the same state. In 2010, the Center for American Progress completed an analysis that looked at the academic outcomes of individual school districts by their spending levels (Boser, 2011). As with the previous study, this study was careful to establish criteria that would control for factors outside a district's

control, such as cost of living and students with special needs (English language learners, special education, low income). The following table examines the range of funding at the school district level within selected states. Specifically it looks at the average per-pupil spending for each state's 10 school districts with the least funding, and the 10 school districts with the greatest funding. The states selected for discussion were three of the four top states in terms of per-pupil spending in 2008 (Alaska had no school district data), three from the mid-range, and the three states with the lowest per-pupil spending.

Table 10.
Disparity of per-pupil funding between individual school districts within selected states

STATE	Average PPS for 10 Lowest Funded School Districts	Average PPS for 10 Highest Funded School Districts	PPS Difference Between Lowest and Highest	% Difference
HIGH PPS STATES				
Wyoming	$11,367	$18,161	$6,794	60%
New York	$9,649	$21,756	$12,107	125%
New Jersey	$7,896	$15,070	$7,174	91%
MID PPS STATES				
New Mexico	$6,512	$12,628	$6,116	94%
West Virginia	$7,352	$9,306	$1,954	27%
Illinois	$4,971	$10,733	$5,762	116%
LOW PPS STATES				
Oklahoma	$5,141	$11,747	$6,606	128%
Utah	$4,551	$8,482	$3,931	86%
Tennessee	$5,010	$7,953	$2,943	59%

Data are drawn from Center for American Progress, 2011.

These data suggest significant differences in funding at the school district level within states, ranging from a 27% difference in West Virginia to a 128% difference in Oklahoma. They also suggest that these differences occur irrespective of overall level of per-pupil spending in a state. The average disparity among school districts was 92% in the three states with the highest per-pupil spending, 79% in the three mid-range states, and 91% in the three states with the lowest per-pupil spending. The complexities of state funding

formulas clearly result in inequitable funding at the school district level within states.

Table 11 provides a more detailed analysis of funding by school districts in one state, California, in 2008.

Table 11.
Disparity of per-pupil spending among individual school districts in California

Number of Districts	Adjusted Per-Pupil Spending Range
13	$3,747 – $4,392
102	$5,046 – $5,985
109	$6,001 – $6,989
28	$7,001 – $7,940
18	$8,034 – $8,988
13	$ 9,113 – $9,979
7	$10,002 – $10,945
4	$11,382 – $12,663
7	$13,192 – $14,517
3	$17,099 – $19,168

Data are drawn from Center for American Progress, 2011.

The results are startling. The annual per-pupil spending in individual districts ranged from $3,747 to $19,168. The statistics speak for themselves: 13 districts spent less than $5,000 per pupil annually; 115 spent less than $6,000; and 21 spent more than $10,000. Even among the districts that spent over $10,000 per pupil there was considerable disparity: The 3 districts at the highest end of the range spent almost twice as much as the 7 districts spending just over $10,000. It is hard to justify such an inequitable distribution of resources.

The Center for American Progress issued a 2012 report (Spatig-Amerikaner, 2012) documenting the fact that funding inequalities also exist between individual schools within school districts. It offered the following conclusions:

- While 59% of the inequity was the result of funding differences between school districts within a state, 41% was the result of differential funding of individual schools within a district.
- The per-pupil spending difference attributed to individual schools varied

from state to state, from 77% in South Carolina to 9% in Arizona.

- The primary mechanism driving this phenomenon was the district placing the least experienced, lowest paid teachers in high-minority, high-poverty schools.

The macro benchmark data reflect an education system that is rife with inequities in funding at all levels of the system, leading to the question of whether these inequities are random or systematic.

FUNDING PROGRESSIVITY (EQUITY ACROSS ETHNICITY AND SOCIO-ECONOMIC METRICS)

Funding equity: Federal education law mandates that services provided by a district receiving state and local funds be made available to all attendance areas and all children without discrimination. In particular, it references funding for Title I schools (schools in which 40% or more of students qualify for FRLP). "Title I of the *Elementary and Secondary Education Act of 1965 (ESEA)* requires that schools receiving funds under Title I receive state- and locally-funded services that, taken as a whole, are at least comparable to the state- and locally-funded services provided to non–Title I schools" (Heuer & Stullich, 2011).

Funding progressivity: In several of its research reviews, PISA noted that "many of the world's successful education systems…invest money where the challenges are greatest, rather than making the resources that are devoted to schools dependent on the wealth of the local communities in which schools are located" (OECD, 2011). Specifically, "With the exception of Israel, Turkey, and the United States, where socio-economically disadvantaged schools also tend to be deprived of basic resources, such as favorable student-staff ratios, OECD countries try to place at least an equal, if not larger, number of teachers in socio-economically disadvantaged schools as they do in advantaged schools" (OECD, 2010b).

In its study on school funding, the Education Law Center identified one of its fairness measures as the extent to which state funding systems are sensitive to changes in the rate of poverty. In progressively funded states, poor districts get more funding than wealthy districts. In regressively funded states, poor districts receive less than wealthy districts (Baker, Sciarra, & Farrie, 2012).

We have already seen that funding is not equal at any level of the education funding system: state, school district, individual schools. The funding progressivity question is as follows: How much of the inequity in funding is related to poverty level and to student ethnicity? The Education Law Center examined state funding in the context of each school district's poverty level (percentage of students at the poverty level). It defined progressively funded states as those with a higher ratio of funding for poor districts than wealthy districts, and

regressively funded states as those with a lower ratio of funding for poor districts than wealthy districts. Table 12 shows results for the five most "progressive" states and the five most "regressive" states.

Table 12.
The states with the most progressive and the most regressive education funding systems as shown by funding distribution according to poverty level, 2009

	At 0% Poverty	At 10% Poverty	At 20% Poverty	At 30% Poverty	Ratio High/ Low
Five States with the MOST PROGRESSIVE Funding Ratios					
Utah	$5,772	$6,732	$7,851	$9,157	159%
New Jersey	$13,961	$15,687	$17,626	$19,805	142%
Ohio	$8,993	$9,983	$11,082	$12,301	137%
Minnesota	$10,026	$10,945	$11,948	$13,043	130%
Massachusetts	$12,598	$13,513	$14,496	$15,550	123%
Five States with the MOST REGRESSIVE Funding Ratios					
Nevada	$10,561	$9,617	$8,757	$7,974	76%
Illinois	$11,312	$10,367	$9,501	$8,707	77%
North Carolina	$11,111	$10,240	$9,438	$8,699	78%
New Hampshire	$13,958	$12,833	$11,799	$10,849	78%
North Dakota	$10,774	$9,985	$9,254	$8,577	80%

Adapted from *Is School Funding Fair? A National Report Card* (pp. 14–15), by B. Baker, D. Sciarra, and D. Farrie, 2012, Newark, NJ: Education Law Center. Copyright 2012 by Education Law Center.

The data collected for 2009 highlight a number of points:

- There is a wide variation in education funding within states at the school district level relative to the percentage of students who live in poverty.
- The most progressive states spent significantly more (between 23% and 59% more) on high-poverty school districts than on low-poverty school districts.
- The most regressive states spent significantly less on high-poverty school districts, allocating just 76% to 80% of the amount they sent to low-poverty districts.
- These variations in funding patterns were not related to the overall

per-pupil spending of the state (e.g., Utah has a progressive funding model despite the fact it has one of the lowest overall PPS levels of all states, New Hampshire has a regressive model while ranking in the top third of overall PPS).

- The clear linearity of the data that followed each level of poverty was even clearer when graphed. There was a consistent relationship between poverty level and funding in the five most progressive and the five most regressive states (Figures 26 and 27).

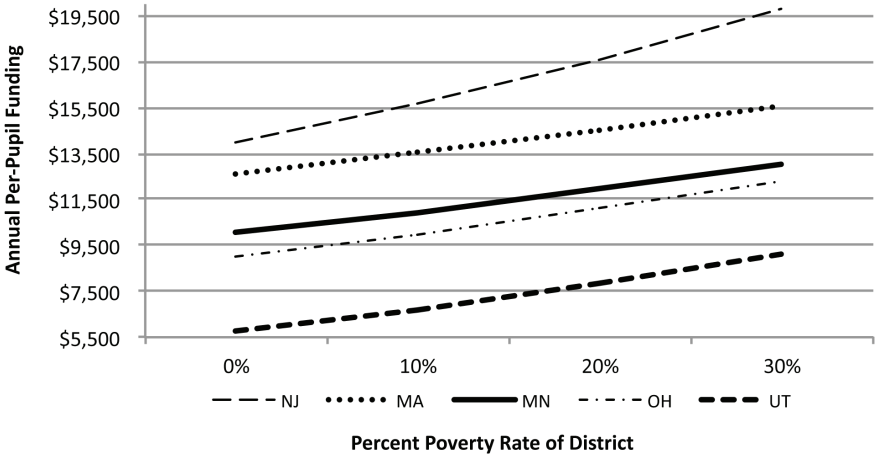

Figure 26. Funding by school district poverty levels in the five most progressive states, 2009.

Data are drawn from Baker, Sciarra, and Farrie (2012, p. 14).

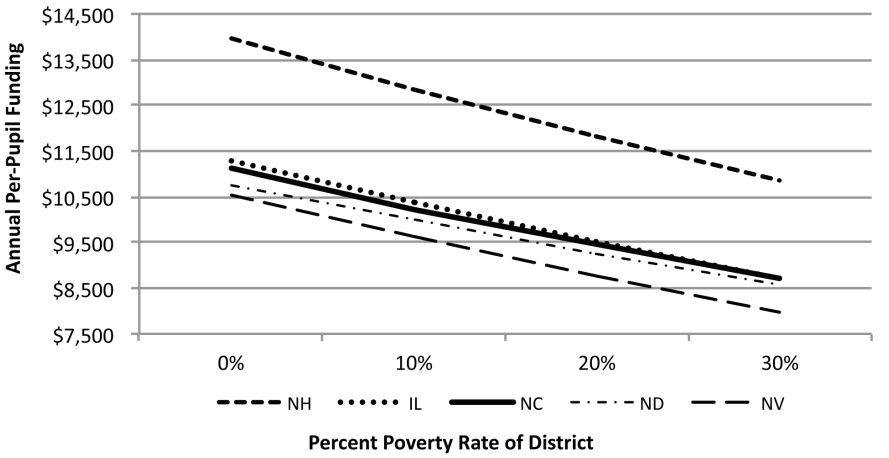

Figure 27. Funding by school district poverty levels in the five most regressive states, 2009.

Data are drawn from Baker, Sciarra, and Farrie (2012, p. 14).

The Education Law Center's report concluded that only 17 states had progressive funding systems, in which high-poverty school districts received more funding than low-poverty districts; 15 states had flat funding systems, in which there was no appreciable difference in the amount of funding; and 16 states had regressive funding systems, in which high-poverty school districts received less funding than low-poverty districts. (Hawaii and the District of Columbia were excluded because each has only one school district; Alaska was excluded from the within-state distribution analysis because of its unique geography and sparse population) (Baker et al., 2012).

This analysis was replicated by Baker and Corcoran (2012). They identified the same five most regressive states as the Education Law Center report did, with comparable, although sometimes even lower funding for high-poverty schools than low-poverty schools. In New Hampshire, high-poverty school districts received only 64% of the funding that low-poverty districts did. In Nevada, poor districts received 67% of the amount that went to wealthy districts, and in North Carolina that figure was 73%, in Illinois 81%, and in North Dakota 81%.

Spatig-Amerikaner (2012) identified a disturbing trend in the inequity of school funding linked to race. His study came to the following conclusions about data collected for 2009:

- Schools across the nation spent $334 more per White student than per non-White student. (That amount represented approximately 8% of the median per-pupil spending nationwide.)

- Schools whose enrollment was more than 90% White spent $733 more per student than schools whose enrollment was more than 90% non-White students. (That amount represented approximately 18% of the median per-pupil spending nationwide.)
- Each increase of 10% in a school's non-White students was associated with a decrease in spending of $75 per student.
- The primary mechanism driving this funding inequity based on race was the district placing the least experienced, lowest paid teachers in high minority, high-poverty schools.

RESOURCE QUALITY

In terms of quality of resources, the primary metric relates to the quality of teachers. Various analyses have demonstrated that minority students and students from lower socio-economic backgrounds are much more likely to have less experienced teachers with higher turnover rates. Figure 28 examines the percentage of first-year teachers in high-poverty versus low-poverty schools.

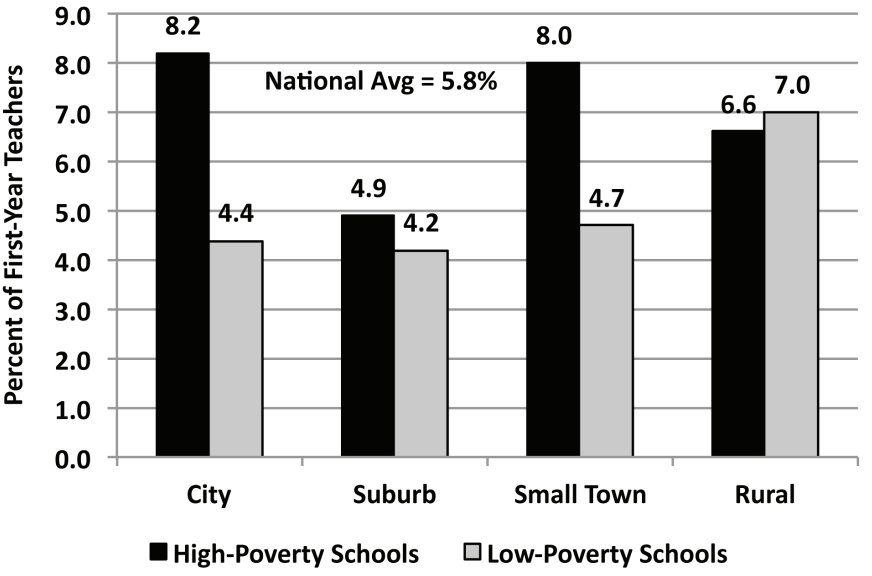

Figure 28. Percent of first-year teachers by school poverty level.

Adapted from *Not Prepared for Class: High-Poverty Schools Continue to Have Fewer In-Field Teachers* (p. 21), by S. Almy and C. Theokas, 2010, Washington, DC: The Education Trust.

First-year teachers are often the least experienced and least effective teachers. They are also the least equipped to work with children who have greater educational needs. Yet, in cities and small towns, the percentage of first-year teachers in high-poverty schools was almost twice as high as in low-poverty schools during 2007–08. In suburbs and rural areas, the percentages of first-year teachers in high-poverty and low-poverty schools were more similar during that same year.

Another critical metric for measuring teacher effectiveness is teacher turnover. Figures 29 analyzes teacher turnover by schools based on their percentage of students qualifying for FRPL. Figure 30 looks at teacher turnover by schools based on their percentage of minority students.

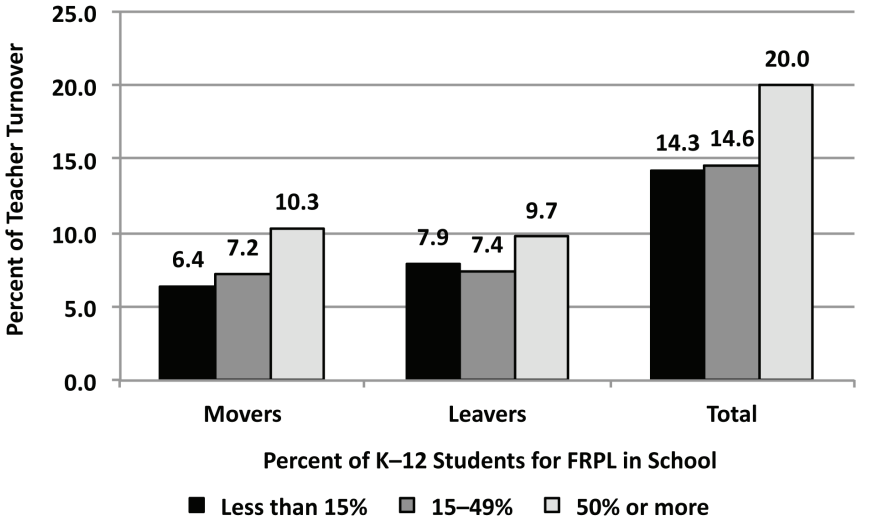

Figure 29. Teacher turnover by K–12 students qualifying for FRPL. Data are drawn from Marvel, Lyter, Peltola, Strizek, and Morton (2006, p. 9).

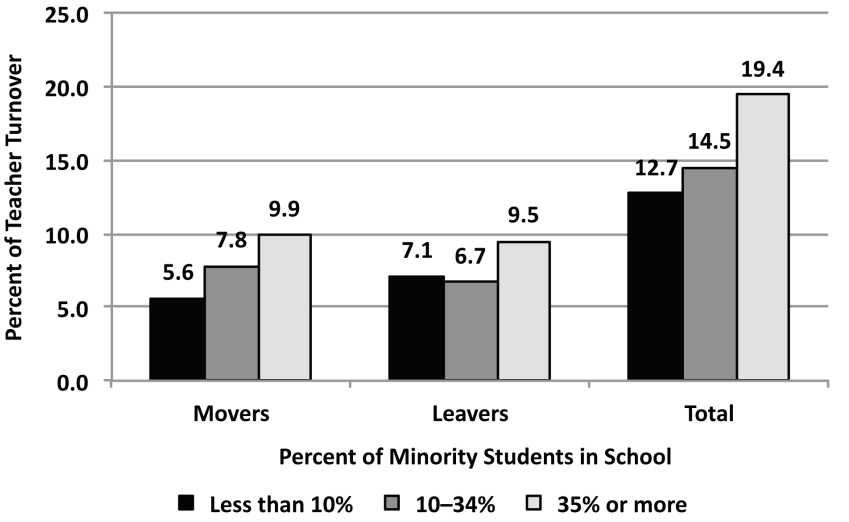

Figure 30. Teacher turnover by percent of minority students in K–12. Data are drawn from Marvel, Lyter, Peltola, Strizek, and Morton (2006, p. 9).

As with funding equity, there is a linear relationship between teacher turnover and a school's poverty level. In 2003–04, schools with 50% or more of their children qualifying for free or reduced-price lunches had a 40% greater teacher turnover than schools with less than 15% of their children qualifying for FRPL (20% versus 14.3%). The same linear relationship was evident in a school's student ethnicity. Schools with 35% or more minority students had a 53% greater teacher turnover (19.4% versus 12.7%). While some small percentage of turnover may be beneficial if the least effective teachers leave, this level of turnover certainly affects the overall quality of the teaching staff at a school. It also raises the question of why teachers were more likely to leave these schools.

A more thorough study of the issue of teacher quality was conducted by the Illinois Education Research Council, which developed the Teacher Quality Index. The TQI measures the quality of teachers in a school using teacher attributes that research suggests affect student performance. The council amassed a TQI database of all teachers by school and assigned each school a TQI rating. It then ranked schools into four quartiles. The schools in the top quartile had the highest ranking (teachers who were better educated and more experienced and whose academic skills were stronger). The schools in the fourth quartile had the lowest ranking. The council cross-referenced all the

schools by each school's percentage of minority students and level of poverty, resulting in the data shown in Figures 31 and 32.

Figure 31 displays student minority data for the schools in the bottom TQI quartile (schools with the weakest teachers). Figure 32 displays school minority data for the schools in the top TQI quartile (schools with the strongest teachers). Each figure leads to the same conclusion: As a school's minority enrollment increases teacher quality decreases (Peske & Haycock, 2006).

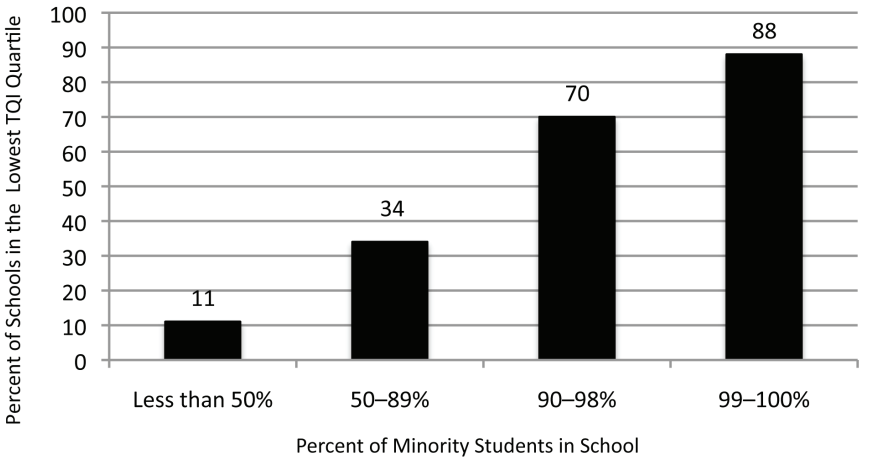

Figure 31. Percent of student minority in schools with the lowest teacher quality.

Adapted from *Teacher Inequality: How Poor and Minority Students Are Shortchanged on Teacher Quality* (p. 7), by H. G. Peske and K. Haycock, 2006, Washington, DC: The Education Trust.

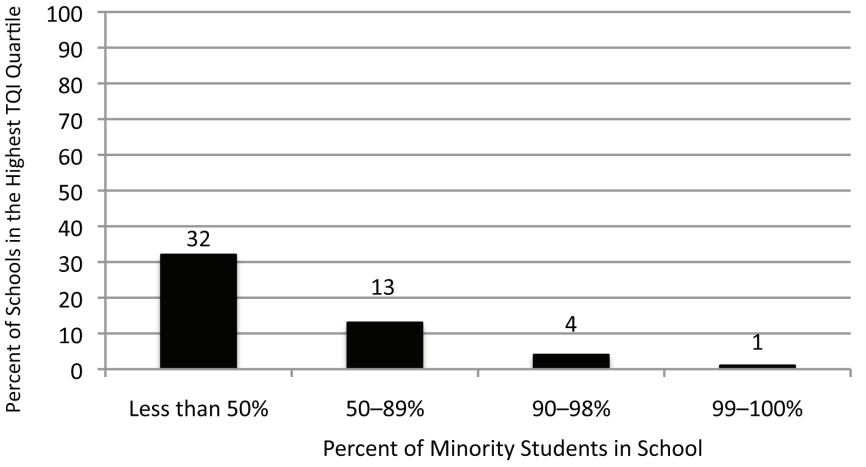

Figure 32. Percent of student minority in schools with the highest teacher quality.

Adapted from *Teacher Inequality: How Poor and Minority Students Are Shortchanged on Teacher Quality* (p. 7), by H. G. Peske and K. Haycock, 2006, Washington, DC: The Education Trust.

The following figures examine the same teacher quality data in the context of school poverty levels. Figure 33 displays school poverty data for the schools in the bottom TQI quartile (schools with the weakest teachers). Figure 34 displays school poverty data for the schools in the top TQI quartile (schools with the strongest teachers). As with minority levels, the conclusion is clear: As a school's poverty enrollment increases teacher quality decreases (Peske & Haycock, 2006).

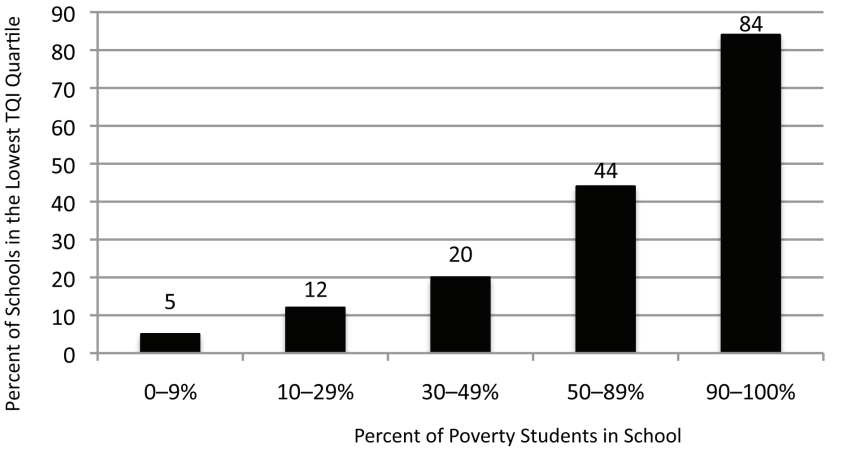

Figure 33. Percent of poverty students in schools with the lowest teacher quality.

Adapted from *Teacher Inequality: How Poor and Minority Students Are Shortchanged on Teacher Quality* (p. 7), by H. G. Peske and K. Haycock, 2006, Washington, DC: The Education Trust.

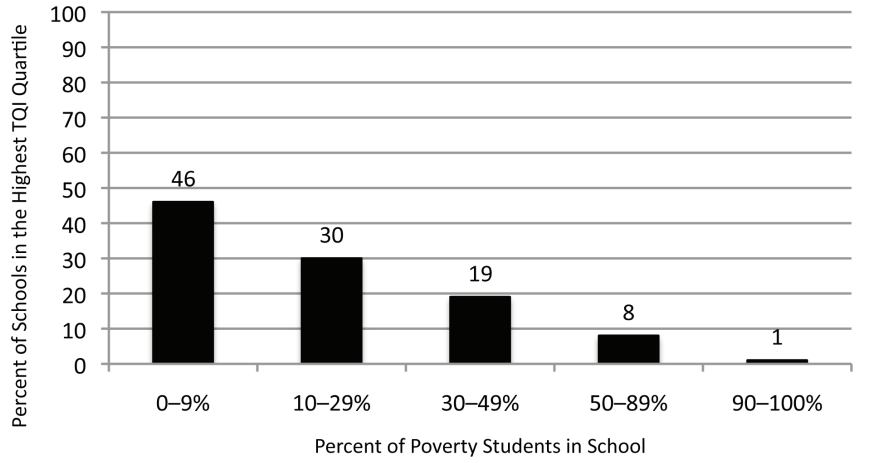

Figure 34. Percent of poverty students in schools with the highest teacher quality.

Adapted from *Teacher Inequality: How Poor and Minority Students Are Shortchanged on Teacher Quality* (p. 7), by H. G. Peske and K. Haycock, 2006, Washington, DC: The Education Trust.

The linear nature of the data makes a strong case that schools with higher percentages of minority or poor students do not have equal access to quality resources where it matters most, the point of delivery.

PISA also examined the question of equity in distribution of educational resources. Its analysis looked at (a) whether or not all schools received equal access to educational resources (quality and quantity) regardless of socio-economic background, (b) whether or not more and better resources were devoted to more advantaged schools, or (c) whether or not more and better resources were devoted to disadvantaged schools.

In particular, the PISA analysis looked at the index of teacher shortage and the index of quality of educational resources. In both categories, in 2009 the United States fell far behind other OECD countries:

In terms of equity in access to resources (allocation of teachers per students to disadvantaged schools), the United States ranked 30th of 34 OECD countries (OECD, 2011).

In terms of equity in quality of educational resources across all schools, the United States ranked 28th of 34 OECD countries (OECD, 2011).

The PISA report noted that "…in 16 OECD countries, the student-teacher ratio relates positively to the socio-economic background of schools. In these countries, more disadvantaged schools tend to have more teachers in comparison with the number of students, which signals that around half of OECD countries try to allocate more teachers to socio-economically disadvantaged schools, presumably with the objective of moderating that disadvantage….Among OECD countries, only in Turkey, Slovenia, Israel, and the United States are socio-economically disadvantaged schools characterized by higher student-teacher ratios; that is, in these countries disadvantaged schools tend to be worse off in the availability of teachers" (OECD, 2010b).

High equity benchmark review

The macro data in this benchmark portray an education system that clearly is not equitable in its learning outcomes, or in the allocation of funding and human capital resources (teachers and principals). The system is particularly inequitable in dealing with students of certain races and from lower socio-economic backgrounds.

The preponderance of evidence suggests the following conclusions regarding learning outcome equity:

- Students who are Black or Hispanic have significantly lower high school

graduation rates, higher dropout rates, and lower test scores than White students.
- Students from higher poverty families have significantly higher dropout rates and lower test scores than those who are from lower poverty families.
- Neither the level of disparity nor the trend has improved in any meaningful way since the early 1990s, when data first became available.
- There is an increasing clustering of students into high-poverty, high-minority schools, which receive less funding and have a higher record of underperformance than low-poverty schools.

The preponderance of evidence suggests the following conclusions regarding resource equity:

- While the United States spends significantly more money on K–12 education than the majority of the other nations, it is not equitably spent across states, school districts, or schools.
- In one third of the states, there is an inverse relationship between funding and a school district's poverty level; that is, the higher the poverty level, the lower the funding.
- Across the nation, schools spend an average of 8% less on non-White students than on White students.
- High-poverty and/or high-minority schools are significantly more likely to have teachers who are less experienced and less effective than are low-poverty and/or low-minority schools.

The overwhelming preponderance of evidence strongly argues that the United States education system is failing the high equity benchmark.

High efficiency benchmark

The high efficiency benchmark looks at the issue of education productivity. In the business world, this is often referred to as return on investment (ROI). How are education outcomes related to spending? Which education systems get the greatest results for dollars spent? The data suggest that spending alone does not necessarily result in improved outcomes. An increasing amount of evidence reinforces the conclusion that there is no direct relationship between funding and school success. The lack of a direct link between per-pupil spending and education outcome is apparent at all levels of macro analysis—international, national, state, and school district.

International

Internationally, the United States spends more per student than any other nations except Luxembourg, Switzerland, and Norway. How does that compare with our ranking on PISA test scores? The following tables provide data on each country's lifetime spending on K–12 education along with its 2009 PISA test scores. Table 13 sorts countries by their 2009 PISA reading test scores and also shows their spending levels. Table 14 sorts the same countries by their 2009 PISA mathematics test scores.

Table 13.
International spending by 2009 PISA reading test scores

Country	2009 Lifetime Per-Pupil Spending	2009 PISA Reading Score
Korea	$73,854	539
Finland	$83,774	536
Canada	$89,966	524
New Zealand	$70,090	521
Japan	$82,857	520
Australia	$89,113	515
Netherlands	$94,678	508
Belgium	$93,146	506
Norway	$120,349	503
Estonia	$58,390	501
Switzerland	$122,797	501
Iceland	$100,022	500
Poland	$52,038	500
United States	**$116,268**	**500**
Germany	$75,259	497
Sweden	$91,763	497
France	$81,121	496
Ireland	$90,743	496
Denmark	$109,017	495
United Kingdom	$94,583	494
Portugal	$68,931	489
Italy	$88,992	486
Slovenia	$91,883	483
Spain	$85,117	481
Czech Republic	$55,168	478
Slovak Republic	$48,712	477
Israel	$54,580	474
Luxembourg	$176,013	472
Austria	$115,563	470
Chile	$29,456	449
Mexico	$22,688	425

Data are drawn from Organisation for Economic Co-operation and Development (OECD) (2012a, p. 228), and Organisation for Economic Co-operation and Development (OECD) (2010c, p. 15).

Table 14.
International spending by 2009 PISA math test scores

Country	2009 Lifetime Per-Pupil Spending	2009 PISA Math Score
Korea	$73,854	546
Finland	$83,774	541
Switzerland	$122,797	534
Japan	$82,857	529
Canada	$89,966	527
Netherlands	$94,678	526
New Zealand	$70,090	519
Belgium	$93,146	515
Australia	$89,113	514
Germany	$75,259	513
Estonia	$58,390	512
Iceland	$100,022	507
Denmark	$109,017	503
Slovenia	$91,883	501
Norway	$120,349	498
France	$81,121	497
Slovak Republic	$48,712	497
Austria	$115,563	496
Poland	$52,038	495
Sweden	$91,763	494
Czech Republic	$55,168	493
United Kingdom	$94,583	492
Luxembourg	$176,013	489
Ireland	$90,743	487
Portugal	$68,931	487
United States	**$116,268**	**487**
Italy	$88,992	483
Spain	$85,117	483
Israel	$54,580	447
Chile	$29,456	421
Mexico	$22,688	419

Data are drawn from Organisation for Economic Co-operation and Development (OECD) (2012a, p. 228), and Organisation for Economic Co-operation and Development (OECD) (2010c, p. 15).

The United States' greater level of spending has not produced commensurate results. Most of the nations with higher scores in reading spent a fraction of what the United States did. The top five nations in reading scores spent an average of $80,108 on lifetime per-pupil spending, which was 69% of what the United States expended. New Zealand spent approximately 60% of the United States' total, with superior results. The data on spending and math scores in Table 14 show a similar pattern. Most of the higher performing nations spent a fraction of what the United States spent, with substantially better results.

National

At the national level, the same disconnect can be seen. As was noted in the previous section, K–12 per-pupil spending has increased steadily over the past 40 years with virtually no gain in two of the most critical performance benchmarks: NAEP test scores in reading and mathematics and high school graduation rates.

State

State funding shows much the same pattern. There is very little correlation between the amount of money spent and student outcomes. The next two tables look at selected states' K–12 per-pupil spending in relation to student outcomes: Table 15 analyzes high school graduation rates, and Table 16 examines fourth-grade NAEP reading scores. In each table, 20 states are grouped by comparable student outcomes:

- group one: the five highest performing states
- group two and three: two groups of five states with virtually identical performance outcomes
- group four: the five lowest performing states.

The question is, are states spending comparable amounts to achieve comparable outcomes?

Table 15.
State per-pupil spending by graduation rates, 2008–09

State	Annual PPS 2008–09	High School Grad Rate 2008–09	Average Scores by Groups	
GROUP ONE				
Wisconsin	$10,807	90.7%		
Vermont	$15,020	89.6%	**score avg.**	88.2%
North Dakota	$9,542	87.4%	**PPS avg.**	$11,533
Minnesota	$11,533	87.4%	**PPS range**	$5,478
Iowa	$10,764	85.7%	**high/low diff.**	57%
GROUP TWO				
Idaho	$7,509	80.6%		
Pennsylvania	$12,976	80.5%	**score avg.**	80.3%
Kansas	$11,060	80.2%	**PPS avg.**	$11,435
Maryland	$13,505	80.1%	**PPS range**	$5,996
Maine	$12,125	79.9%	**high/low diff.**	80%
GROUP THREE				
Texas	$8,862	75.4%		
Connecticut	$15,693	75.4%	**score avg.**	75.4%
Michigan	$9,611	75.3%	**PPS avg.**	$11,932
Rhode Island	$13,047	75.3%	**PPS range**	$6,831
Hawaii	$12,445	75.3%	**high/low diff.**	77%
GROUP FOUR				
South Carolina	$9,657	66.0%		
New Mexico	$10,113	64.8%	**score avg.**	62.3%
Dist. of Columbia	$14,596	62.4%	**PPS avg.**	$10,278
Mississippi	$7,930	62.0%	**PPS range**	$6,666
Nevada	$9,094	56.3%	**high/low diff.**	84%

Data are drawn from Baker, Sciarra, and Farrie (2012, p. 12) and Snyder and Dillow (2012a).

Table 16.
State per-pupil spending by fourth-grade NAEP reading scores, 2008–09

State	Annual PPS 2008–09	4th-Grade NAEP Reading 2009	Average Scores by Groups	
GROUP ONE				
Massachusetts	$14,091	234		
New Jersey	$16,817	229	**score avg.**	230
New Hampshire	$12,206	229	**PPS avg.**	$14,765
Connecticut	$15,693	229	**PPS range**	$4,611
Vermont	$15,020	229	**high/low diff.**	38%
GROUP TWO				
New York	$17,375	224		
Kansas	$11,060	224	**score avg.**	224
Missouri	$9,163	224	**PPS avg.**	$12,540
Maine	$12,125	224	**PPS range**	$8,212
Pennsylvania	$12,976	224	**high/low diff.**	90%
GROUP THREE				
Wisconsin	$10,807	220		
North Carolina	$9,754	219	**score avg.**	219
Utah	$7,379	219	**PPS avg.**	$9,329
Illinois	$9,841	219	**PPS range**	$3,428
Texas	$8,862	219	**high/low diff.**	46%
GROUP FOUR				
Arizona	$7,899	210		
California	$8,897	210	**score avg.**	207
New Mexico	$10,113	208	**PPS avg.**	$10,359
Louisiana	$10,289	207	**PPS range**	$6,697
Dist. of Columbia	$14,596	202	**high/low diff.**	85%

Data are drawn from Baker, Sciarra, and Farrie (2012, p. 12) and National Center for Education Statistics (NCES) (2011b).

The data in Table 15 suggest very little relationship between per-pupil spending and high school graduation rates. Group One consists of the five states with the highest graduation rates, yet each state spent very different amounts of money to achieve similar results. There was an annual per-pupil spending difference of 57%, or $5,478, between Vermont, which spent the most ($15,020) and North Dakota, which spent the least ($9,542). The graduation rate data from Groups Two and Three were virtually identical within each group, yet the funding difference within Group Two was 80% (Maryland at $13,505 and Idaho at $7,509), and within Group Three 77% (Connecticut at $15,693 and Texas at $8,862). And, finally, in Group Four, the lowest performing five states had a funding difference of 84%, with the District of Columbia spending $14,596 per year compared with Mississippi's annual spending of $7,930. Another way to look at the data is this: North Dakota spent virtually the same amount per pupil in 2008–09 as South Carolina, yet its graduation rate was 21.4 percentage points higher (87.4% versus 66%).

The pattern is repeated in Table 16, where state per-pupil spending is compared with fourth-grade NAEP reading scores. The funding difference within each similar performing group ranged from 38% to 90%. Across both tables, the average disparity in funding within these subgroups of comparable performance was 70%. In other words, within each group, the state spending the most spent an average of 70% more than the state spending the least to achieve similar performance outcomes.

This analysis admittedly relies on gross measures. While calculation of state per-pupil spending accounts for the major variations in economic indicators across states and the performance measures are standard, different states have different opportunities and limitations. However, the size and consistency of the pattern in which state funding seemingly is unrelated to student performance outcome is of such a scale that it is hard to draw any other conclusion than that there is no correlation.

School districts

Finally, there is evidence that differential funding at the district level does not produce commensurate outcomes. In 2011, the Center for American Progress conducted an analysis of the academic outcomes of individual school districts by their spending levels, controlling for factors outside a district's control. The result was a comparative analysis of the relationship between spending and outcomes across individual school districts. The analysis looked at the following two sets of data:

Academic achievement index: The average percentage of students across grades designated at or above proficiency on state assessments in reading and math.

Cost: Current expenditures including salaries, services, and supplies (and excluding capital expenses).

The study published performance (achievement index) and spending (adjusted per-pupil spending) data for every school district across the nation. Table 18 summarizes the data for California's 304 school districts, which are grouped by their achievement index in 10% groupings. For example, 9 districts scored 90% or higher on the achievement index, 27 between 80% and 89%, and so on. The spending gap between the highest spending school district and lowest spending one within a performance group was reported.

Table 17.
California school district return on investment

Achievement Index (2008)	Number of Districts	Adjusted Per-Pupil Spending Range (2008)	Max. Diff.	High/Low
90%	9	$6,425 – $17,572	$11,147	273%
80%	27	$6,043 – $13,486	$7,443	223%
70%	52	$4,493 – $17,099	$12,606	381%
60%	81	$4,358 – $19,168	$14,810	440%
50%	91	$4,747 – $14,517	$9,770	306%
40%	37	$4,992 – $13,257	$8,265	248%
30%	7	$4,527 – $13,739	$9,212	303%

Data are drawn from Center for American Progress, 2011.

As with the previous analyses, there is little correlation between per-pupil spending and student performance. Of the nine school districts that demonstrated an achievement index of 90% or above, per-pupil spending ranged from $6,425 to $17,572. At the opposite end, seven school districts at the 30% achievement index had a per-pupil spending range of $4,527 to $13,739. Each level of student achievement showed similar ranges in spending, with some districts achieving much better student performance with much less spending. The disconnect between spending and student outcomes is as dramatic as it is alarming.

High efficiency benchmark review

The previous analyses highlight both the limitation and the value of return on investment (efficiency) data analysis at the macro level. On the limitation side, this type of data analysis doesn't evaluate the return on investment of specific interventions. On the value side, it does highlight the disconnect between levels of funding and student performance outcomes. This disconnect is apparent across multiple levels of analysis: international, national, state, and school district. Simply spending more money does not necessarily improve student performance.

WHERE DO WE GO FROM HERE?

The performance data in all four benchmark categories—participation, quality, equity, and efficiency—are as compelling as they are distressing. Taken together, the education components (federal, state, school district, local school) constituting our national education system are failing our children and society. An unacceptable number of students do not participate fully in school, let alone graduate. Those who do graduate are not likely to have gained proficiency in reading and mathematics. Children of color and/or from lower socio-economic families are significantly more likely to have fewer and lower quality resources and to perform substantially worse than children who are White and from higher socio-economic families. Despite the fact that the United States spends more money on education than most other nations, researchers can show little if any correlation between funding level and student outcomes. And there is little sign of improvement despite significant reform initiatives over the past 40 years. All of which lead to the questions: Now what? Where do we go from here?

Poor performance outcomes highlight two glaring flaws in our education system: (a) the lack of clear education benchmarks that reflect the outcomes we want as a society and (b) the absence of systematic feedback to inform our decisions. We have a long history of neglecting both critical components.

Our nation has provided public education for over 100 years without clearly resolving the issues of what we expect from our education system in terms of socially relevant outcomes. The four selected benchmark categories in this chapter came from OECD, but hopefully reflect core values that will further the debate and transcend ideology, politics, philosophies, and fads. Whatever one's perspective, it is hard to argue against the importance of evaluating our system by its ability to serve all students, produce quality outcomes, treat everyone equitably, and get the most return on our resources. The question then becomes one of how we measure these benchmarks, which always generates

much more debate.

As discussed at the beginning of this chapter, macro-level metrics (standardized tests, graduation rates, dropout rates, per-pupil spending, education resources) have their limitations. They are blunt instruments that move slowly and show only large-scale outcomes and trends. They reflect the overall outcomes of countless interventions across numerous system levels. While we know that the most important feedback is at the micro level—immediate response that drives short-term behavior and affects students directly—there is a critical role for feedback at the macro level. These are the data that tell us how well we are serving *all* our students, which is the ultimate purpose of a national education system.

REFERENCES

Almy, S., & Theokas, C. (2010). *Not prepared for class: High-poverty schools continue to have fewer in-field teachers.* Retrieved from The Education Trust website: http://www.edtrust.org/dc/publication/not-prepared-for-class-high-poverty-schools-continue-to-have-fewer-in-field-teachers

Aud, S., Hussar, W., Johnson, F., Kena, G., Roth, E., Manning, E., … Zhang, J. (2012). *The condition of education 2012.* (NCES 2012-045). Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2012045

Aud, S., Hussar, W., Kena, G., Bianco, K., Frohlich, L., Kemp, J., & Tahan, K. (2011). *The condition of education 2011.* (NCES 2011-033). Retrieved from http://nces.ed.gov/pubs2011/2011033.pdf

Baker, B., Sciarra, D., & Farrie, D. (2012). *Is school funding fair? A national report card.* Retrieved from Education Law Center website: http://www.schoolfundingfairness.org/National_Report_Card_2012.pdf

Baker, B. D., & Corcoran, S. P. (2012). *The stealth inequities of school funding: How state and local school finance systems perpetuate inequitable student spending.* Retrieved from Center for American Progress website: http://www.americanprogress.org/wp-content/uploads/2012/09/StealthInequities.pdf

Bandeira de Mello, V. (2011). *Mapping state proficiency standards onto the NAEP scales: Variation and change in state standards for reading and mathematics, 2005–2009.* (NCES 2011-458). Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011458

Boser, U. (2011). *Return on educational investment: A district-by-district evaluation of U.S. educational productivity.* Retrieved from Center for American Progress website: http://www.americanprogress.org/issues/2011/01/pdf/dwwroi.pdf

Center for American Progress. (2011). *Return on educational investment: A district-by-district evaluation of educational productivity.* [Interactive Map]. Retrieved from http://www.americanprogress.org/issues/education/news/2011/01/19/8877/interactive-map-return-on-educational-investment/

Cronin, J., Dahlin, M., Xiang, Y., & McCahon, D. (2009). *The accountability illusion.* Retrieved from http://www.edexcellencemedia.net/publications/2009/200902_accountabilityillusion/2009_AccountabilityIllusion_WholeReport.pdf

Education Finance Statistics Center (EDFIN). (2009). *Percentage distribution of revenues for public elementary and secondary education in the United States, by source: 2008–09.* [Graph 4]. Retrieved from http://nces.ed.gov/edfin/graph_topic.asp?INDEX=4

Egerter, S., Braveman, P., Cubbin, C., Dekker, M., Sadegh-Nobari, T., An, J., & Grossman-Kahn, R. (2009). *Reaching America's health potential: A state-by-state look at adult health.* Retrieved from Robert Wood Johnson Foundation website: http://www.commissiononhealth.org/PDF/d472ee9d-f880-4d04-8358-397977f93447/CBHA_AdultHealthChartbook.pdf

Fleischman, H. L., Hopstock, P. J., Pelczar, M. P., & Shelley, B. E. (2010). *Highlights from PISA 2009: Performance of U.S. 15-year-old students in reading, mathematics, and science literacy in an international context.* (NCES 2011-004). Retrieved from National Center for Education Statistics website:http://nces.ed.gov/pubs2011/2011004.pdf

Gardner, D. P., Larsen, Y. W., Baker, W. O., Campbell, A., Crosby, E. A., Foster, C. A., Jr., ...Wallace, R. (1983). *A nation at risk: The imperative for educational reform. An open letter to the American people. A report to the nation and the secretary of education.* Retrieved from http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED226006

Hall, D. (2005). *Getting honest about grad rates: How states play the numbers and students lose.* Retrieved from The Education Trust website: http://www.edtrust.org/sites/edtrust.org/files/Getting_Honest.pdf

Heuer, R., & Stullich, S. (2011). *Comparability of state and local expenditures among schools within districts: A report from the study of school-level expenditures.* Retrieved from U.S. Department of Education website: http://www2.ed.gov/rschstat/eval/title-i/school-level-expenditures/school-level-expenditures.pdf

Marvel, J., Lyter, D. M., Peltola, P., Strizek, G. A., & Morton, B. A. (2006). *Teacher attrition and mobility: Results from the 2004–05 teacher follow-up survey.* (NCES 2007-307). Retrieved from National Center for Education Statistics website: http://nces.ed.gov/pubs2007/2007307.pdf

National Assessment of Educational Progress (NAEP). (2011a). *The nation's report card: Math grade 4 national results.* Retrieved from http://nationsreportcard.gov/math_2011/gr4_national.asp?subtab_id=Tab_3&tab_id=tab2#chart

National Assessment of Educational Progress (NAEP). (2011b). *The nation's report card: Reading grade 12 national results.* Retrieved from http://nationsreportcard.gov/reading_2009/gr12_national.asp?subtab_id=Tab_3&tab_id=tab2#

National Center for Education Statistics (NCES). (2010a). *An introduction to NAEP.* (NCES 2010-468). Retrieved from National Center for Education Statistics website: http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2010468

National Center for Education Statistics (NCES). (2010b). *The nation's report card: Grade 12 reading and mathematics 2009 national and pilot state results.* (NCES 2011-455). Retrieved http://nces.ed.gov/nationsreportcard/pdf/main2009/2011455.pdf

National Center for Education Statistics (NCES). (2011a). *Data explorer for long-term trend.* [Data file]. Retrieved from http://nces.ed.gov/nationsreportcard/lttdata/

National Center for Education Statistics (NCES). (2011b). *Data explorer for main NDE.* [Data file]. Retrieved from http://nces.ed.gov/nationsreportcard/naepdata/dataset.aspx

National Center for Education Statistics (NCES). (2011c). *The NAEP glossary of terms.* Retrieved from http://nces.ed.gov/nationsreportcard/glossary.asp#basic

National Center for Education Statistics (NCES). (2011d). *The nation's report card: mathematics 2011.* (NCES 2012-458). Retrieved from http://nces.ed.gov/nationsreportcard/pdf/main2011/2012458.pdf

National Center for Education Statistics (NCES). (2011e). *The nation's report card: Reading 2011.* (NCES 2012-457). Retrieved from http://nces.ed.gov/nationsreportcard/pdf/main2011/2012457.pdf

National Center for Education Statistics (NCES). (2011f). *Students meeting state proficiency standards and performing at or above the NAEP proficient level: 2009.* Retrieved from http://nces.ed.gov/nationsreportcard/studies/statemapping/2009_naep_state_table.asp

Orfield, G. (2009). *Reviving the goal of an integrated society: A 21st century challenge.* Retrieved from The Civil Rights Project/Proyecto Derechos Civiles at UCLA website: http://civilrightsproject.ucla.edu/research/k-12-education/integration-and-diversity/reviving-the-goal-of-an-integrated-society-a-21st-century-challenge

Organisation for Economic Co-operation and Development (OECD). (2006). *PISA 2006 technical report.* Retrieved from http://www.oecd.org/pisa/pisaproducts/pisa2006/42025182.pdf

Organisation for Economic Co-operation and Development (OECD). (2010a). *PISA 2009 results: Learning trends–Changes in student performance since 2000 (Volume V).* Retrieved from http://dx.doi.org/10.1787/9789264091580-en

Organisation for Economic Co-operation and Development (OECD). (2010b). *PISA 2009 results: Overcoming social background–Equity in learning opportunities and outcomes (Volume II).* Retrieved from http://dx.doi.org/10.1787/9789264091504-en

Organisation for Economic Co-operation and Development (OECD). (2010c). *PISA 2009 results: What students know and can do–Student performance in reading, mathematics and science (Volume I).* Retrieved from http://dx.doi.org/10.1787/9789264091450-en

Organisation for Economic Co-operation and Development (OECD). (2011). *Lessons from PISA for the United States–Strong performers and successful reformers in education.* OECD Publishing. doi:http://dx.doi.org/10.1787/9789264096660-en

Organisation for Economic Co-operation and Development (OECD). (2012a). *Education at a Glance 2012: OECD Indicators.* OECD Publishing. doi:10.1787/eag-2012-en

Organisation for Economic Co-operation and Development (OECD). (2012b). *Education at a Glance 2012: OECD Indicators–Chapter B: Financial and human resources invested in education–Indicators.* [Table B1.3b]. Retrieved from http://dx.doi.org/10.1787/888932665943

Peske, H. G., & Haycock, K. (2006). *Teacher inequality: How poor and minority students are shortchanged on teacher quality.* Retrieved from The Education Trust website: http://www.edtrust.org/dc/publication/teaching-inequality-how-poor-and-minority-students-are-shortchanged-on-teacher-qualit

Snyder, T. D., & Dillow, S. A. (2012a). *Averaged freshman graduation rates for public secondary schools, by state or jurisdiction: Selected years, 1990–91 through 2008–09.* [Table 113]. Retrieved from http://nces.ed.gov/programs/digest/d11/tables/dt11_113.asp

Snyder, T. D., & Dillow, S. A. (2012b). *Digest of education statistics 2011.* (NCES 2012-001). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2012001

Spatig-Amerikaner, A. (2012). *Unequal education.* Retrieved from Center for American Progress website: http://www.americanprogress.org/wp-content/uploads/2012/08/UnequalEduation-1.pdf

Stillwell, R., Sable, J., & Plotts, C. (2011). *Public school graduates and dropouts from the common core of data: School year 2008–09.* (NCES 2011-312). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011312

Sum, A., Khatiwada, I., McLaughlin, J., & Palma, S. (2009). *The consequences of dropping out of high school.* (Paper 23). Retrieved from Center for Labor Market Studies Publications website: http://hdl.handle.net/2047/d20000596

U.S. Bureau of Labor Statistics. (2013). *Employment status of the civilian population 25 years and over by educational attainment.* [Table A-4]. Retrieved from http://data.bls.gov/cgi-bin/print.pl/news.release/empsit.t04.htm

U.S. Census Bureau. (2011). *Abstract of the United States: 2012 (131st Edition).* [Table 232]. Retrieved from http://www.census.gov/compendia/statab/2012/tables/12s0232.pdf

U.S. Department of Education. (2008). *A uniform, comparable graduation rate: How the final regulations for Title I hold schools, districts, and states accountable for improving graduation rates.* Retrieved from https://www2.ed.gov/policy/elsec/reg/proposal/uniform-grad-rate.html

Chapter 2

Feedback in Education: On Whom and for What

AUBREY DANIELS
Aubrey Daniels International

ABSTRACT: *If education systems are to be effective it is necessary to define the important outcomes they are to produce. Once the outcomes are defined then it is necessary to describe what educators should do that will produce the student outcomes. Neither students nor educators are likely to improve their performance if they do not receive feedback about it. In order to provide feedback it is necessary to measure what students and educators are doing. This chapter reviews what is known about effective feedback, distinguishes between feedback and reinforcement, and provides an outline for how education can begin to reform to be more efficient and effective.*

> *"A culture that is not willing to accept scientific advances in the understanding of human behavior, together with the technology that emerges from these advances, will eventually be replaced by a culture that is."*
> *B. F. Skinner*
> *The Shame of American Education*

A memory that has stuck with me from my undergraduate education involves an experiment I assisted a psychology professor conduct on the effects of feedback on learning. The experiment was very simple. Subjects were asked to turn a knob 180 degrees. Although I was able to see the precise number of degrees they turned the knob, the subjects themselves did not know how far they turned the knob as their hands were shielded from their view.

The results of the experiment were consistent and predictable: Without any knowledge of the results, the subjects were unable to improve. When given feedback (knowledge of results), they all improved.

Between undergraduate and graduate school, I served in the U.S. Army in Korea as an artillery officer. I was taught how to adjust artillery rounds to hit designated targets at distances of up to 7 miles. The process was fairly simple. We were instructed to use binoculars as a feedback device. If the first round landed above the target, we were taught to adjust the second round so that it fell below the target. When the target was bracketed, the distances were halved until the target was hit. Using this feedback procedure, we became proficient in directing the gun crews to hit the target with a minimum number of rounds. As in my psychology experiment, feedback made our adjustments efficient and effective.

In graduate school, while training to be a clinical psychologist, I discovered that traditional psychotherapy treatment provided little feedback for either the patient or the therapist. The closest thing to feedback was the verbal report of the patient. Patients reported that they "felt better," "about the same," or "worse." Although some patients professed good results, many seemed to make little progress and others abandoned treatment after only a few therapy sessions.

Interestingly, lack of progress was always blamed on the patient. They were "resistive," "uncooperative," "neurotic," or "crazy." While it didn't seem right to blame the patient for being a patient, I was to learn later that this approach was also common in other fields such as education, where poor results are blamed on the student, the parent, the government, or the community. Assigning blame to the person seeking assistance was very unsatisfying to me, and I began to wonder if I had made a bad career move. I was sure there was a better way. Fortunately, during my final year of graduate training, I was introduced to behavioral treatment methods.

My first job as a Ph.D. was at the newly opened Georgia Mental Health Institute, a facility designed to train professionals for the new mental health hospitals that were being built around the state. The treatment programs were conducted under the auspices of the department of psychiatry of Emory University Hospital. Although the Institute was designed to train all mental health disciplines, psychiatrists had the legal responsibility for treatment and all treatment was conducted under their supervision. Because the department of psychiatry was psychoanalytically oriented, treatment was protracted and only a small number of patients could be treated.

Within a few months of the Institute's opening, the directors began to receive heavy pressure from the Department of Mental Health to treat more patients. This meant that primary treatment responsibilities had to be given to other disciplines. Although the social workers, chaplains, and, to a certain extent, vocational rehabilitation counselors had some patient treatment responsibility, the clinical psychologists, also holding the title "Doctor," quickly assumed a full load of patients with little oversight from the unit psychiatrists. Because I was eager to use behavioral treatment methods, I volunteered to treat any patient whom the unit psychiatrist or psychiatric residents did not want or with

whom the psychoanalytically oriented treatment had not been effective.

My early patients were those diagnosed as chronic schizophrenics or chronic depressives, or who had long-term incapacitating phobias. Since behavioral treatment relies on data to evaluate its effectiveness and for determining when reinforcement is earned, all of my patients were on an individualized point and token system. Patients earned points and tokens for behavior demonstrated during their therapy sessions with me and from the nursing staff for behavior on the residential unit during the rest of the day. This method of treatment allowed a patient to monitor his or her progress continually, and allowed the therapist and other staff to know the patient's progress, or lack of it, at any point in time. All patients on token plans made rapid progress compared to those not on such plans.

Because phobics are difficult to treat with psychoanalytic treatment and since I had volunteered to treat them, typically they were assigned to me on admission to our unit. The treatment method I used, developed by Dr. Joseph Wolpe, is called *systematic desensitization*. Phobic behavior is treated as learned maladaptive behavior rather than the result of some sexual or oedipal trauma.

Measurement is easy because phobic behavior is easy to see. For patients who had agoraphobia—the fear of outdoors or open spaces—I simply took them to the front door and asked them to see how far they could go down the walkway before having to come back into the building. I then counted the steps or measured the number of feet traversed. Often the baseline for individual patients was zero. Most were unable to even crack the door without expressing considerable anxiety. However, given that baseline, when the patient was able to stand in the open doorway, I was able to provide positive reinforcement for that achievement. This was repeated with every step in the desensitization protocol, and usually within a matter of days most patients were able to walk around the campus unescorted. With feedback and reinforcement paired this way, patients who had been in treatment for periods ranging from 15 to 25 years were discharged in as few as 90 days. (I maintained contact with several of these patients for many years and they remained symptom free.) Following this initial success, we were given more patients to treat. The psychology staff and interns were soon treating most of the patients assigned to the unit and with good results.

The clinical director took notice, and when he was promoted to superintendent of a new 500-bed regional hospital in the Atlanta area, he asked me to join him as head of psychology, education, and training.

At Georgia Regional Hospital I was allowed to create a computerized system-wide token economy in which all 500 patients had individualized treatment plans that were reviewed at least weekly. Every patient had a point card and knew what he or she had to do to earn points and the number of points needed to pay for merchandise, trips home, and other reinforcing events and activities. While this may not sound like leading-edge treatment today,

it was revolutionary in the 1960s. The results were revolutionary also. Rehospitalization was reduced from approximately 75% to 11% in the first year, and the average length of stay in the hospital was reduced from months and years to weeks.

GOING TO SCHOOL

In 1969, Jim Grenade, a vocational rehabilitation counselor, and I wrote a grant entitled "Innovative Grant for the Behavior Disorders." Truancy was a national concern at that time and this grant was designed as an experiment to see if we could create a method to keep truant students in school. Most of the participants in this project had been arrested one or more times for committing crimes serious enough for them to be sent to the Fulton County Juvenile Detention Center.

Juveniles assigned to this project were released from the detention center as long as they remained in school. Grenade was the homeroom and study hall teacher. When not in his class, students attended regular classes. In that era of social promotion, it was common for our students to be functionally illiterate even in the 10th grade. The 28 students in the program were, on average, 3 to 4 years behind their academic grade level.

Following the model used at Georgia Regional Hospital, we developed a point system in which each student was assigned behavioral and academic pinpoints performance criteria for earning privileges and tangible reinforcers (tickets to sports events were the most popular). Grenade was a master in dealing with these kids. He was tough in that he was not deterred by the many excuses or threats the students made when they failed to earn the points required to keep them out of the detention center for the weekend. He was focused on creating successful students, and success was assessed by academic measures and student referrals to the program. Even the toughest, most recalcitrant students came to respect and trust Grenade because he was probably the only person in their lives who followed through on promises.

In addition to the points Grenade awarded, classroom teachers gave points for various academic behaviors in their classes. Since we needed a way to measure progress, all students were sent to a local tutoring center, where they also earned points. We chose a company called Learning Foundations to tutor our students in basic studies because its remedial tutoring centers in Atlanta were the only ones to use teaching machines. Data from the machines allowed us to give the students a report card every day. The result was that over 90% of the students remained in school, improving several academic years during one school year. The changes in social and academic success were so dramatic that the director of Learning Foundations asked if it was possible to develop a

similar program for its centers. Of course I said that it was, and the centers saw good results with a non-delinquent population.[1]

An executive vice president with Tarkenton Ventures, the parent company of Learning Foundations, approached me with a problem after seeing how the behavior of the students changed with the use of the point system. One of his companies was involved in a government program administered by the National Alliance of Businessmen called JOBs 70s. The program targeted the hard-core unemployed.

To encourage companies to hire participants from this group of unemployed people, the government paid for recruitment, training, child care, medical care, and job training. The problem the executive vice president presented was that while his company was able to find, recruit, and train chronically unemployed people, when they were assigned jobs in a textile factory, supervisors fired them in a matter of days or weeks. He asked me simply, "Do you think you could teach the supervisors to *keep* them?" Of course I said yes, and that was the beginning of what was to become Aubrey Daniels International (ADI). Ninety-day turnover (terminations) was reduced by half in 90 days. As a result, our business exploded and we subsequently worked with all the major U.S. textile companies.

For the past 34 years, ADI has used behavioral methods to improve business performance by upgrading management skills and management systems in virtually every kind of business in over 30 countries. Even though our work has been primarily in the private sector, we are increasingly being asked to work with non-profit organizations.

WHY DO SCHOOLS EXIST?

Before addressing the issue of feedback in education, a more basic question needs to be answered: Why do schools exist? Although this may seem an unnecessary question, if it had actually been addressed on a practical level, many of the current problems in education would not exist. I suggest that it must be answered in order to fix schools. Once that question is answered, the answers to other questions will come easily.

The most apparent answer is, to educate students. But what is an educated student? How do we tell an educated student from an uneducated one? Once we define the word "educated," we will know how to measure the effectiveness of a school and a teacher. The measure is simply the number of students who are

[1] During this time I also helped a prevocational training center for mildly retarded students and a vocational rehabilitation center at Cherry Hospital in North Carolina develop similar programs, all with good success.

educated according to a pinpointed (highly specific) definition.

The problem starts with the fact that educators cannot agree on what the outcome of education is, or should be. Is it knowledge or skill, or both? If both, what is the balance? Is knowledge more important than skill, and, if so, how much more important? While I will not address these issues in this paper, I submit that until attempts are made to define and measure the elements that constitute a good education, nothing will change.

We can measure teacher effectiveness, but the typical measures used have all too often resulted in more punishment than positive reinforcement. That is why after centuries we still cannot agree on what an education should be. If measures are well constructed, teachers can be motivated to meet or exceed them. If they are not constructed properly or administered well, educators will naturally resist them to the detriment of the educational system and ultimately to the detriment of students.

Here's the rub: If you don't have data, you don't know what you are doing. We need data, but on whom and for what. More importantly, how will the data be used?

For a moment, let's assume that the mission of a school is to create successful students. The classroom teacher's role is to create successful students in the subject matter of the class. If teacher accountability is to create successful students, what is the accountability of other staff? It is actually simple. Every staff member's charge is to help teachers be successful. The only reason any education staff exists at the school, county, state, or federal level is to help teachers educate children more effectively. Right away you can see that accountability in these jobs should be primarily for *valuable* behaviors that have a direct link or connection to increased student learning. Look at the following figures.

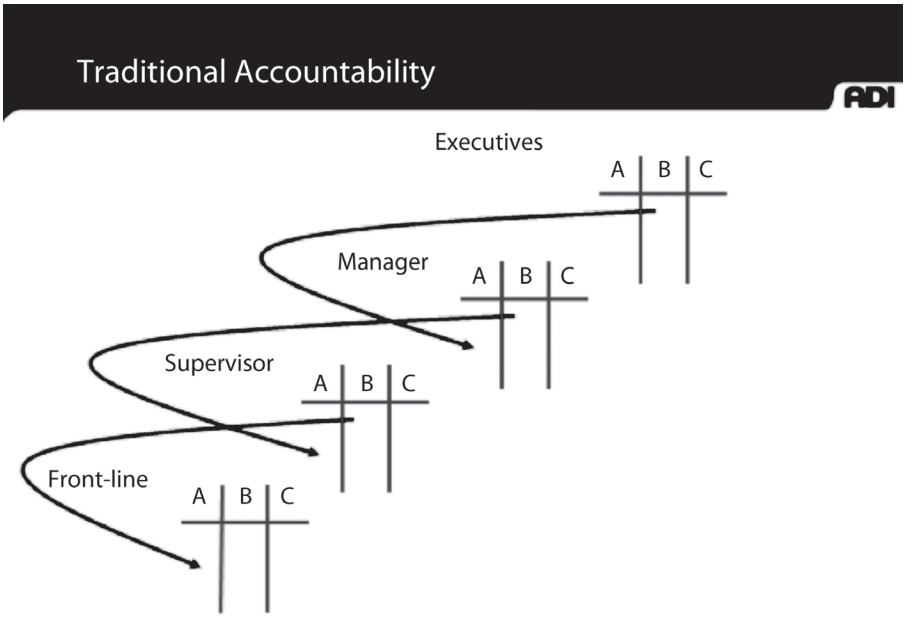

Figure 1. Traditional accountability.

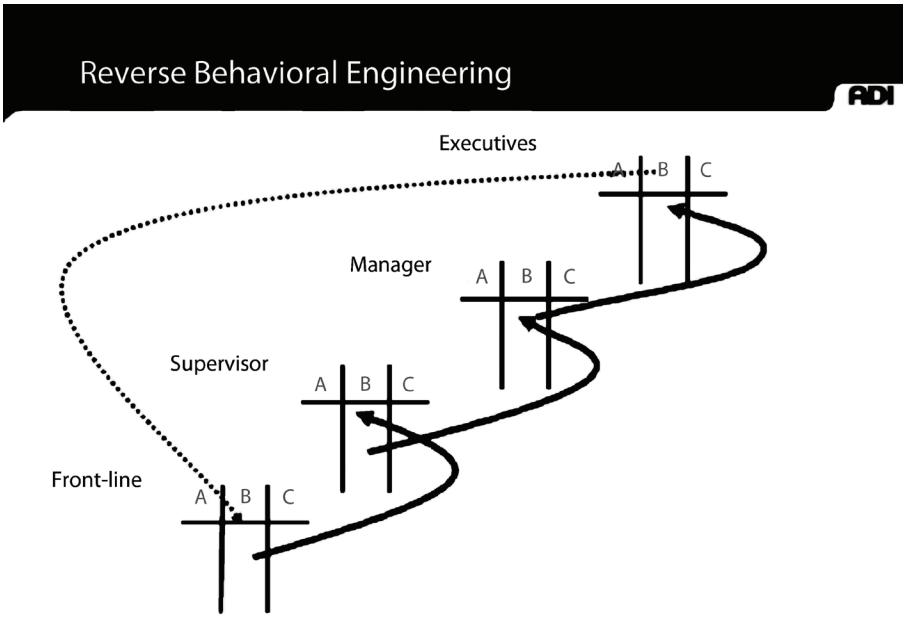

Figure 2. Reverse behavioral engineering.

Figure 1 is a traditional accountability chart. When you start at the top and cascade down, everyone has a different job. When improvements are needed, the jobholders above the front-line employee level (teachers, in this case) usually ask for more money and more staff, typically resulting in more meetings and more paperwork.

Rather than start at the top, the reverse behavioral engineering accountability chart in Figure 2 starts with the front-line employee (teacher) and asks, "What does a teacher need to do to help students learn?" That question is repeated at each level and for all administrative and support personnel—in other words, "What does a department head do to help teachers teach more effectively?" "What does the principal do to help a department head create more successful teachers?" In this model, no one can be successful if students don't learn.

However, even if the children learn, some employees in the chain may not be successful if they cannot demonstrate behavior that helped the level below to be successful. If a department head cannot show behavior that was beneficial to a teacher's effectiveness, then that person was not effective and the teacher was successful in spite of the department head.

This model of accountability flushes out redundancy, ineffectiveness, and incompetence. Once you know what the job requirement is, then the issue of feedback can be more focused on valuable behavior and outcomes at every level of school staff. As you will see, feedback is needed at every level but will be different at every level.

WHAT IS FEEDBACK?

If someone walks up to you and asks, "May I give you some feedback?" I advise that you excuse yourself and leave quickly because, in the popular vernacular, the term "feedback" is code for "May I criticize you?" That kind of information is rarely helpful and falls in the category of what a friend of mine refers to as "more unsolicited advice." The function of feedback should be to provide information that will promote success—in this case, increase teacher effectiveness and improve student learning. Therefore, I will confine my remarks to what I call performance feedback.

In any context where the goal is to help someone improve a skill or social behavior, performance feedback is information about performance that will allow, for example, a student to change, or improve. Allowing the recipient to change is an important part of this definition of feedback, because much information that is presented as feedback does *not* help a person improve.

For feedback to be performance feedback, it must be information that allows the performer to adjust his or her behavior toward more effective or efficient performance. For example, most people would say that seeing the flight of a

golf ball is feedback for a golfer. It is certainly immediate information about how the club hit the ball. Watching a ball careen out of bounds with a vicious curve (slice) motivates every golfer to make changes to correct the flight path of the ball. However, I know of some golfers who have been playing for 50 years and still cannot hit the ball straight. For average players, seeing the flight of the ball is not information that will allow them to improve. They don't usually consider the angle of the clubface when it hits the ball. Furthermore, if they do, they typically don't know what to change about their stance, grip, position of legs and arms, and so on—all factors that affect the angle of the clubface when it comes in contact with the ball.

Just as information about the flight of the ball doesn't help the average golfer to improve, a failing grade doesn't tell a student how not to fail the next test. Therefore, in those instances, the flight of the ball and a failing grade do not fit the definition of performance feedback. However, the flight of the ball is performance feedback for a professional golfer who, on seeing a slice, knows what caused it and is able to change the offending behavior on the next swing. Likewise, being able to recognize and convey what a student must do to improve is the role of a professional teacher.

For many students, getting an answer wrong does not necessarily help them improve, since they may not know what specific thing they did that caused the answer to be wrong. Therefore, performance feedback is limited to a particular kind of information. When a teacher says, "That is wrong—try it again," the student may either repeat the mistake or engage in highly variable behavior far from the behavior that will result in improvement. It would be more helpful to say something like, "Try this next time," followed by relevant and helpful instruction.

The point is that performance feedback is more than a score, and effective feedback for one student may not be effective for another. For expert mechanics, hearing a noise in an engine is performance feedback as they will know exactly what to fix. For expert musicians, hearing the sound produced by the musical instrument is performance feedback because they will know precisely how to correct their performance. Teachers should then always strive to know the specific form of feedback that will help each student.

PUT IT ON A GRAPH

In over 30 years of helping companies improve performance, ADI has taught managers and supervisors to "put it on a graph." Let's assume you are presenting data that meet the definition of performance feedback. A graphic display of performance that allows performers to see where they are relative to where they started and where they are going is often highly effective in helping them

improve. Such data have been associated with tremendous improvement in situations where poor performance has existed for years and where managers thought individuals or groups could not, or would not, change. However, this is not the whole story.

If you want to improve some aspect of your own performance, start by tracking it graphically. If you want to lose weight, weigh yourself every day and record the result on a graph. If you want to stop smoking, graph the number of cigarettes you smoke daily. If you want to exercise more, graph the number of minutes you exercise or the number of repetitions of specific exercises. Chances are high that you will lose a couple of pounds, smoke fewer cigarettes, and exercise some more. However, the changes will be small and temporary. You may lose 1 or 2 pounds, smoke one or two fewer cigarettes, and exercise a couple of times more than normal before you return to your former weight, addiction, or exercise routine.

I learned this early in my business consulting when visiting facilities where the supervisors had previously made significant improvement in quality or production only to find that the improvement graph had not been updated for several weeks. When I asked why, the response was, "The graph quit working." All they had done was to graphically track the performance, and although it resulted in some quick improvement, the upswing rarely lasted.

NECESSARY BUT NOT SUFFICIENT

Improvement in performance is practically impossible without some knowledge of the results of behavior. You can't learn to talk, walk, write your name, or ride a bicycle without some form of feedback from the environment about your behavior. The famous Helen Keller, a blind and deaf child, learned to communicate but only because she had a teacher who found a way to create effective feedback for her verbal behavior. (It is interesting that Helen Keller is well known, but almost no one knows the name of her teacher—Annie Sullivan—the real heroine in Helen's achievements.) Another problem with performance feedback is that it does not change behavior. I have frequently heard it said about some undesirable behavior, "I have given him feedback on that many times and nothing changes." In fact, the person may know what to do but doesn't do it. Why?

Although feedback is often paired with consequences, the feedback does not change performance—*the consequences do*. This fact is poorly understood in education. Situations in which the measure is the number of wrong answers not only fails to motivate, but often does the opposite. As simpleminded as it may seem, feedback on the number of correct answers (although the reciprocal can be inferred from the number of wrong answers) is more motivating. The

number of correct answers focuses students on what they accomplished rather than what they failed to do correctly.

I have said many times that the best job you will ever have is one where you know at the end of the day how well you did. Most students don't have that job; most teachers don't have that job. However, teachers can create that job for students and for themselves. What if every student went home every day knowing what he or she accomplished that day? How motivating would that be? Shouldn't every teacher know what he or she wants every student to learn each day? How motivating would it be for a teacher to go home knowing that every student learned everything that the teacher targeted to teach that day?

While this seems impractical to most teachers, it is a reality at Morningside Academy in Seattle, Washington. Founded by Dr. Kent Johnson in 1980, Morningside gives all parents a written, money-back guarantee that a student who is behind grade level will gain at least two academic years per year of instruction in his or her worst subject. In over 30 years of operation, Morningside has refunded less than 1% of tuition. All of Morningside's teaching materials and teaching methods have been thoroughly researched in order to make sure that they increase student learning.

A typical class hour at Morningside consists of 10 minutes of instruction, 40 minutes of practice on the instruction received, and a 10-minute break. A distinguishing feature of Morningside, in addition to assigning no homework, is that every student receives a report card every day! This sounds reasonable when you consider that the teacher should know what he or she needs to accomplish each day.

A cartoon in one of my books depicts a caveman standing in front of a progress status graph scratched on the cave wall. The caveman says to his friend, "It's just something the kids scratched out, but for some reason I feel good when I look at it." Changes in the data that make a person feel good are almost always associated with improvement. The trick is to create conditions in which seeing the graph makes the person who has improved feel good about the improvement. Creating those conditions requires knowledge of consequences. Although feedback is necessary for improvement, consequences change behavior.

CONSEQUENCES

Of the four behavioral consequences—positive reinforcement, negative reinforcement, punishment, and penalty—only two are of concern in improving the performance of students: positive and negative reinforcement. Positive reinforcement is clearly the most powerful interpersonal tool known, but at the same time it is the most misunderstood and misused.

Negative reinforcement, by far the most frequent consequence in schools and businesses, occurs when a person increases a behavior in order to escape or avoid some form of punishment. With negative reinforcement, people improve because they "have to." If students are told that they cannot go to recess until their work is completed, it is likely that the teacher will see an increase in behavior, because the students want to enjoy a full recess period. While resulting in improvement, the negative reinforcement will elicit only enough behavior to enable the students to go to recess. If all that a student ever does is what the teacher assigns and no more, then neither the teacher nor the parents have made the subject positively reinforcing. Unfortunately, a large part of education is accomplished through negative reinforcement. While negative reinforcement gets a minimum of improvement, it never captures the discretionary effort that characterizes love of learning. The only way to do that is with positive reinforcement.

Figure 3. Behavioral consequences and their effect.

POSITIVE REINFORCEMENT

Surprising to many teachers, positive reinforcement is not a pat on the back, telling the student he or she is smart, saying "good job," or giving the student a sticker or gold star. While all of these examples *may* be positive reinforcers to a given child in a given situation, they could also be punishers. The impact on the behavior is the proof. If the behavior increases, it is a positive reinforcer; if the behavior decreases, it is a punisher. Students let the teacher know what their reinforcers are, not by what they say, but by the way they respond behaviorally. It is quite possible that a student will tell a teacher that he or she likes something, but when the teacher tries it, the student's performance doesn't improve.

One thing teachers can count on is that nothing is positively reinforcing to all students and that everything is reinforcing to some students. There is no substitute for finding the unique reinforcers for each and every student. While the process can be time consuming in the short run, it will be efficient in the long run after the elimination of many false starts. Finding a student's positive reinforcers may require trial and error. Although teachers can make mistakes in what they choose as reinforcers, if the worst error they make in teaching is to try something as a positive reinforcer that turns out not to be one, they will have made the best mistake possible. The worst mistake is to try something as a punisher, such as negative attention, only to discover that it is a positive reinforcer to the recipient.

Finding an effective positive reinforcer is only the beginning of effective teaching. Reinforcers that are immediate are more effective than delayed reinforcers. Reinforcers lose a significant part of their value within minutes of the occurrence of the behavior. Not only do they lose their value, but when they are delivered later they may fall on a behavior that is not productive and thus may increase the unwanted or unproductive behavior. While some people have a difficult time believing this, there would be no superstitious behavior if it were not so. People develop superstitious behavior when there is a coincidental pairing of a reinforcer with some unrelated behavior.

Of course, I realize that teachers cannot see or be in a position to reinforce every occurrence of a behavior. That is why students must be taught the proper way to reinforce peers. People who are positively reinforced will reinforce others more often. In addition, those who are reinforced for some improvement or accomplishment can reinforce themselves in similar situations. A student who receives a teacher's praise for a creative production is likely to look to the teacher when completing another production. When the teacher is not present, the student knows that the teacher would like the present creation because of what happened before. You cannot be proud of yourself until someone has been proud of you. Students who have been praised for some accomplishment will tell other students as well as their parents and grandparents about the

accomplishment. This multiplies the reinforcement associated with the behavior. Although part of this reinforcement is delayed, it has some effect on subsequent behavior as parents, or others, are likely to ask the students how they did it or to show them what they did. All of this provides reinforcement for the demonstrated behavior.

One of the biggest problems with reinforcement in schools is that it is delivered non-contingently. Simply stated, often the positive reinforcer is not earned. In other words, there is no real accomplishment. My grandson played T-ball when he was 6. He knew little about baseball and spent more time in the outfield looking for four-leaf clovers or animals in the clouds than looking at the batter. When the season ended, everyone received a trophy. On the way home he asked his mother, "Mama, why did I get a trophy?" His mother had to think quickly and replied, "Well, Elijah, you were at every practice."

The practice of giving rewards when they are not earned is a bad practice. It not only creates a mentality of entitlement, but it also robs children and adults of the joy of accomplishment. Teachers cannot give students self-esteem; the students must earn it. However, teachers can create the conditions in which every student can earn rewards. Even the smallest improvement is an accomplishment. Teachers who make improvement a reinforcer will benefit their students for a lifetime.

People have told me over the years, "I reinforced her but she didn't change." My response is, "One positive reinforcer will not change your life." B. F. Skinner estimated that it requires 50,000 contingencies to teach basic math. He was wrong because with modern technology we are now able to track contingencies involved in learning much more accurately than in Skinner's day. By a "contingency," I mean an opportunity to do something right or wrong and where a correct response provides an opportunity for positive reinforcement. At Morningside Academy a child may receive 50,000 contingencies a month and may do over 100 math facts per minute.

Some teachers do not understand the value of rapid responding. They think it is nothing but repetition, and that went out of instruction many years ago. Repetition without reinforcement is resisted for good reason—it's boring! Yet, repetition with reinforcement is exciting and energizing. I suggest that the lack of repetition paired with reinforcement is one of the reasons academic achievement is so dismal in our schools today. Repetition with positive reinforcement is also the reason that computer games are so much more attractive to students and adults than traditional instruction. When playing video games, players often receive as many as 200 reinforcers per minute. In how many classes do students receive as many as 200 reinforcers a minute? This is the reason that it may take 12 years to become fluent in subjects that could be taught in a matter of several weeks using modern technology. Maximizing the learning opportunities requires many reinforcers by teachers, the material, peers, and parents. It is the teacher's responsibility to manage reinforcers.

Although performance feedback and positive reinforcement are necessary at every level of the educational system, the frequency and form of both vary from job to job.

WHAT I WOULD DO

If I took on the job of making schools more effective, I would do the following: I would first examine every job using the following criterion: "How does this job help students learn?" When a clear link cannot be *demonstrated*, either the job should be eliminated or the position used to create a job for someone who will help the teacher teach more effectively. I am convinced that more than half of all administrative and support jobs could be eliminated or reassigned.

Then I would train all school personnel in the science of behavior. Since teaching is about changing behavior, all teachers should be fluent in applying the science of behavior in the classroom. Teachers need to know more than the basics of the science; they need to know the science in depth, as every child presents a unique opportunity to apply it. However, to create a culture in which everyone participates in student learning, directly or indirectly, I believe it is important that everyone understand how to make sure that only productive behavior is reinforced.

Next I would hire coaches and assign them to all schools to help teachers teach more effectively. The coaches would spend most of their time in the classroom. Their accountability would be to create successful teachers. They would pinpoint behavior that each teacher needed to do, track it, and reinforce it appropriately. A coach would be assigned several teachers, the number of teachers per coach determined by how far the teachers are from performing at desired levels. Coaches would probably be assistant principals.

Then I would develop a bonus plan in which all teachers who are successful would receive financial remuneration commensurate with their level of success. Success would be defined by the rate of improvement and the number of students who improved.

Finally, I would hire Dr. Kent Johnson to install the Generative Instruction Model used so successfully at Morningside Academy.

Many years ago, Dr. Fred Keller, the pioneering behavioral educator, said, "If the student didn't learn, the teacher didn't teach." You can't blame students for not learning, because they are *students* after all. They don't know what it is that they are to be taught.

The movie *Stand and Deliver* chronicled Jaime Escalante's efforts to teach calculus at Garfield High School, an inner city school in East Los Angeles. When he started teaching, he found that students were worse than poor in math skills. Knowing that businesses needed employees who had knowledge of math,

he set out to teach it. This type of story is usually found only in the movies, but Escalante's success along with that of his students was a real-life triumph under the most difficult circumstances. His problem was not the students—he won them over quickly—but the administration and other teachers. His success with students apparently caused such problems with the other teachers that he finally left the school. He was criticized for coming in early, staying late, and teaching too many students per class.

Escalante started at Garfield in 1974 and by 1978, with Ben Jiménez, a fellow teacher he recruited, taught calculus to five students, two of whom passed the Advanced Placement (AP) Calculus test. In 1982, Escalante came into the national spotlight when 18 of his students passed the AP exam. In 1983, 33 students took the exam and 30 passed. By 1987, 73 students passed the AP Calculus AB exam and another 12 passed the more challenging BC version of the test. By 1991, when Escalante left the school, 570 students took AP Calculus tests. This inner city school at one time had more students pass the tests than any other school in California (Mathews, 1989).

While this example is certainly noteworthy, almost all schools have some exemplary teachers. It is unfortunate, but if not for the positive reinforcement they receive from students and parents, they, like Escalante, might leave the system—and many have. I have had a number of teachers tell me that they love to teach but hate where they have to do it. They are not referring to the geographical location of their jobs or the physical conditions, although many times those two aspects leave much to be desired. They speak of the negative atmosphere created by ineffective administrators, time-consuming paperwork, an abundance of conflicting regulations, useless meetings, irrelevant in-service training, and, of course, some uncooperative parents and students. Yet, few jobs in our society are capable of generating more positive reinforcement than teaching. Seeing a student's response to learning a simple fact, his or her curiosity about class material, and the joy in an accomplishment can overcome many of the negatives in "the system." If not for that, education would be in an even bigger mess.

COACHING TEACHERS

More than two decades ago, Bennett (1987) demonstrated that coaching in the classroom was up to 19 times more effective than the usual ways of training teachers. Therefore, a classroom coach for teachers is a cost-effective addition to the faculty. This person should be in the classroom long enough to sample teacher performance and give real-time feedback to teachers on behavior, methods, organization, and planning. The coach would not be an evaluator but instead employed for the sole purpose of increasing teacher effectiveness. Some teachers

require more coaching than others, but over time a school will need fewer classroom coaches. Coaches can then be assigned to work with students who require more individual attention. A primary outcome of classroom coaching is more positive reinforcement for the teacher. Initially, the reinforcement comes from the coach, but if successful this approach will result in increased reinforcement from parents, students, administrators, and other teachers.

In my opinion, these coaches should be behavior analysts and adept at working with teachers as well as special needs students. A coach should work primarily with classroom teachers, and secondarily with individual students.

Every job in the system, from the superintendent to the school custodian, should first be examined from the vantage point of how the job facilitates learning in the classroom. If a connection between a job and student learning is determined, then the tasks involved in the job should be examined to see if they are relevant to student learning. I estimate that a reduction in staff positions of as much as 70% is possible while increasing the rate of student learning. Paperwork, rules, and regulations should similarly be analyzed to see whether they advance learning. If a direct link between a requirement and learning cannot be demonstrated, trash the requirement. I realize that some federal paperwork and requirements could not initially be eliminated, although if they really add no value they should be targeted for eventual elimination. This would be part of the superintendent's responsibility. In education as in business, many systems, processes, and management behaviors waste time and money. Eliminating them will save valuable resources as well as free employees to spend more time and effort on the task of student learning.

Pay for performance has been an issue in schools for many years. Numerous systems have pay for education (the teacher's education, that is) but not pay for student performance. The assumption is that more highly educated teachers will produce better student performance. Consequently, obtaining an advanced degree results in a higher pay grade whether the teacher demonstrates increased effectiveness in the classroom or not. This has long been done in schools, and it is obviously not working. The assumption that higher pay for a higher education constitutes pay for performance is simply wrong. Resistance to pay for performance comes from attempts to use merit pay as an incentive. These systems as used in education are just as flawed as most performance pay plans in business.

In a true pay for performance system, bonuses are triggered by student performance. However, pay for performance should be based on individual student improvement, not average or grade-level performance, because doing otherwise encourages subterfuge and possibly grade alterations as occurred in the recent Atlanta Public Schools scandal. Some goal would be set, and a teacher would begin to earn a bonus only after hitting that goal. The maximum bonus would be paid only if all students met their learning goals. Teachers with a greater number of students would have an opportunity to earn a larger bonus.

The bottom line is that teachers are employed to transfer knowledge and skill to students. Those who excel at it are clearly more valuable to schools than those who don't, and pay should be based on their effectiveness. New Jersey's Governor Christie recently asserted that science teachers should earn more than physical education teachers. His point was that some teachers are more valuable than others. While Christie is focused on the value of subject matter—that science is, in his opinion, more valuable than physical education—I think the focus should be on the value created by a teacher regardless of subject matter. The fact is that some teachers succeed even with minimal resources and support and under the most adverse physical and social circumstances.

In an interview on *60 Minutes*, Michelle Rhee, one-time chancellor of Washington, D.C., public schools, observed during a school visit that most of the classes had very few students in them. When she asked where the students were, she was told that it was Friday and, in addition, it was raining! However, in one class she visited, all the desks were occupied and students were even sitting on the radiators in the back of the room. Later that morning she observed several of the students she remembered from this class leaving school early. When she asked why they were leaving, they said that the only class they found interesting was the class of the teacher she had observed. They came in early for that class and then left for the day.

Let's return to the movie *Stand and Deliver*. The success of Escalante's teaching methods seems to have been lost on today's educators, although he was teaching until the 1990s. Ironically, his phenomenal success in teaching inner city students calculus was his undoing. His success created too much pressure on other, less successful teachers who used every traditional excuse for not being effective teachers: too many students, too little money, lack of parental involvement, and so forth to explain their poor results. He demonstrated that none of that mattered. Although he had many barriers to overcome, they did not prohibit his success.

Escalante did not have the advantage of computers, and his methods involved much repetition. To help students develop fluency in calculus, he also taught during the summer. Think of inner city students volunteering to come to summer school to learn calculus! By 1991 he left, as did his colleague Jiménez, citing faculty politics and petty jealousies. Today the very successful program he started is practically non-existent. I think most people would agree that Escalante should have earned the maximum salary and bonuses allowed by the system whether he was teaching math or basket weaving, because he was a great teacher proven to be so by the measure of his students' success.

Peter Drucker, the influential management consultant, said, "If you can't measure your job, stop doing it and see what changes. The things that change are the measure. If nothing changes, eliminate the job." While I am sure that such a procedure will encounter much resistance in education, it may well identify many excesses.

SUMMARY

Looking at the problems in education, you will likely discover that schools are no better at managing teachers than teachers are at managing students. The current methods of accountability are misguided and ineffective. Evaluating schools on overall learning is resisted at every turn. Teachers and administrators think that because of differences in schools—inner city, suburban, rural—it is not fair to evaluate schools or teachers on whether they meet one-size-fits-all academic standards. I agree. The measure currently used is average improvement by class or school. The problem is that no school is average and no student is average.

The "average" student may be 60% female, 10% Asian, 15% Hispanic, and 15% Black. Who meets those criteria? No one! Parents send children to school to be taught. Whether the average improvement in a class met or exceeded some standard is irrelevant to parents whose child failed. Teachers should be evaluated on the number of students who improved, not some average in which several students scored very high, skewing the class score but leaving behind many students who made no improvement or some minimal increase.

Just because behavior is measured doesn't mean it will be changed. Effective feedback only provides the opportunity to deliver consequences in an effective manner for the right behavior at the right time with the right frequency. Truly effective schools will have plenty of charts showing progress at the student, class, and school levels. Performance feedback is a necessary part of academic success but is in no way sufficient. With a scientific understanding of the proper functions of feedback and consequences, schools can make much progress. Without that understanding, they will make little progress.

REFERENCES

Bennett, B. (1987). The effectiveness of staff development training practices: A meta-analysis. *Dissertation Abstracts International*, *48*(7), 1739.

Mathews, J. (1989) *Escalante: The best teacher in America*. New York:Henry Hold and Co.

Skinner, B. F. (1984). The shame of American education. *American Psychologist, 39*(9), 947.

Chapter 3

Seeking the Magic Metric: Using Evidence to Identify and Track School System Quality

MARY BETH CELIO
Northwest Decision Resources

ABSTRACT: *With the increasing demands for accountability, many educators are searching for a single measure that will inform them about how well their system is serving students-a magic metric. Unfortunately, there is no magic metric that can guide educators decision- making. This chapter reviews three common measurement systems that have developed over the last decade and describes what they can do for educators and their limitations. A set of indicators that can guide educators through the myriad of data that schools collect is described. Finally, an indicator system is proposed that measures the important features of an educational system. The data from this system can be used to guide educator decision-making as they seek to improve educational services for all students*

The problems confronting educational institutions today, at a time of massive budget cuts, are myriad and complex. Increasing pressure for accountability adds to the challenges facing school superintendents, school boards, governors, and the millions of parents who are concerned about the educational future of their children.

Amid the chaos that often characterizes debates about public education, the natural desire for a silver bullet—a single intervention or a policy or program that can trigger all the necessary changes that will improve the system for all children—is strong. Clearly, some politicians and educators thought No Child Left Behind (NCLB) might offer a version of the silver bullet: that by setting ambitious goals, offering incentives and consequences, and tying these efforts to standards-based testing, the federal government could focus local efforts and effect change. Few people who have any familiarity with NCLB retain these

early hopes.

Although NCLB seems to have failed in many of its objectives, it has brought into sharp focus a gap in the conversation about educational reform: There is no universally accepted measuring rod for gauging success that can be used to compare individuals, classrooms, schools, districts, and states. What is missing is a metric similar to those in use in other sectors of society.

In the business world, a metric is any type of measurement used to gauge some quantifiable component of a company's performance. No one metric is universally accepted across industries, but some metrics are ubiquitous because they tell managers and stockholders about the health and performance of companies from small shops to major conglomerates: return on investment, employee and customer churn rates, revenue versus debt. Other metrics summarize important economic factors: the unemployment rate, the cost of living index, a borrower's credit score, a country's GNP. What all of these metrics have in common is that they summarize key aspects of a entity (person, state, industry, company, or nation) in a single number that lets the observer know the relative success of that entity on a scale.

The established metric for NCLB—adequate yearly progress—quickly lost its credibility and value. A school that didn't meet the criterion could be put on probation, restructured, or even closed permanently. Another commonly used metric, graduation or dropout rates, has the virtue of measuring a commonly accepted benchmark, but such rates are a summary measure, not a progress report. State-mandated standards-based tests at various grades can work as a metric within a given state, but those tests differ significantly from state to state. The National Assessment of Educational Progress (NAEP) allows for state-by-state comparisons, but they aren't available below the district level.

As is true for health care and other industry leaders, education leaders at all levels want an effective metric that can measure how well they are doing and monitor their improvement across time: a metric that captures the key elements of an institution in a concise and compelling way and points toward a goal that is generally accepted, if not perfectly articulated. Even more, they want what might be termed a "magic metric"—a universally accepted metric that is brief, understandable, and measurable, but one that can be used with units in a system of similar units (e.g., retail stores or restaurants in a chain, hospitals in a network, high schools in a district) as well as the system as a whole.

Three examples of magic metrics currently in use are presented here in order to highlight the advantages and drawbacks of each. Two are state-specific metrics and the third was developed for a national study. The first state-specific index of school quality and improvement, one that has been in use for about a decade, is the California Academic Performance Index, or API. The second state-specific metric is the more recently initiated Washington Achievement Index. The third metric is used in a national study called Return on Educational Investment.

API calculates a single score for each school in California using that school's test results on the state's learning standards as they compare to prior year test results. The state's goal is a score of 800 for each school, but the actual scores in 2011 ranged from 200 to 1,000. Although API definitely has the virtue of being universally applied and concise, one parent information website, GreatSchools, noted that "Educators and parents alike struggle to understand where the API comes from, how it's calculated and what exactly it means." In fact, all it can convey is the relative position of a given school on a scale determined entirely by scores on a single set of tests. The bottom line, according to the same website: "The API is based on test scores and is calculated in a way that encourages schools to raise the test scores of the lowest-scoring students." However, nothing about the scores or their display (a list of scores in similar schools, districts, or counties) can provide any information to decision makers on what elements of school or student behavior need to be addressed to improve those scores.

Washington state has attempted to expand the usefulness of an API-type index through adoption in 2009 of the state legislature-mandated Washington Achievement Index. It is currently used to select schools for the Washington Achievement Award. The index measures how all schools and districts are performing in five key areas (reading, writing, math, science, and graduation rate). Each of the key areas are evaluated across 4 indicators (achievement by non-low income students, achievement of low income students, achievement vs peers, and improvement from the previous year). The five outcomes and four indicators result in a five by four matrix with 20 measures. Each cell of the matrix contains a score, and the index is the average of the ratings across the 20 outcomes. By using the average, schools without data for some indicators are still included in the system and a separate system is not needed for different types of schools. The Washington state index retains the benefit of a single index number (the average of scores, ranging from 1–7) and the additional virtue of including more than a single outcome measure. However, it is, again, a lagging indicator—telling us something about outcomes but providing nothing that can help to identify which aspects of student or school performance need to be addressed. For more details about the Washington Achievement Index, please visit their website (http://www.sumner.wednet.edu/studentfamilyservices/academics/pages/achievementindex.html).

The final example of a potential magic matrix was published by the Center for American Progress (Boser, 2011). Called the Return on Educational Investment, the method calculates how much learning a district produces for every dollar spent, after controlling for factors such as cost of living and students in poverty. A vast majority of public school districts in the United States were evaluated using this method, and the interactive display mechanism provides viewers with the ability to see a district's basic ROI (return on investment) compared to other districts in the state. The color-coded evaluation

matrix is used on both a state map and a matrix plotting the state achievement index against adjusted per-pupil spending (to see this interactive map, please go to http://www.americanprogress.org/issues/education/news/2011/01/19/8877/interactive-map-return-on-educational-investment/). This makes it possible to easily identify particularly effective or ineffective districts.

The education ROI provides a metric that has an appeal beyond either of the state-specific metrics discussed earlier, both because it takes into account factors that aren't directly addressed elsewhere (like cost of living and level of poverty) and uses graphic displays that illuminate relationships and encourage further investigation. The biggest disadvantage with this particular measure is that it is designed for whole districts rather than individual schools. Focus on the district level is largely due to the fact that most school districts do not calculate per-pupil spending for individual schools, clinging to the comfortable fiction that all schools in a district have essentially the same resources (see Keyworth, Detrich, and States, this volume, for an analysis of the inequities across schools within a district). However, using methods pioneered by researchers at the Center on Reinventing Public Education at the University of Washington, it should be possible to develop such per-school numbers, thus producing an ROI for individual schools (Roza & Miles, 2002; Roza & Hill, 2004).

While being more intuitively and visually appealing than the other education metrics discussed, the ROI shares a major drawback. The measure can identify districts and potentially schools that are at great risk of failing their students, but it can provide little to help decision makers at the state, district, or school level to get beyond a score that might be considered a final grade. It shows where action is needed; it doesn't provide the information necessary to diagnose the underlying weaknesses and to intervene and put the schools back on track.

IF NOT A MAGIC METRIC, THEN WHAT?

Clearly, a metric that assigns a final grade or acts as a lagging indicator can provide valuable triage information, identifying which schools are in the greatest distress and which are progressing well on their own. However, data for the next steps—intervention and remediation—must be immediately available. What would be helpful at this stage are indicators that identify schools or districts failing to meet the educational needs of their students and also provide decision makers with the information they need to address the issues that led to the failing grade. Unfortunately, these next steps have often involved the collection and presentation of masses of scattershot and unfocused data on district websites and district-produced school report cards. Numbers collected and presented in

this way provide little guidance about what the information means and what to do with it. The result, as one study of district data needs and uses (Roza, 2004) puts it, is that "Most urban cities lack the strategic information to successfully identify and implement a district reform strategy."

This paper describes a set of indicators that can serve as a guide through the data wilderness and can help school officials and community members make sense of the mountains of data. It is not so much a cookbook as it is a blueprint to action, and it begins with a more refined definition of "indicator." Innes (1990; see also Norris, Atkisson, et al., 1997; Innes & Booher, 2000) has described an indicator as "simply a set of rules for gathering and organizing data so they can be assigned meaning." They are often single items or indices of data that provide information about an underlying characteristic. The readings on automobile speedometers and gas gauges are indicators. A fever thermometer reading is an indicator. New factory orders and housing starts are indicators. So are rates of unemployment and hospital morbidity. The point is that whenever we are unable to view a large system in its totality—whether an automobile, the human body, the national economy, a local community, a hospital, or a school system—indicators can provide a general sense of how well the system is functioning. The trick is to find educational indicators that have meaning, are easy to read, have been validated by research as related to student learning, and can be presented comprehensibly in a graph or chart or in a page or two of text rather than in a volume.

In developing a set of indicators of school quality and improvement, Celio and Harvey (2005) adopted several basic principles that are explained below:

Indicators should be neither top-down (developed by experts or those in leadership) nor built from the grassroots but rather evidence-based.

There is a long history of conflict between advocates of the two traditional approaches to indicator development. Community organizations and city governments have conducted listening sessions and focus groups to identify indicators of effectiveness in different areas. On the other hand, government agencies have developed metrics that don't always make much sense to those who have to live with them (e.g., adequate yearly progress.) This question is actually moot; what is important is whether the indicators are linked by reliable evidence to the quality being sought. In other words, does the indicator measure what it is intended to measure.

As far as indicators are concerned, experience has shown that less may be more, but one is not enough.

The single metric (magic or otherwise) cannot possibly provide enough information beyond that needed to identify the schools most in need of attention. A single metric, no matter how attractive, makes it difficult to understand its implications or motivate to action. On the other hand, schools and school systems are now awash in data. A school or district report card made up of dozens of data items for dozens of subgroups of students confuses rather than enlightens; the human mind has a limited capacity to absorb unlimited data if the data are not organized in a meaningful way.

Parsimony and power must be respected.

The number of indicators can spiral out of control when developers try to cover all bases and please all stakeholders. Success rests in parsimoniously selecting a limited number of indicators and judging their power to communicate useful information plainly and succinctly.

Current status data are necessary but not sufficient.

Knowing the graduation rate of a particular school can tell you how a single group of students at that school fared, but not much else. The data are out of context. There are, of course, year-to-year fluctuations and these can confuse matters, but trend data are crucial for understanding the overall trajectory of the particular institution. Is the trend generally positive? Negative? Stable?

Proxies for key elements such as adequacy of funding or teacher effectiveness are inevitable.

It would be wonderful if school and community leadership could have immediate access to information about school-level factors that have been found to affect school quality (e.g., school culture, the effectiveness of the teachers, and the connectedness of students to the school). However, such information is either not readily or not consistently available in most districts. If an indicator has been shown to be connected to school quality but is not readily available, then it might be necessary to identify proxies for that indicator. Areas for which there are no universally accepted indicators cannot be excused from assessment and reporting for that reason.

Presentation cannot be an afterthought.

Every school board member, principal, and teacher is familiar with the reams of computer output supposedly designed to convey critical information on how students and schools are doing.

Often these reports are an excellent illustration of the truth of an observation made generations ago: "Getting information from a table is like extracting sunbeams from a cucumber" (Farquhar & Farquhar, 1891, cited in Wainer, 2004).

The heightened emphasis on school accountability, along with the realization that matrices and tables seldom have the desired impact has caused software developers to flood the market with tools for the conversion of input and output data to dashboards, sample report cards, and other display mechanisms. Many of these displays are very colorful but are not easily readable or understandable. The plethora of what Edward Tufte (1983) called "chartjunk" has inspired volumes to educate the reader in effective graphic techniques.[1] These books concentrate on presentation of content, making it clear that the way information is presented is critical to its usefulness.

THE BIRTH OF AN INDICATOR SYSTEM

Outside medicine, few fields are subject to such intense public analysis as education. Given the sheer volume of data about schools and the hundreds of articles published each year detailing evidence of school effectiveness or ineffectiveness, it should be possible to develop a parsimonious set of educational indicators that contain great power in terms of data, proxy value, and communications utility (Marzano, 2000) In fact, the proposed indicator system grew from an extensive study of the literature on school effectiveness and reform. It also included research into the ways in which school district leaders use data in making decisions. It was informed by analyses of how leaders in other areas of community life try to understand how the public institutions for which they are responsible measure up against public institutions elsewhere. Finally, it was designed to assist leaders of individual schools, school districts, and state education systems improve school management. The foundation of the indicators is work completed over a 6-year period at the University of Washington's Center on Reinventing Public Education, much of it supported by

[1] Edward R. Tufte coined this term in *The Visual Display of Quantitative Information* and expanded on the concept in later books. A good example of a workbook designed to correct for the most egregious designs is Stephen Few's I*nformation Dashboard Design: The Effective Visual Communication of Data*.

the Wallace Foundation. This work focused on the achievement gap, national and statewide dropout statistics, national studies of school superintendents and principals, and extensive work on school and school district reform, including school finance.

Based on the work described above, seven indicators were selected, for which both status and change measures were identified.

1. Student achievement (scores on standards-based math and reading tests);
2. Elimination of the achievement gap (status and change in reading and math achievement for subgroups of students by race, economic status, English language facility, etc. where there are adequate numbers within a subgroup for comparison);
3. Student attraction (ability of the school/district to attract students where there is opportunity for choice by parents/students);
4. Student engagement with school (proxy measure of school engagement, including attendance, tardiness, and involvement in school activities);
5. Student retention and completion (retention of students during the school year and completion of the requirements appropriate at each school level: elementary, middle, and high);
6. Teacher attraction and retention (proxy measure of teacher attraction using applications per job opening and non-retirement turnover); and
7. Funding equity and efficiency (proxy measure using amount of funding per student expected by policy and amount actually received; return on investment using calculated per-student funding).

A sample of the display mechanism designed for this indicator system is presented on the next page which is not unlike the displays used in consumer product review publications. Each school level in a district (in the example, all schools in the fictitious Rebel Valley School District) is rated on status and change in the seven areas. The status indicators are a snapshot of how the schools perform right now in comparison to a comparison group, in this case, other middle schools in the district.

	Indicators		Rebel Valley Elementary Schools		Rebel Valley Middle Schools		Rebel Valley High Schools	
			Compared to state	Compared to other urban districts	Compared to state	Compared to other urban districts	Compared to state	Compared to other urban districts
Status	Student achievement	Math	◒	◒	●	○	○	◒
		Reading	◒	◒	●	●	◒	◒
	Elimination of achievement gap	Math	○	◉	◒	○	◒	◒
		Reading	○	◒	◒	◒	◒	○
	Student attraction		○	○	◒	◒	○	○
	Student engagement with school		○	○	◒	◒	○	○
	Student retention/completion		◒	◒	◒	◉	◒	○
	Teacher attraction/retention		◒	◒	●	○	○	◒
	Funding equity/efficiency		✶	✶	✶	✶	✶	✶
Change	Student achievement change from 2005	Math	◒	◒	◒	○	◒	◒
		Reading	◒	◒	◒	○	◒	◒
	Reduction in achievement gap, change from 2005	Math	◒	○	◒	○	◒	○
		Reading	○	○	◒	○	○	○
	Student attraction, change from 2005		○	○	◒	◒	◒	○
	Student retention/completion, change from 2005		◒	◒	◒	◒	○	◒
	Student engagement with school, change from 2005		◒	◒	◒	○	◒	◒
	Teacher attraction/retention, change from 2005		●	◒	◒	◉	○	◒
	Funding equity/efficiency, change from 2005		✶	✶	✶	✶	✶	✶

● = In bottom 10% of comparison group
◒ = In bottom third of comparison group, but > bottom 10%
○ = Within 15% (+/-) of comparison group
◓ = In top third of comparison group, but < top 10%
◉ = In top 10% of comparison group
✶ = Not available for comparison group

Figure 1. Sample indicator system.

The reason for and brief explanation of each of the indicators is provided below.

Student achievement

Standards-based test scores have become something of a lightning rod in contemporary education. Some experts see them as necessary measures of the effectiveness of a school or school system, while others view them as a force that limits the creativity of educators and pupils and pushes students out in the

end like widgets on a production line. Whether a bane or a boon, test scores are essential to any indicator system (Wainer, 1992; Wainer & Brown, 2004).[2] Two items need brief discussion here: the use of test scale scores instead of the ubiquitous "percent meeting standard" and the reporting of only math and reading scores.

Although there is a satisfying directness in the use of a single number to characterize a given school or group of students (i.e., percent meeting standard or classified as proficient), such an approach ignores the fact that scores below or above the cutoff may be distributed in very different ways. If most of the "below standard" scores are clustered close to the cutoff point, the approach to raising achievement would be quite different from the approach required if the "below standard" scores were found primarily at the bottom end of the test score distribution. Richard Rothstein (*New York Times*, 2002) made an impassioned plea for using scale scores in reporting criterion-referenced test performance, noting that the cut-point used to determine the standard is simply a predetermined point on the scale score distribution, not a magic number. Thus, moving the cut-point in one direction or another could make a radical difference in the percent meeting standards (Shaw, 2004).[3]

A scale score is neither the raw score a student earns (i.e., the number of correct answers) or a percentage of correct answers. It is a number on a scale that is derived from the raw score but takes into account differences on the forms of the test students take. A well-known example of scale scores is the SAT. For both verbal and math portions of the test, the scale runs from 200 to 800, and the two scale scores are added together to get a total SAT score. Using scale scores and research on what different scale scores mean in terms of acquisition of required knowledge and skill, most states set two or more cutoff points along their scales, with the most important division being between students who are considered proficient and those who are not. In Washington state, the two lowest categories ("not proficient") were originally titled "well below standard" and "below standard" but are now called "below standard" and "approaching standard." The two highest classifications ("proficient") are

[2] The indicator presented in this report uses scale scores unadjusted for the racial or economic composition of the student body or the geographic location of the school. Wainer has made a strong argument that using unadjusted scores creates a situation called the Simpson Paradox, in which average scores for subgroups may actually be higher than the average for the group as a whole because the subgroups are of different sizes. When adjustments are made for the racial/ethnic constitution of the schools in the database used for the report presented here, the effect on average school scores is often considerable, with many of the differences equivalent to an effect size that would be classified as "high." Adjusting average school achievement scores for student body composition can have as much effect on apparent achievement as most educational interventions.

[3] In fact, Washington state changed the cutoff points for fourth- and sixth-grade Washington Assessment of Student Learning (WASL) scores in math and reading because of concerns that they had been set too high. Thus, without any major changes in the actual test performance of students, there was a significant jump in the number and percentage of students who met the standards at both levels and in both subjects.

labeled "meets standard" and "exceeds standard."

For leaders to see and understand how students in a school are actually performing, it is not enough simply to know how many students fall to one side or another of an arbitrarily drawn line. Using scale scores permits educational leaders to understand where their students stand as they monitor their efforts to improve achievement or close the achievement gap. "Percent meeting standard" provides no such guidance. Scale scores also permit leaders to detect change over the entire range of scores. For example, an annual increase of 2% or 3% in the proportion of students meeting standard is certainly cause for celebration, no matter how it is achieved. But if that 2% or 3% represents students already close to the standard who were levered over the bar, that is not nearly as impressive an accomplishment as if some portion of the newly successful students came from the bottom of the distribution barrel. Indeed, districts congratulating themselves on annual increases of 2% or 3% in those meeting standard might find themselves with tougher challenges ahead—if most of the students remaining below standard are substantially below the bar.

The academic subjects used in the indicator system are math and reading, generally accepted as the two basic skills without which a student is unlikely to do well on other criterion-referenced tests like writing, physical sciences, social sciences, and so forth. The correlation among the scores is very high and statistically significant.[4] It is conceivable that one of the two scores might be used by itself. However, reading and math scores together are the scores most generally accepted as meaningful.

Status*:* The circles on the display in the achievement row represent scale score data from schools based on what is known as relative distribution and density analysis. Hancock and Morris (1999) explicated in detail this method of analysis and presentation, which was developed specifically to study and report on the achievement gap between groups in society with special attention to income variations. They wanted to provide a full picture of the distribution of different measures rather than simple summary measures like means, modes, or "percent meeting standard." Since the method was specifically developed to show the relationship of one group to another (e.g., scores of Hispanic students and White students, earned incomes of male and female employees), it does not rely on the assumed distribution of scores as represented by the standard bell-shaped curve.

Change: Rate of change analysis provides information on how scores have changed over time, in our example, over 5 years. Changes from year to year are likely to be highly unstable but potentially indicative of progress toward

[4] For example, correlations among the scores on the WASL were very high for reading-math (around 0.76), but much more modest for listening-writing (around 0.40). Correlations were very similar across the grades tested (4th, 7th, and 10th) and across years of testing (1998–2003).

academic achievement across the spectrum of students (Kane & Staiger, 2001).[5] Achievement change graphs are not cohort charts, which would show progress of the same group of students as they move through school. Still, they provide a picture of what is happening within a given school building, at a particular grade level, from year to year. In this example, 2005 is the base year. Each subsequent year shows the percentage of change from 2005.

Elimination of the achievement gap

One of the great accomplishments of the accountability movement of recent years has been the insistence that data on average student achievement be disaggregated so that low achievement among particular subgroups (e.g., ethnic and low-income groups) is not concealed within overall averages. In the last 5 years, the importance of closing the achievement gap has taken on an urgency never seen before in the United States. There is no doubt that in education, the achievement gap is a sizzling hot issue. Educators need to see what is happening with respect to the achievement gap, both at the district level and within individual schools.

The achievement gap presented in this chapter defines the gap in terms of racial and ethnic groups. Those descriptors were the only student-level data available in the fictitious school district being analyzed. However, some districts also collect information on free or reduced-price lunch status and family composition (e.g., single parent household) that could be used to analyze student achievement. The same approach could be taken, independent of the descriptors used.

Status: The indicator system uses an analytic, graphic method based on the cumulative distribution of individual student scores—a method of presentation often used in such areas as medical research, marketing, and insurance. This approach has been suggested to the Educational Testing Service for use with data from the National Assessment of Educational Progress (NAEP) in the sponsoring organization's long search for effective ways to present NAEP results to the public (Olson, 2002).

Change: The gap status that can be displayed in the graph recommended by Olson (2012) shows a particular moment in time. This functions to inform educators how a specific school is performing relative to a comparison school. There are three possible outcomes from these data: The school of interest can

[5] In *Improving School Accountability Measures*, Kane and Staiger emphasized the imprecision of school-level test score means. They estimated that "28% of the variance in 5th grade reading scores is due to sampling variation and about 10% is due to other non-persistent sources and that less than half of the variance in the mean gain in reading performance between 4th and 5th grade is due to persistent differences between schools." Based on their study, I estimate that the confidence interval for the average fifth-grade reading score in a school with 60 students per grade level would extend from roughly the 25th to the 75th percentile!

narrow the achievement gap with the comparison school; there is no difference between the two schools; or the school of interest is outperforming the comparison school.

Student attraction

Makers of toothpaste, producers of TV shows, and designers of teen clothing all conduct extensive research into what makes a product attractive to potential consumers. Some public school districts, and many private schools, do similar market research to determine what parents and students are looking for in a school. In the absence of intensive polling, one way of knowing what aspects of a school are attractive to its target market is to look at families' choices when choices are available. Many public school districts now offer some level of choice for parents, ranging from magnet or alternative schools that are open (usually by lottery) to all students in the district to permitting parents to rank their school choices from among all schools within the system. Where some level of choice is available, the indicator system uses an indicator called "student attraction," which differs in definition according to district policies.

Status and change: In a school district where parents and students can designate any school in the system as first, second, or third choice and where the choices are relatively equal (e.g., every middle school is equally attractive), then the percentage of students should be comparable to the capacity of the school. Therefore, if a school has the capacity to enroll 20% of the sixth graders in the district, and all schools have the same ability to attract students, the district could expect that about 20% of the incoming sixth graders would choose that school.

What does this have to do with districts in which school choice is limited or not available at all? Even here proxies can be developed. Districts without an established policy of school choice can learn a great deal by comparing the number of children living within a school's catchment area with the number of students actually enrolled in the school. U.S. Census Bureau data, available down to the block level, are a good place to start. A large number of "missing" students may indicate that parents have taken an exit strategy, such as private school enrollment, to find appropriate schools for their children.

Student engagement with school

One of the most potent behavioral predictors of failure in school and subsequent dropping out is attendance patterns (Celio, 1989; National Research Council, 2001; Hale, 1998;). In a summary of national research on the issue,

DeKalb (1999) reported the following:

> The Los Angeles County Office of Education identifies truancy as the most powerful predictor of delinquency. … When Van Nuys, California, officials conducted a three-week sweep for truants on the streets, shoplifting arrests dropped by 60 percent (Garry, 1996). Absenteeism is detrimental to students' achievement, promotion, graduation, self-esteem, and employment potential. Clearly, students who miss school fall behind their peers in the classroom. This, in turn, leads to low self-esteem and increases the likelihood that at-risk students will drop out of school."

On the other side of the coin, a potent predictor of persistence and success in school is engagement with the school, defined as involvement in school clubs, sports, and other extracurricular activities.

Engagement has been viewed as both a cause and an effect of other risk factors, but there is clear evidence that the school itself has a strong influence on student participation and sense of belonging (Bryk & Thum, 1989; Fine, 1986; Finn & Voelkl, 1993; Johnson, Crosnoe, & Elder, 2001). An international study conducted by Programme for International Student Assessment (PISA) in 2000 was designed to assess the sense of belonging and participation of students in school.[6] In a report of the study, Willms (2003) noted that the term "engagement" as used in international research refers to "the extent to which students identify with and value schooling outcomes, and participate in academic and non-academic school activities."

Status and change: As with most assessments of engagement, the primary source of data for the PISA study was self-reported surveys of students. This is the ideal and has been used extensively by the Chicago Consortium in tracking the progress of educational reform in the Chicago Public Schools. However, few school districts possess such data, especially data that make it possible to identify trends over several years. The indicator system, therefore, relies on the use of behavioral data (attendance/absenteeism, tardiness, and membership in school-sponsored activities) to build this indicator. The information available includes average daily attendance, average class attendance (at high school

[6] The two measures used to assess the sense of belonging were based on responses to six items describing the students' personal feelings about belonging, acceptance by peers, and support from teachers, along with frequency of absence, class skipping, and late arrival at school during the 2 weeks prior to the survey. Measures used in other studies include time spent on homework, participation in classroom discussions, and involvement in sports and other extracurricular activities, but because of the need for cross-cultural applicability, these potentially more sensitive and meaningful measures weren't used in the PISA study. Instead, absenteeism was used as the most important aspect of participation.

level), percentage tardiness, and percentage of students belonging to school-sponsored activities.

Student retention and completion

The retention and completion indicator assesses "leaks" from the system at each school level—elementary, middle, and high. Here, the indicator system shows the proportion of students who enter the system at the beginning of a cycle—first grade, sixth grade, and ninth grade—and who are still in school at the end of the cycle. Although the "completion rate" commonly reported now is just for high school students, the retention and completion indicator can provide valuable information to school leaders at every grade span. If a particular cohort of students begins sixth grade together and only 70% of that number are present at the end of eighth grade, this may carry a warning to district leaders: For some reason, parents are abandoning the school and departing students are not being replaced, as would normally occur with routine movement across a district. Exploring the reason for these losses could inform school leaders about potential problems in the school. Breaking the changes down by subgroups of students could also help illuminate the situation.

Ideally, it should be possible for a school and a district to track each student through the system, letting them know whether an individual who started at a particular middle school remained there through 3 years. However, only 31 of 50 states (and the District of Columbia) have individual student identifiers, and many of these states do not yet have data systems flexible enough to track students easily or economically for the purposes of districtwide monitoring (Celio & Harvey, 2005). Thus, the ideal (i.e., knowing where students are in the system over their years of schooling) may not be practicable within the immediate future. In the absence of the ideal, what most districts do, at least for high school students, is report dropout rates. Considerable controversy has surrounded these reports, however; the way the data are collected and computed makes all the difference in their usefulness as indicators of school health.

As a measure of a school's ability to retain its students, the completion rate recommended by Greene and Winters (2002) of the Manhattan Institute has some real advantages over the traditional dropout rate. Greene demonstrated that the completion rate is straightforward and stark. The Manhattan Institute approach is used in the indicator system. The retention-completion rate measures school completion in the aggregate, comparing the number of students who graduate in a given year in a particular school, district, or state with the number who started at an earlier transition point such as 9th grade. Every state conducts some version of an October count each year, and some states complete counts at other points during the year. The data are usually broken down by gender and race, but with little additional information about

individual students.

It is unlikely that the retention and completion rates can explain much about why a school's population is increasing or decreasing over time. But the data can provide a convenient and highly useful way to determine whether changes deserve more attention. If only one or two schools are experiencing significant changes in student enrollment, leadership may need to look closely at what is going on in those schools. Changes across the district may require a more general response.

Status: In most schools, we might expect the number of students in a grade to be approximately the same each year. Barring an obvious situation such as increased housing density in a neighborhood, significant variation in grade-level enrollment might alert leadership to possible changes in the school that might negatively affect its quality and attractiveness.

Change: To see how a school's ability to retain its students changes over time, it is necessary to look at cohorts of students. The National Governors Association has recommended a cohort analysis approach to graduation rates. Most states have indicated their commitment to implementing this method. If the recommendation is implemented, many districts will have data for use in a retention and completion indicator system.

Teacher attraction and retention

Teacher effectiveness is not an element listed in the indicator system because currently no way of assessing it independent of student performance exists. There is a circular quality to many analyses of teaching effectiveness. The best teachers are identified as those whose students gain the most from their time with those teachers. To date, no research has been able to identify the characteristics that help effective teachers produce those student results. No external characteristic of the teacher—not years of teaching experience, type of certification, or having a major in the subjects taught—has been shown to be reliably related to significantly greater student achievement.

Systems that rate teacher effectiveness have traditionally looked retrospectively at teachers to see how their students did over time. Such approaches are of little help when examining an entire district to determine which schools require assistance; a concentration of less effective teachers may account for the poor performance of students, but so also may changes in demographic makeup of the school, innovative curricular approaches, or new leadership. At the moment, there is no direct way to measure the effectiveness of all teachers, or specific teachers, within a school or district. Until a measure of teacher effectiveness can be developed and added to the indicator system, teacher attraction and retention are suggested as proxies.

Although not an ideal indicator of the important role teachers play in school

effectiveness, the measurement of teacher attraction and retention is useful. Just as it would be useful to know what parents are looking for in a school before they enroll their children, it would be helpful to know how potential employees view a school before they accept positions. It would be valuable, also, to get a better understanding of what draws the most highly qualified teachers to particular schools.

When a school is perceived as unsupported, in trouble, or failing, it is unlikely that many teachers—especially those with the most experience and expertise—will be interested in applying for positions there. An unattractive school might be on either pole of several variables: a weak or a domineering principal, uninterested or overly controlling parents, extremely needy or overindulged students. Whatever the cause, teachers' perceptions of a school can play an important role in determining whether that school will attract the desired number and quality of teachers.

It should be possible to identify attractive and unattractive schools by surveying current and prospective teachers, but such an effort is costly. Another approach is to quantify teacher attraction and retention. Roza (2004) reported that calling around to a few schools in Los Angeles revealed that some schools received, at best, 1 to 3 applications per position while others had up to 130 applicants. Similar studies in the Seattle Public School District found much the same pattern: Some schools had a handful of applicants per opening while others received dozens (Roza, 2004). It should not be a surprise that schools with few applicants wind up doing the best they can when hiring while schools with many applicants can be choosier, selecting teachers with the qualifications and qualities they seek.

Attraction and retention are not the same thing. A school might attract many more applicants per opening than neighboring schools while experiencing higher than average teacher turnover, even for several years in a row. This situation could develop because potential applicants have not yet heard that the school is not a good place to work. There could also be more positive explanations, for example, a core of experienced and highly qualified teachers retiring or moving around in the system to take on master teacher or administrative roles. Either way, this indicator, like others in the indicator system, cannot provide a diagnosis or prescription, but it can act as an early warning system, alerting leaders to look closely at what is happening in a school. Also, as with other indicators, a particular rating on the teacher attraction and retention indicator can call for additional attention from the school superintendent and school board.

Funding equity/efficiency

Although the achievement gap between groups of students has received a lot of public and professional attention recently, another type of gap has elicited little comment: a funding gap between school districts and even schools within a district. A gap between districts cannot be solved by districts, but a disparity in funding among schools in the same district can be.

The Center on Reinventing Public Education (CRPE) has conducted detailed studies of budgeting practices in more than a half dozen major school districts over the past 10 years (Roza & Miles, 2002; Roza & Hill, 2004). This research revealed substantial differences in the actual (as opposed to the budgeted) funding levels of schools within each of these districts. Surprisingly, these differences are often largely invisible not just to the public eye but to the eyes of district leaders.

The differences fly beneath the radar of both district leaders and the general public because schools are "resourced" rather than funded. That is to say, they are provided with a certain number of teachers depending on enrollment and not on funds to pay for teachers, much less the categorical funding that is intended for particular groups of students such as special education, English language learners, and low socio-economic status students. The budgeting process in many districts makes it difficult to determine exactly how much funding is going to each school.

When CRPE researchers examined school funding, they found that disparities were related both to the way school districts budget for teachers (the single largest expense for schools) and the way funds do (or do not) follow the students for whom they were intended.[7] The result in all the districts studied was that the most needy schools tended to receive lower per-student funding than schools with fewer needs. In effect, schools with the most challenges were subsidizing schools with the fewest.

Status and change: The CRPE research developed two measures of funding equity that are used in the indicator system. The first uses actual versus budgeted teacher salaries and the second uses a weighted index of resource allocation to compare expected funding with actual funding for schools.

The first measure (teacher equity) compares what the district budgets for teacher salaries in a particular school with actual teacher salaries in that school. That is to say, for each school this measure compares how the district budgets its money for teachers (the district's average teacher salary multiplied by the number of teachers assigned to the school) with how it spends the money (the

[7] The CRPE researchers found that few districts have developed the capacity to track real dollar spending on a per-pupil basis, using real teacher salaries. They noted that the necessary data management and computational methods have been published in a tool kit by the Annenberg Task Force for School Communities that Work. This tool kit is designed to help analyze district data and is not overwhelmingly technical.

real salaries of teachers in the school). Some of the teachers in the school may make the minimum salary while others may be paid at the top of the scale. This measure throws light on a finance fiction—namely that budgeting (or "resourcing") schools on the basis of average teacher salaries represents what is spent on teacher salaries in individual schools.

The CRPE researchers found that every district queried about the effects of average versus actual teacher salaries was convinced that the average teacher salary within the schools would closely match the average district salary (Roza & Hill, 2004). That was not the case in any single district studied. In fact, the disparity among schools within a district in real teacher salaries amounted to a gain of as much as $1 million in some schools. This could only be made up with corresponding losses in other district schools.

This disparity might not be of great significance if all teachers possessed equal experience and ability. In practice, what happens is that very needy schools tend to be staffed largely with new and inexperienced teachers, at the bottom of the salary scale. Once those teachers get a few years of experience, they tend to take their increased capability to a more attractive school—or leave teaching altogether. In summarizing the effects, Roza and Hill (2004) concluded that "there is good reason to believe that schools with higher average salaries have more capable teachers."

The second measure of funding equity, a weighted index of expected allocation, was developed to look at how student-based budgeting would affect school-level funding (Miles, Ware, & Roza, 2003; Miller, Roza, & Swartz, 2004; Miles & Roza, 2006). If support were attached to students rather than buildings, a school serving a large number of low-income, educationally vulnerable students should receive more funding than one serving children of upper-income, professional families. The disadvantaged students not only need more assistance, but districts can draw on specific sources of state and federal funds (e.g., Title I) to help them.

As with other indicators, the funding equity indicator cannot tell how a particular situation came about or how to address it, but it can provide educational leaders with a tool that lets them see a meaningful summary of complex data. What they do with these insights is up to them.

CONCLUSIONS AND RECOMMENDATIONS

The indicators presented here can't tell school, district, and state leaders everything about an educational system, but they serve as a mechanism for providing feedback about a system that might otherwise be too large and cumbersome to understand. Like the unemployment rate, the poverty index, and the Dow Jones Average, these indicators provide insight into complex modern

systems, offering leverage points for thinking about what a large system needs when it's in distress. They also offer a center of gravity for educators and citizens faced with mountains of data. Indicators can't diagnose problems or prescribe solutions. They won't tell school superintendents, board members, or other leaders what is wrong, but they will instantly warn when something is wrong and offer those in leadership positions some preliminary information about where to begin and what to examine.

The indicator system consists of both status indicators and trend indicators in seven areas. Each indicator tells part of the story, but even taken together they cannot possibly tell the whole story. However, the indicators are based on what research tells us about school and student characteristics associated with improved educational outcomes. Some are more thoroughly researched and powerful than others, but each provides a unique piece of the story that can act cumulatively as either a wakeup call (to shock, enlighten, and jump-start) or a guide to the goal or standard to be attained, or both.

Most school districts are already collecting the data underlying these indicators. Much of that information is also available to members of the public, who are likely to find it even more difficult to comprehend than teachers and principals. The nation's educational data problem today is not that not enough data exist. Quite the contrary. The problem is that educators and parents are awash in data they find hard to understand. The indicator system described here promises to create a center of gravity for data usage, a focal point around which to organize data so as to identify both critical problems and promising opportunities.

REFERENCES

Boser, U. (2011). *Return on educational investment: A district-by-district evaluation of U.S. educational productivity*. Center for American Progress.

Bryk, A. S., & Thum, Y. M. (1989). The effects of high school organization on dropping out: An exploratory investigation. *American Educational Research Journal*, *26*(3), 353–383.

Celio, M. B. (1989). *Falling through the cracks: A study of dropouts in the Seattle Public Schools.* A Report for the Seattle–King County Economic Development Council.

Celio, M. B., & Harvey, J. (2005). *Buried treasure: Developing a management guide from mountains of educational data*. Seattle, WA: Center on Reinventing Public Education. http://www.crpe.org/publications/buried-treasure-developing-management-guide-mountains-school-data

Council of Chief State School Officers. (2006) *Implementing the NGA graduation rate compact: State-level issues.* http://www.dataqualitycampaign.org/files/Publications-Implementing_the_NGA_Graduation_Rate_Compact_State-level_Issues.pdf

DeKalb, J. (1999). Student truancy. *ERIC Digest*, #125. (ED429334.)

Few, S. (2006). *Information dashboard design: The effective visual communication of data*. Sebastopol, CA: O'Reilly Media.

Fine, M. (1986). Why urban adolescents drop into and out of public school. *Teachers College Record*, *87,* 393–409.

Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research*, *59*(2), 117–142.

Finn, J. D., & Voelkl, K. E. (1993). School characteristics related to student engagement. *Journal of Negro Education*, *62*(3), 249–268.

Garry, E. (1996). *Truancy: First step to a lifetime of problems*. Washington, DC: U.S. Department of Justice, Office of Juvenile Justice and Delinquency Prevention. (ED 408 666.)

GreatSchools. http://www.greatschools.org/students/local-facts-resources/31-the-api-10-things-parents-should-know.gs

Greene, J. P., & Winters, M. A. (2002, November). *Public school graduation rates in the United States* (Civic Report No.31). New York: The Manhattan Institute for Policy Research.

Hale, L. (1998). *High school dropouts cost everyone something!* Bethesda, MD: National Association of School Psychologists.

Handcock, M. S., & Morris, M. (1999). *Relative distribution methods in the social sciences*. New York: Springer-Verlag.

Holland, P. W. (2002). Two measures of change in the gaps between the CDFs of test-score distributions. *Journal of Educational and Behavioral Statistics, 27(*1), 3–17.

Innes, J. E. (1990). *Knowledge and public policy: The search for meaningful indicators*. (2nd ed.). New Brunswick, NJ: Transaction Publishers.

Innes, J. E., & Booher, D. E. (2000). Indicators for sustainable communities: A strategy building on complexity theory and distributed intelligence. *Planning Theory and Practice*, *1*(2), 173–186.

Johnson, M. K., Crosnoe, R., & Elder, G. H. (2001). Students' attachment and academic engagement: The role of race and ethnicity. *Sociology of Education 74*, 318–340.

Kane, T. J., & Staiger, D. O. (2001). *Improving school accountability measures*. (Working Paper #8156.) Cambridge, MA: National Bureau of Economic Research.

Marzano, R. J. (2000). *A new era of school reform: Going where the research takes us*. Aurora, CO: Mid-continent Research for Education and Learning.

Miles, K., & Roza, M. (2006). Understanding Student-Weighted Allocation as a Means to Greater School Resource Equity. *Peabody Journal Of Education, 81*(3), 39-62.

Miles, K. H., Ware, K., & Roza, M. (2003). Leveling the playing field: Creating funding equity through student-based budgeting. *Phi Delta Kappan*, *85*(2), 114–119. http://www.kappanmagazine.org/content/85/2/114.abstract

Miller, L., Roza, M., & Swartz, C. (2004). *A cost allocation model for shared district resources: A means for comparing spending across schools.* Working Paper. Seattle, WA: Center for Reinventing Public Education.

National Research Council. (2001). *Understanding dropouts: Statistics, strategies, and high-stakes testing.* A. Beatty, U. Neisser, W. T. Trent, & J. P. Heubert, J. P., Eds. Board on Testing and Assessment, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Norris, T., Atkisson, A., et al. (1997). *Community indicators handbook: Measuring progress toward healthy and sustainable communities*. Oakland, CA: Redefining Progress.

Olson, L. (2002). Testing experts develop new method of presenting achievement gap data. *Education Week, 21*(26), 11.

Rothstein, R. (2002, May 22). Lessons: Testing reaches a fork in the road. *New York Times*. http://www.nytimes.com/2002/05/22/nyregion/lessons-testing-reaches-a-fork-in-the-road.html

Roza, M. (2004). Rethinking data capacity. In P. T. Hill & H. James (Eds.), *Making school reform work: New partnerships for real change* (pp. 105–106). Washington, DC: Brookings Institution Press.

Roza, M., & Hill, P. (2004). How Within-District Spending Inequities Help Some Schools to Fail. *Brookings Papers On Education Policy*, 201-227.

Roza, M., & Miles, K.H. (2002). *A new look at inequities in school funding.* Seattle, WA: Center on Reinventing Public Education.

State of Washington, Office of Superintendent of Public Instruction. School Achievement Index. http://www.k12.wa.us/accountability/

Shaw, L. (2004, September 2). Improvement on WASL carries asterisk. *Seattle Times*.

Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.

Wainer, H. (2004). *Graphic discovery: A trout in the milk and other visual adventures*. Princeton, NJ: Princeton University Press.

Wainer, H. (1994). On the academic performance of New Jersey's public school children: Fourth and eighth grade mathematics in 1992. *Education Policy Analysis Archives*, *2*(10).

Wainer, H., & Brown, L. (2004). Two statistical paradoxes in the interpretation of group differences: Illustrated with medical school admission and licensing data. *The American Statistician, 58,* 117–123. http://www.statlit.org/pdf/2004wainer_threeparadoxes.pdf

Willms, J. D. (2003). *Student engagement at school: A sense of belonging and participation*. Results from PISA 2000. Organisation for Economic Co-operation and Development.

Chapter 4

Are We Making the Differences That Matter in Education?

AMANDA VANDERHEYDEN
Educational Research and Consulting, Inc.

ABSTRACT: *This paper argues that ineffective practices in schools carry a high price for consumers and suggests that school systems consider the measurable yield in terms of gains in student achievement for their schooling effort. Student performance data can be used to evaluate efforts, set instructional targets, and plan instructional changes. The routine reporting of student achievement gains is also a very powerful way to solve the tension that can exist between stakeholders in the schools and those who must run the schools using limited resources. The data are transparent for all to interpret and any corrective action can be evaluated by all interested parties. This paper contends that the structure of multi-tiered intervention services or response to intervention (RtI) systems in schools (that is, screening, providing intervention, and monitoring progress to verify that the interventions worked as planned) offers great opportunity for determining whether or not an educational effort changes the odds of student success. Schools can and should examine whether the use of assessments and interventions in their schools reduces risk of learning failures over time for all students and for students who are thought to be especially vulnerable. Given the historically great investments that have been made in education and the current economic climate pushing for spending reductions, policy makers and local decision makers must avoid the "more is better" logic and instead seek information about which investments (assessments, interventions) yield the greatest return in student learning. Systems must also consistently engage in those actions that are demonstrated to yield a high return in terms of student learning.*

THE HIGH COST OF INEFFECTIVE PRACTICES

By my calculations, in my own son's school, taxpayers spent about $20 per student per day of the school year last year. That is an extraordinary sum of money when you stop to think about it. Between 1970 and 2007, average spending in the United States increased from $4,210 to $10, 041 per pupil per year (Snyder & Dillow, 2010). Given the well-documented history of rapidly increasing investments in public education, it is puzzling that so many Americans suggest that perhaps we are not spending enough money. Spending more money is often offered up as the best solution for making our schools more effective. What seems lost in the debate over whether we are spending too much or too little is the relevance of what we are spending the money on. There seems to be little attention paid to whether or not we are funding what works.

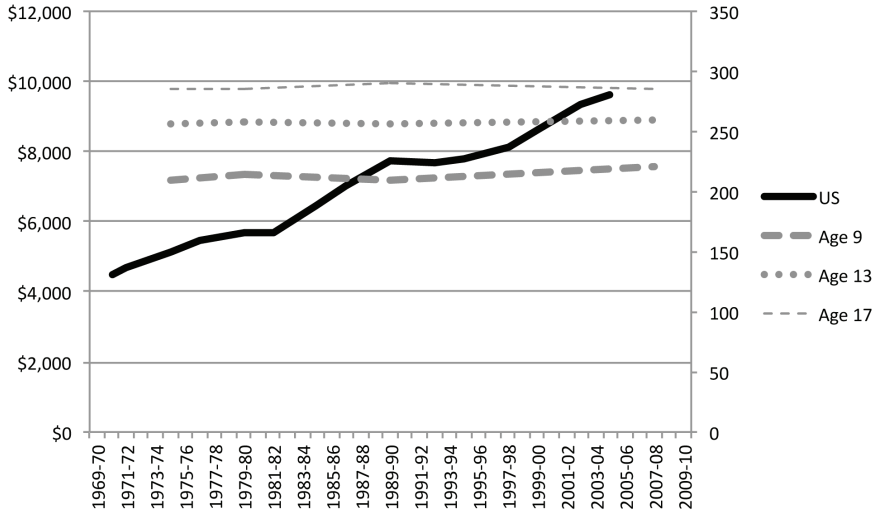

Figure 1. The relationship between spending and reading scores.

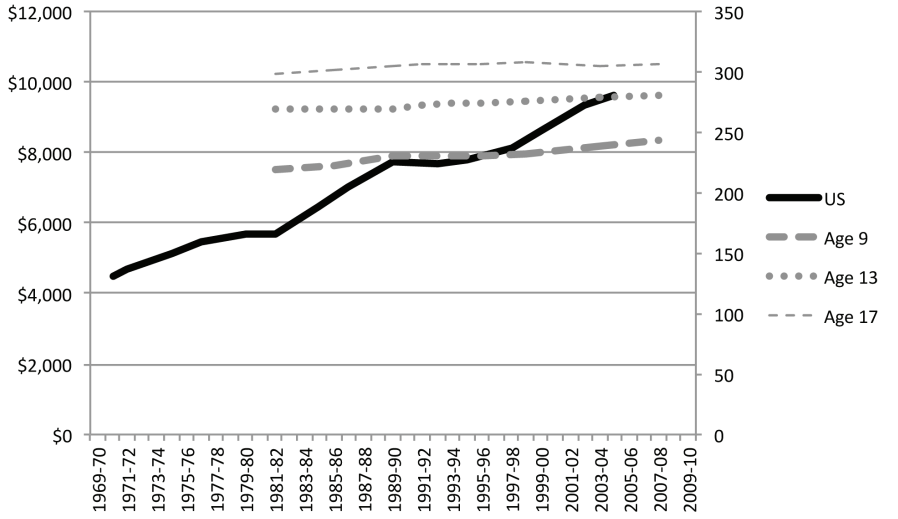

Figure 2. The relation between math scores and spending.

Part of the trouble with having a discussion about whether we are funding what works is that the goals of educational efforts have often been poorly defined (Bushell & Baer, 1994). That is, we have not determined what it means when schooling is successful or what results we want. So we place undue focus on the process and almost completely neglect the outcome, and this leads us to superficial solutions that are not necessarily causally related to improved student outcomes (e.g., reducing class size, increasing time in school). The focus on process and the neglect of outcome have fueled debate that is often filled with vitriol and passion, but is of little use to the students the debaters are supposed to be concerned about.

Focusing on process as opposed to outcome has also fueled tension between school systems and parents. Mistrust is bred and communication suffers when parents approach the school with an agenda of wanting a particular service and view the school as either giving them what they want or not. Similarly, schools may be slow to share information with parents because they do not want parents to interfere with the process they wish to use. This arrangement does not effectively serve the goal of improved learning for students.

One very common scenario involves parents pushing for a special education eligibility evaluation and special education services for their child who is struggling to learn to read, because the parents believe that special education will lead to a better outcome for their child. Unfortunately, this belief does not reflect the realities of special education in schools; special education has not

been shown to have a significant effect on the learning outcomes of students served under the category of Specific Learning Disability (Kavale & Forness, 1999). Another example involves communities advocating for smaller class size. On the surface, smaller class size sound good to everyone, but when you view the available resources as a single pie that must be divided for the greatest good, then something that seems desirable might not be worth the cost if it means not being able to implement another strategy that has been shown to improve achievement. When resources are allocated to one effort, they are not available for another effort.

Student outcome data can take the heat out of these debates about resource allocation, because any resource allocation decision simply becomes a hypothesis to be tested, and the action will be continued only if it returns the desired results. Using student data to inform resource allocation decisions has a number of important effects. First, it focuses decision makers on attaining improved learning outcomes. Second, it increases the probability that the decisions will favor actions that have been shown to successfully improve learning in the past or in other schools. Third, and perhaps most important, it creates an opportunity for decision makers to make midstream adjustments to implemented strategies to ensure that they return the desired effect. Selecting something that is likely to work is a good first step, but once something is implemented, the most important function of leadership is ensuring that desired outcomes are reached and sustained over time.

Student learning is the most fundamental outcome of schooling

Student learning is the outcome that schools and communities should prioritize. The purpose of the school is to ensure learning. This purpose is not at odds with big-picture questions that parents might care about: "Is my child happy at school?" "Does my child like learning?" "Is my child developing positive relationships with teachers and students and learning how to function well away from our home environment?" Rather, learning and growth of students is a powerful—perhaps the most powerful—indicator the school is a healthy, productive environment that supports students in engaging in learning tasks they can successfully complete (Hattie, 2009). Being successful at learning in school fosters a sense of well-being in the student and improves parent-school bonding. When resource allocation is based on data and the effective actions are emphasized, precious time is preserved and is thus available for socialization, recreation, and rest during the school day. Children have a right to effective instruction and a well-rounded schooling experience that fosters the development of the whole child. Many argue that effective instruction is the best path to that end (Barrett et al., 1991).

Focusing on a simple, measurable outcome like learning gives consumers, teachers, and other decision makers a compass. Thus, activities that promote learning become valued activities that warrant further investments of time and resources. Activities that do not promote learning receive less priority. Measuring process targets such as number of hours allocated to math instruction is much less meaningful and direct than tracking whether or not students are learning and growing in math proficiency. The value of the school's effort can be evaluated in terms of student skill proficiency, growth in achievement over time, and reduction of performance gaps between groups of students at baseline or when instruction begins. The yield of the effort can be computed as the positive effect on learning divided by the cost in per-pupil spending.

Selecting achievement as the fundamental outcome of schooling is logical and viable (Barrett et al., 1991; Hattie, 2009). Consensus for achievement as the primary outcome has emerged over the past three decades and is reflected in policy efforts promoting standard content expectations (National Mathematics Advisory Panel, 2008; National Reading Panel, 2000) and accountability legislation (No Child Left Behind Act [NCLB], 2001) that are intended to demonstrate that educational services enhance student outcomes over time. Research trends reflect a shift from correlational (where conclusions about causal relationships cannot be reached) to experimental (where conclusions about causal relationships can be reached) research, and syntheses of existing research studies (e.g., Hattie, 2009; Slavin & Lake, 2008) provide excellent direction for practitioners who wish to use educational strategies that will be of highest yield for students.

Bad decisions are not benign

When decisions are made to allocate educational resources in ways that do not yield achievement gains, the cost is greater than consumers might suspect. When a school chooses to use an ineffective strategy, it bears the tangible cost of materials and training for the new strategy. But the cost does not stop there. There is also the cost of lost opportunity to do something that would have better served the achievement goal; for example, lost instructional time, teacher absence from the classroom to participate in professional development for the new strategy, and substitution of the new strategy for an existing strategy that may have been higher yield. But perhaps the greatest cost comes in creating a legacy in the school that teachers will be asked to use unproven strategies, and when those strategies fail the program will be abandoned and replaced with a new initiative. This approach creates a culture of "attempt-attack-abandon" (D. Deshler, personal communication, August 23, 2008) that is highly detrimental to a productive program-improvement system that all schools and districts should use. All educators and community stakeholders must

understand that bad decisions (i.e., decisions to allocate resources in ways that do not return the desired effect) are not benign and can result in an apathetic teaching environment in which teachers just push through some new effort until it is replaced by some new mandate.

In Figure 3, an example of bad decision making is shown. In this particular school, a decision was made to implement a new mathematics program just as the school was experiencing a strong upward trend in mathematics achievement. Of course, the data below are not experimental and no causal conclusions should be reached about the efficacy of the new program, but the need for a new program can and should be rightfully questioned when achievement is trending upward. Similar mistakes happen with great frequency in systems where decision makers decide to adopt a new program without local evidence to show that it can work to serve the needs of students in the district or even that the new program fits the needs of the school.

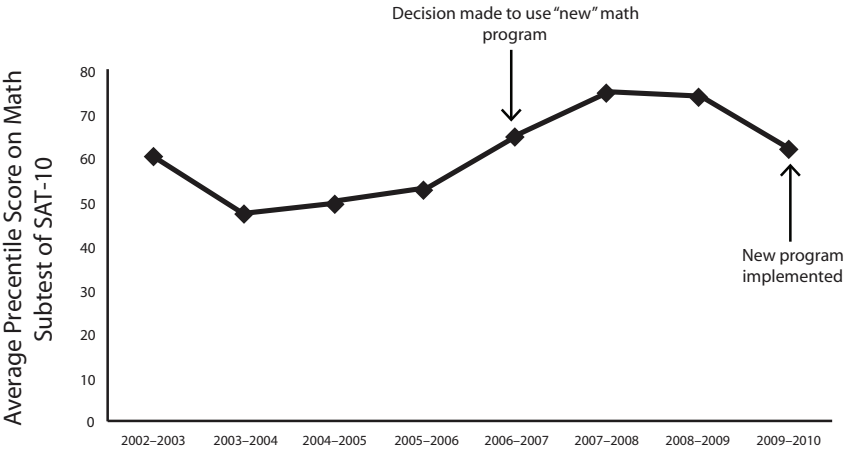

Figure 3. Third-grade math achievement.

Response to Intervention (RtI)

The use of student performance data, collected during the course of instruction, is an ideal basis for determining where resources are needed to improve learning outcomes. Systems of using student performance data to make resource allocation decisions that improve learning for the greatest number of students

are referred to as response to intervention systems. RtI is not a product. It cannot be purchased. It is a decision-making process that uses student performance data as the ever-present arbiter of all instructional decisions. Teaching can be like flying at night in poor visibility without navigational instruments to tell you how far you are from the ground, how far you are from your target, and whether or not you are moving toward or away from your target. When there is no easy way to monitor the effects of instruction and make adjustments, the likelihood is high that the instructor will miss the target altogether for many of the students.

The use of student performance data as a basis for evaluating instructional efforts, setting instructional targets, and planning instructional changes is also a very powerful way to solve the tension that can exist between stakeholders in the schools and those who must run the schools using a limited number of resources. The data are transparent for all to interpret, and any corrective action can be evaluated by all interested parties.

The questions that should guide instruction at the classroom, school, and district level are:

1. Are we making the differences that matter? Are we changing the odds of student success?

2. If we are not changing the odds of student success, what are we going to do about it?

These questions have quantifiable answers. I would like to return to the questions at the end of this article after first describing how student performance data can be used to improve learning through RtI and then describing two of the most common barriers to the effective use of RtI.

RtI has enjoyed widespread popularity as a framework for using student performance data to set system improvement targets and attain system improvements. Implementers must have data to determine risk; identify systemic problems; plan instructional changes systemwide; plan interventions for individuals, small groups, and whole classes as a supplement to core instruction; and evaluate intervention effects and inform future resource allocation decisions. RtI can be used to reduce unnecessary evaluations, initiate and sustain instructional changes that produce the desired improvements in learning, and improve learning outcomes for all students. Most states report partial or full implementation of RtI. However, certainly not all implementations return similar results. Unfortunately, the potential for improved results is often lost to implementation errors. In the section that follows, I will discuss two of the most common errors made during RtI implementation.

Weighing a cow does not make it fatter (or the prevalent problem of overassessment)

It is amazing to consider that 10 to 20 years ago assessment of student learning occurred only rarely. Now children participate in a great deal of assessment. There is no doubt that assessment is necessary to improve learning, and therefore it is not surprising that nearly all prominent policy documents related to improving outcomes in education feature routine student assessment as an essential recommendation (e.g., National Reading Panel, National Mathematics Advisory Panel). Installing technically adequate and well-implemented student assessments into schools is the first stage of RtI implementation. This effort has been speeded along in most school systems via requirements and funding provided through a statewide Reading First initiative and year-end accountability assessment. Many instructional products integrate student assessment into their materials and procedures. Hence, routine assessment of student performance is now commonplace in most schools.

Assessment is absolutely essential to make instructional decisions that improve instructional targets, but *too much assessment is detrimental to instructional systems because assessment alone will never improve achievement*. Frequent assessment is useful when it leads to a different instructional action the next day as a result of the assessment. That is, when the data are used formatively, it is reasonable to expect achievement gains (Fuchs, Fuchs, Mathes, & Simmons, 1997; Yeh, 2007). Yeh computed effect sizes for student achievement under the following conditions: frequent student assessment (two to five times per week), 10% increase in per-pupil spending, voucher programs, charter schools, and increased accountability. He then computed the cost for each approach. Frequent student assessment was 4 times as effective as increased spending per pupil, 6 times as effective as vouchers, 64 times as effective as charter schools, and 6 times as effective as increased accountability, even after accounting for the increased costs associated with conducting frequent assessments. Hattie (2009) found that formative evaluation is one of the most reliable and powerful ways to improve student achievement, yielding an average effect of $d = 0.90$ among the 30 studies included in his analysis.

From the list of actions that must be performed well to use RtI, screening and progress monitoring are tasks that most schools do not struggle to implement (Burns, Peters, & Noell, 2008). In other words, most schools implementing RtI are able to accurately collect screening and progress-monitoring data. Teams struggle to "do something" with the data. That is, they struggle to interpret, plan, and deploy corrective actions, and to evaluate and troubleshoot those actions (Burns et al., 2008). Perhaps the relative ease and competence with which schools collect assessment data contribute to the error of overassessment. There is a tendency among school leaders to think if some is good, then more must be better. When I work with schools and districts to build an action plan

to improve achievement, often the first suggestion from the leadership team is to obtain more assessment data. This suggestion is so prevalent that I routinely use a slide entitled, "Schools are drowning in data and the same students still cannot read."

I recently worked with a kindergarten center and calculated for the teachers how much time they were allocating to assessment. I wish all school and district leadership teams would undertake this exercise. I find that what administrators and teachers say they will do instructionally often is not possible given the available hours of instruction. At the kindergarten center, for example, children attended about 180 days of school. If 6 hours were used solely for instruction in all of those 180 days, then teachers had about 1,080 hours of usable instructional time for the year. Teachers reported spending about 120 hours assessing skills over four reporting periods (two report cards and two midterm reports), 10 hours per year screening, 15 hours per year monitoring progress for low-performing students, and 6 hours per year on end-of-unit tests. Hence, teachers were spending a total of 151 hours per class per year on assessment activities. If we assume that teachers were using 100% of the balance of available time for instruction (which is not possible because teachers must leave some time for transitions, non-instructional routines, and enrichment), then they were spending 14% of available instructional time on assessment. Whether or not this allocation of resources to assessment is an investment that is well spent is a question for which there is a definitive answer, but few schools seem to raise the question or look at their data in this way.

All schools should list all assessments used in the school, identify the decision that will be made from each assessment, and determine which assessments are redundant and which are not actually contributing data needed to inform instructional actions. Overassessment is a costly error that comes at a direct and substantial cost to instruction.

Implementation failures are sentinel events but usually go undetected in education

In medicine, the term "sentinel event" is defined as "an unexpected occurrence involving death or serious physical or psychological injury, or the risk thereof. Serious injury specifically includes loss of limb or function. The phrase, 'or the risk thereof' includes any process variation for which a recurrence would carry a significant chance of a serious adverse outcome. Such events are called 'sentinel' because they signal the need for immediate investigation and response." (The Joint Commission, 2011). The aviation community closely examines failures with the explicit goal of preventing those failures from occurring in the future. Defining and attending to events that come at a high

cost to the stated goals of a profession (e.g., death or injury that could have been prevented in medicine, where the goal is to promote health and well-being) is a testament to the commitment of a profession to attain its stated goals. It is not pleasant to acknowledge, let alone study, our failures, but education would do well to follow the examples of medicine and aviation.

One of the most common RtI failures involves overemphasizing intervention selection and underemphasizing intervention management (VanDerHeyden & Tilly, 2010). In RtI, every decision and action leading up to intervention may occur perfectly, but if the intervention is not implemented correctly for a consistent period of time, the intervention will fail and student learning will not improve. Research tells us that intervention failures should be exceedingly rare events (Torgesen et al., 2001; VanDerHeyden, Witt, & Gilbertson, 2007). A high rate of intervention failure is a sure sign of intervention implementation error. Many research teams have highlighted the persistent challenge of deploying interventions accurately and effectively outside of research settings (Fixsen & Blase, 1993; Witt, Noell, LaFleur, & Mortenson, 1997). Despite these data, research teams have also documented the careless disregard most interventionists and intervention researchers pay to monitoring the degree to which an intervention was correctly implemented when reaching a conclusion about the intervention's effectiveness (Gresham, Gansle, & Noell, 1993; McIntyre, Gresham, DiGennaro, & Reed, 2007).

The lack of attention to implementation accuracy is puzzling given the likelihood that poor intervention integrity will threaten or weaken intervention results and lead to inaccurate conclusions about an intervention's value in improving learning for a student or a class of students. Integrity failures are sentinel events in education. It is a sentinel event because the decision errors lead directly to the allocation of unneeded additional resources, the abandonment of a strategy that might have worked had it been implemented correctly, the use of more costly and probably more restrictive interventions for the student, and an inaccurate belief about a child's capability for learning.

One important lesson from implementation research is that often implementations fail for seemingly simple reasons that would be relatively easy to address if only someone were paying attention to the indicators. Common causes of implementation failure include not having ongoing access to a person who knows how to implement an intervention, the child not being available for intervention sessions due to scheduling problems, intervention error (e.g., modeling too rapidly, failing to give corrective feedback to the student), not having the right materials available, a belief on the part of the implementer that implementation is not being tracked and is not important, and no one tracking and troubleshooting intervention effects. It is important to remember that intervention failures should be rare events. Hence, a very simple approach to monitoring integrity is to track student learning outcomes. Where student

learning outcomes are not improving, implementation error should be investigated and ruled out or repaired before changing the intervention (Gilbertson, Witt, Singletary, & VanDerHeyden, 2008; Witt, VanDerHeyden, & Gilbertson, 2004).

CONCLUDING COMMENTS: ARE WE MAKING A DIFFERENCE?

We know a great deal about how to improve instruction and learning (Hattie, 2009). When we know what works to improve achievement, why do so many school systems struggle to put these strategies into practice in classrooms? I believe our failures have had little to do with measurement or pedagogy or many of the other causes we tend to focus on and discuss. I believe we have consistently failed to use data to guide instruction and then deliver that instruction well. When children fail to learn the skills we expect them to learn, our strong tendency, historically and persisting today, is to attribute those failures to them (Ysseldyke, Algozzine, & Epps, 1983).

To attain improved learning outcomes, implementers should use student performance data to guide resource allocation decisions. RtI systems provide an excellent framework for doing so, but the results obtained depend entirely on how well the system is followed. Implementers must minimize assessment requirements, collecting only the data needed to make the instructional change that will move the students and school system closer to its targets. Implementation must be monitored closely to ensure that the decisions made are high yield. For implementers, smarter decision making will allow them to work with greater quality, intensity, and consistency because they can discontinue unnecessary and unfruitful efforts.

In Figure 4, the progress of a whole class of students can be tracked to ensure that learning gains are being made toward the instructional goal. Students who lag behind once the class as a whole reaches mastery can be provided with small-group or individual intervention.

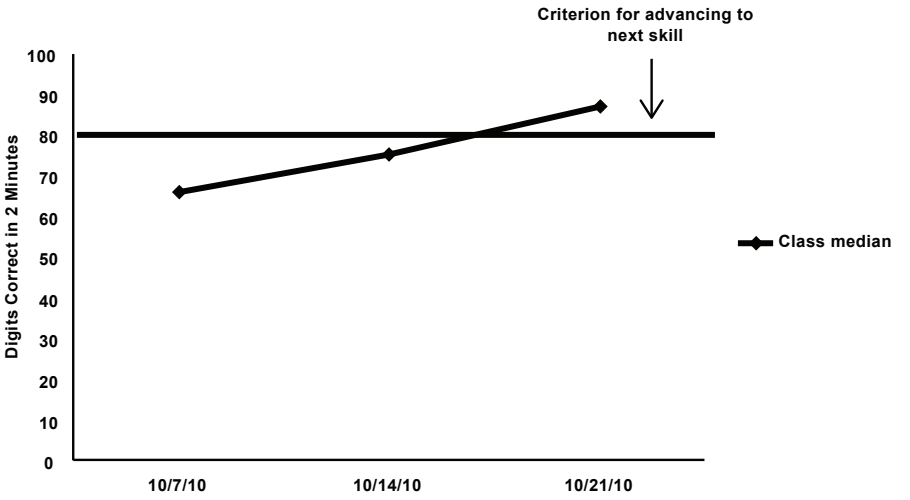

Figure 4. Digits correct with mixed addition and subtraction.

Where progress is monitored consistently across classrooms (e.g., where several classwide learning problems have been detected), the rate of mastery of skills can be tracked across classes to identify classes whose scores are lagging behind other classes participating in similar instruction or whole-class supplemental intervention. An on-site support person (e.g., coach, resource teacher) can go to those classes that are lagging, observe instruction, coach the teacher, and provide performance feedback to improve the efficacy of instruction. In this example (Figure 5), classes 9 to 11 should receive in-class support to improve the efficacy of the intervention.

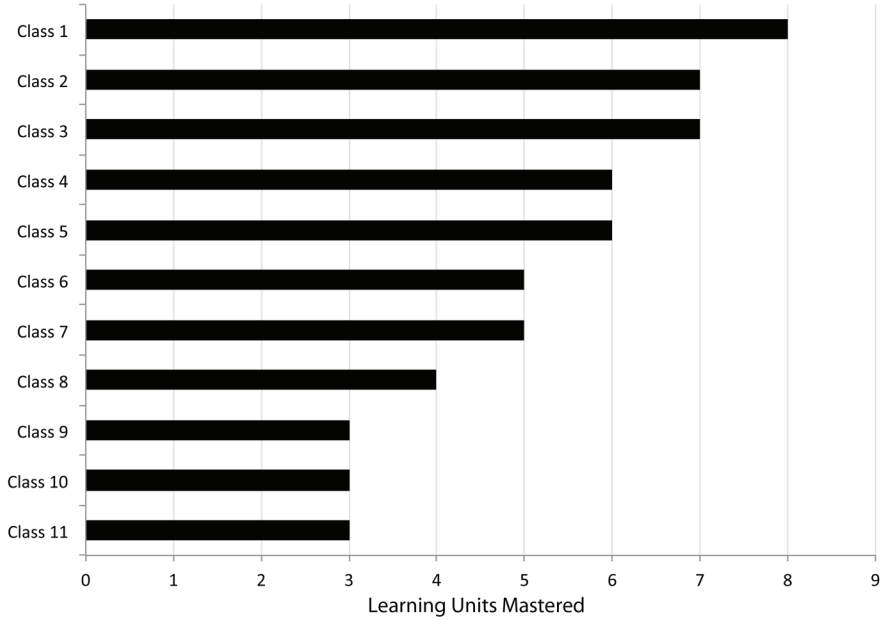

Figure 5. Learning units mastered across classes.

Follow-up screening data can be used to verify that over time instructional efforts are reducing the number of students at risk (Shapiro & Clemens, 2009). In Figure 6, each pair of bars shows the fall screening and winter screening for each teacher at first grade. The dark gray portion of the bar shows the percentage of students performing in the frustrational range, the white portion the percentage of students in the instructional range, and the light gray portion the percentage of students in the mastery range. This type of graph is highly useful to school and district leaders in generating an action plan for improvements (e.g., providing whole-class intervention versus small-group) and for evaluating and troubleshooting the improvement efforts that are already underway (e.g., giving Teacher 6 in-class support because his class is the only one that did not show a marked reduction in the number of students scoring in the frustrational range from fall to winter screening).

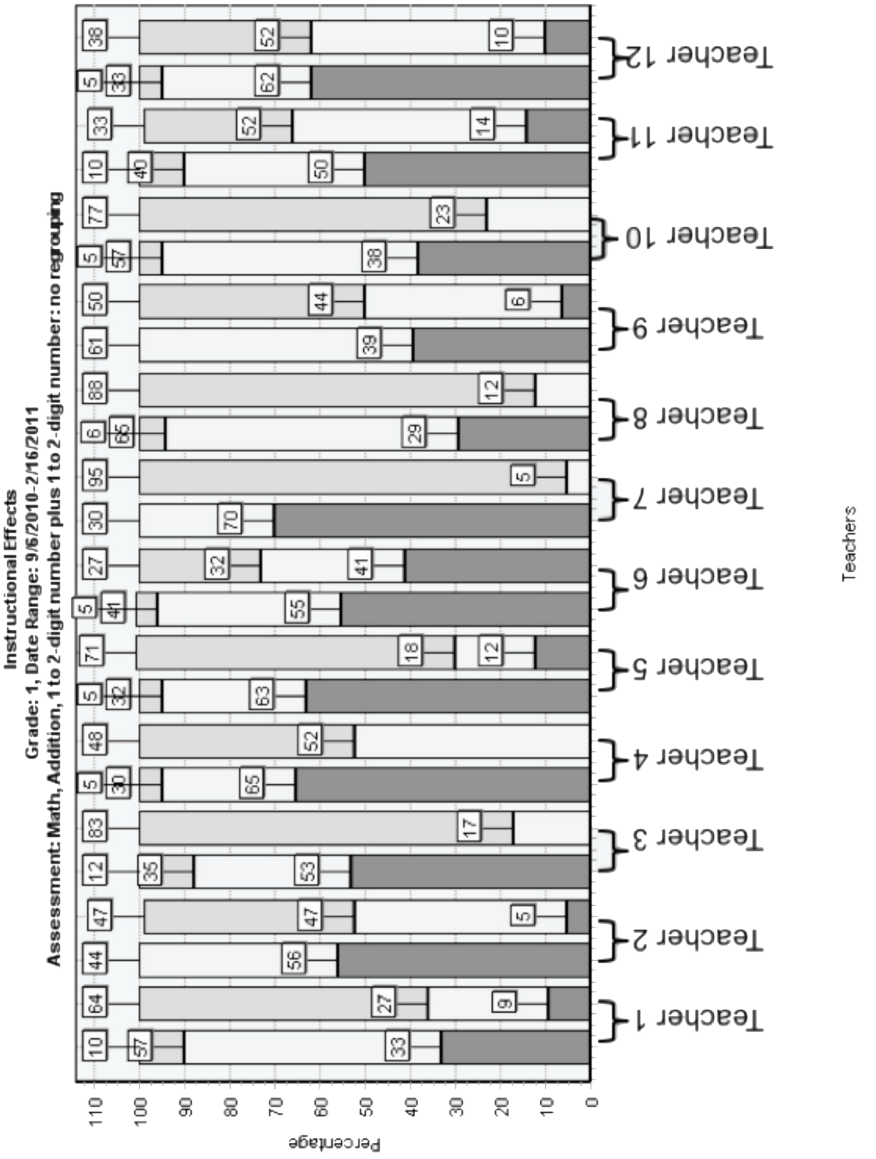

Figure 6. All first grade classes fall and winter screening for reading fluency.

We must have data to answer the two key questions raised earlier in this paper: Are we making differences that matter? And if we are not making a difference, what are we going to do about it? Figures 4 to 6 illustrate the type of data that can be collected to plan corrective actions and to evaluate and ensure the success of those actions over time. With data, any strategy can be tested

and those data can be shared with stakeholders in ways that are transparent and help everyone understand the basis for future actions and resource allocation decisions.

If we consider education as a good or service for which cost and effect can be quantified, we can track the yield of our efforts over time. In Figure 7, it is easy to compare the probability of reading success (dark gray area) and failure (white area) with supplemental intervention (left-hand column) and without supplemental intervention (right-hand column). The top row shows the probabilities of outcomes in a high-achieving school, and the bottom row the probabilities of outcomes in a low-achieving school. The probability of reading proficiency in schools with intervention is greater in both high and low achieving schools. However, decision makers must also consider the cost of providing intervention. Intervention value can be examined by computing yield per cost in each school, with and without supplemental intervention.

In high-achieving schools, the probability of passing the year-end test is 0.80 without supplemental intervention. When supplemental intervention is added (at a cost of 1.5 times the cost of general instruction or 30 instructional minutes added to 60 minutes provided during core instruction), the yield per cost is computed as 0.80 (0.80/[(1 x 100)/100]) probability of passing the year-end test in the school without supplemental intervention and 0.82 (0.90 / [(1.5 x 20) + (1 x 80)/100]) probability of passing the year-end test in the school with supplemental intervention. These data help decision makers understand that the added cost of intervention may be worthwhile since it increases the probability of reading proficiency even after accounting for the cost of providing the intervention.

In the low-achieving schools, the yield per cost of intervention analysis makes the decision very straightforward. If intervention is provided to 50% of students in the low-achieving school, then the yield per cost (expressed as the probability of reading proficiency) is 0.68 (0.85/[1.5 x 50) + (1 x 50)/100] which is superior to the probability of reading proficiency when no intervention is provided (0.60), even after accounting for the added cost of intervention. When only 20% of students are provided with the supplemental intervention (for example, when the system makes efforts to improve the efficacy of core instruction prior to beginning supplemental intervention), then the yield per cost analysis provides stronger evidence of value (0.77 with intervention versus 0.60 without intervention).

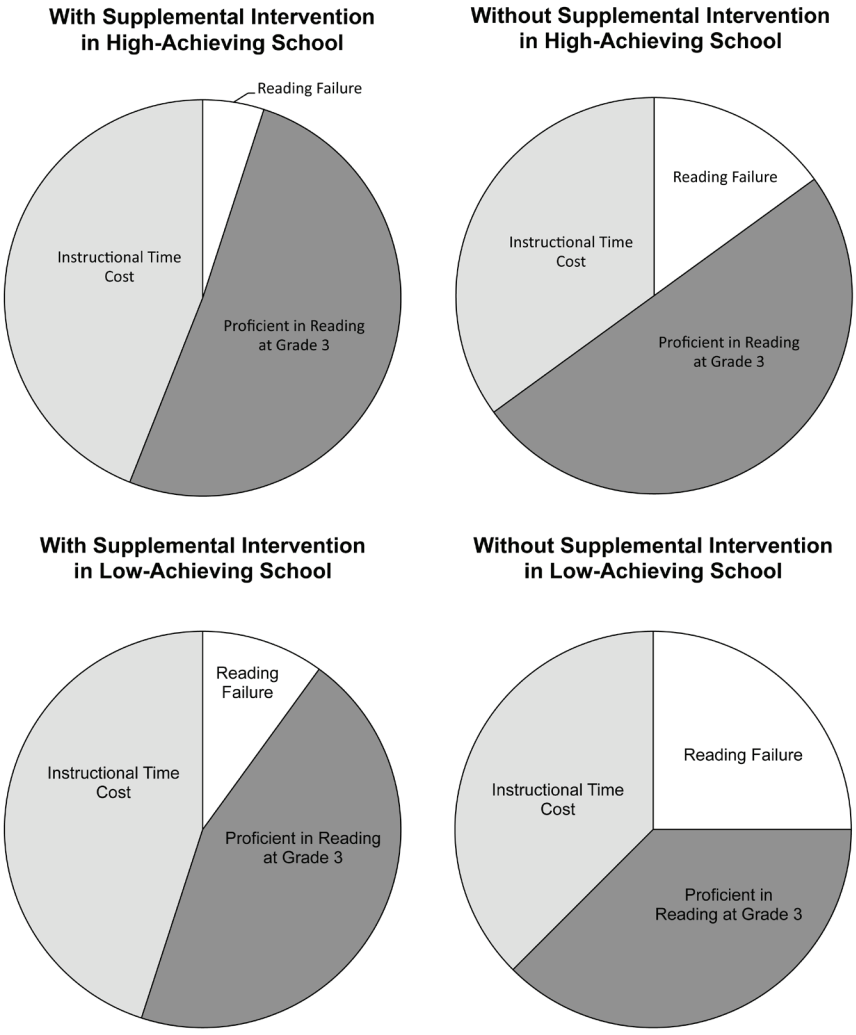

Figure 7. The probability of Grade 3 reading proficiency with and without intervention.

RtI data can be used to advance student outcomes if decision makers collect only the data that are needed to make instructional adjustments, make those adjustments with fidelity, and track their implementation to avoid common implementation errors. A controversial article appeared in the *New York Times* under the title "Can Cancer Ever Be Ignored?" (Brownlee & Lenzer, 2011). It was written in response to an expert medical panel's opinion to not recommend routine prostate screening. At the heart of this issue is the near-universal

belief that if some is good, more must be better—and consequently America's demand for more medical screening and treatment. In medicine, this belief is so prevalent that anything less than patients getting the list of diagnostic tests that they want has been described as "un-American" and compared to rationed health care (Brownlee & Lenzer, 2011).

Consumer-driven assessment and intervention in medicine actually makes for bad medicine because it is equivalent to overassessment and overtreatment. To wit, the rate of false positive errors associated with prostate screening is so high that being exposed to the screening can do more harm than good. Why? Because a positive screening leads to a more invasive medical procedure that can cause impotence, incontinence, or death. And the chances of actually dying of prostate cancer are very low among those who have prostate cancer. Overall, there is a greater risk of harm in having the screening than in not having the screening. Of course, for the individual with an aggressive prostate cancer, early detection matters, but looking for this individual among symptom-free adults causes more overall harm than good. Americans don't need more diagnosis and intervention. We need smarter diagnosis and intervention.

Smart diagnosis and intervention must be guided by four types of data:

1. The prevalence of a condition. This prevalence is the basis for computing the odds of a person having or not having a condition before any assessment or intervention is begun. These odds can be adjusted to reflect increases or decreases in odds given certain symptom profiles with the logic that if a person has a symptom, then the odds of having the condition may be higher, thus changing the utility of various assessment and treatment options. So in the case of prostate screening, determining when screening is likely to be a risk worth taking is a highly individualized decision that probably should be made only after an individual has experienced symptoms that increase that person's probability of having the condition.
2. The associated false positive and false negative error rates of screening measures.
3. The probability of negative outcomes if the condition is not diagnosed and no treatment is provided.
4. The probability of negative outcomes if the condition is diagnosed and treated.

The same scenario has been playing out in education for about 30 years. When a child struggled to learn to read, parents advocated for and sought a battery of psychoeducational assessments and an ultimate diagnosis of learning disability. This diagnosis skyrocketed 260% between 1977 and 2001, hitting a peak in 2001 when 6.1% of American students were identified as having

a learning disability. There were many reasons to question the validity of making a learning disability diagnosis (see VanDerHeyden & Burns, 2010, for a review), but in the end the proliferation of the diagnosis reflected the public's demand for more diagnosis and intervention, and the failure of the diagnosis to change student outcomes has caused people to reconsider the value of making the diagnosis in the first place.

RtI has been touted as an alternative to a learning disability diagnosis that carries the potential for making appreciably positive changes in student learning outcomes over time. When implemented well, RtI can lower false positive errors and reduce the risk of long-term learning failures. Yet, smart decision making is required or RtI may go the way of prostate screening. Implementers must understand that more is not always better and that all decisions carry errors that can and should be quantified to guide future decision making. In RtI, screening should be used only if its use increases the odds of accurate identification of learning problems above those obtained by chance (or prevalence alone).

Universal screening measures commonly used in RtI (e.g., reading curriculum–based measurement and the Dynamic Indicators of Basic Skills) often carry high false positive error rates. Follow-up assessment procedures that can be used in RtI implementations offer a low-risk and practical way to reduce the rate of false positive screening errors. More assessment of *all* students does not improve the accuracy of screening decisions. Rather, gated screening procedures are supported where the sample is filtered and subsets of the original sample participate in additional assessment. Schools implementing RtI can and should examine whether the use of assessments and interventions reduces risk of learning failures over time for all students and for students who are thought to be especially vulnerable. Given the historically great investments that have been made in education and the current economic climate pushing for spending reductions, policy makers and local decision makers must avoid the "more is better" logic and instead seek information about which investments (i.e., assessments, interventions) yield the greatest return in student learning.

REFERENCES

Barrett, B. H., Beck, R., Binder, C., Cook, D. A., Engelmann, S., Greer, R. D., …Watins, C.L. (1991). The right to effective education. *The Behavior Analyst, 14*, 79–82.

Brownlee, S., & Lenzer, J. (2011, October 5). Can cancer ever be ignored? *New York Times.* Retrieved from http://www.nytimes.com/2011/10/09/magazine/can-cancer-ever-be-ignored.html?_r=1

Burns, M. K., Peters, R., & Noell, G. H. (2008). Using performance feedback to enhance implementation integrity of the problem-solving team process. *Journal of School Psychology, 46*, 537–550.

Bushell, D., Jr., & Baer, D. M. (1994). Measurably superior instruction means close, continual contact with the relevant outcome data: Revolutionary! In R. Gardner, III, D. M. Sainato, J. O. Cooper, T. E. Heron, W. L. Heward, J. Eshleman, & T. A. Grossi (Eds.), *Behavior analysis in education: Focus on measurably superior instruction* (pp. 3–10). Monterey, CA: Brooks/Cole.

Fixsen, D. L., & Blase, K. A. (1993). Creating new realities: Program development and dissemination. *Journal of Applied Behavior Analysis, 26*, 597–615.

Fuchs, D., Fuchs, L. S., Mathes, P. G., & Simmons, D. C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal, 34*, 174–206.

Gilbertson, D., Witt, J. C., Singletary, L., & VanDerHeyden, A. M. (2008). Improving teacher use of interventions: Effects of response dependent performance feedback on teacher implementation of a peer tutoring intervention. *Journal of Behavioral Education, 16*, 311–326. doi:10.1007/s10864-007-9043-0

Gresham, F. M., Gansle, K. A., & Noell, G. H. (1993). Treatment integrity in applied behavior analysis with children. *Journal of Applied Behavior Analysis, 26*, 257–263.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.

The Joint Commission, (2011), *Sentinel Event*, Retrieved from http://www.jointcommission.org/sentinel_event.aspx

Kavale, K. A., & Forness, S. R. (1999). Effectiveness of special education. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 984–1024). New York: Wiley.

McIntyre, L. L., Gresham, F. M., DiGennaro, F. D., & Reed, D. D. (2007). Treatment integrity of school-based interventions with children in the *Journal of Applied Behavior Analysis* 1991–2005. *Journal of Applied Behavior Analysis, 40*, 659–672.

National Mathematics Advisory Panel (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*, U.S. Department of Education: Washington, DC.

National Reading Panel (2000). Report of the National Reading Panel. *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

No Child Left Behind Act of 2001. Public Law 107–110, 107th Cong., 115 Stat. 1425.(2002) (enacted).

Shapiro, E. S., & Clemens, N. H. (2009). A conceptual model for evaluating system effects of response to intervention. *Assessment for Effective Intervention, 35*, 3–16. doi: 10.1177/1534508408330080

Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. *Review of Educational Research, 78,* 427–515.

Snyder, T. D., & Dillow, S. A., (2010). *Digest of Education Statistics 2010*. U.S. Department of Education: Washington, DC. Retrieved from http://nces.ed.gov/pubs2011/2011015.pdf

Torgesen, J., Alexander, A., Wagner, R., Rashotte, C., Voeller, K., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities, 34*, 33–58. doi:10.1177/002221940103400104

VanDerHeyden, A. M., & Burns, M. K. (2010). Essentials of response to intervention. In A. S. Kaufman & N. L. Kaufman (Series Eds.), *Essentials of psychological assessment*. Hoboken, NJ: Wiley.

VanDerHeyden, A. M. & Tilly, W. (2010). *Keeping RtI on track: How to identify, repair, and prevent mistakes that derail implementation*. Horsham, PA: LRP Publications.

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of a response to intervention (RTI) model on identification of children for special education. *Journal of School Psychology, 45*, 225–256. doi:10.1016/j.jsp.2006.11.004

Witt, J. C., Noell, G. H., LaFleur, L. H., & Mortenson, B. P. (1997). Teacher use of interventions in general education settings: Measurement and analysis of the independent variable. *Journal of Applied Behavior Analysis, 30*, 693 696. doi:10.1901/jaba.1997.30-693

Witt, J. C., VanDerHeyden, A. M., Gilbertson, D. (2004). Troubleshooting behavioral interventions: A systematic process for finding and eliminating problems. *School Psychology Review, 33,* 363–383.

Yeh, S. S. (2007). The cost effectiveness of five policies for improving student achievement. *Journal of Evaluation, 28*, 416–436. doi:10.1177/1098214007307928

Ysseldyke, J. E., Algozzine, B., & Epps, S. (1983). A logical and empirical analysis of current practice in classifying students as handicapped. *Exceptional Children, 50,* 160–166.

About the Wing Institute

The Wing Institute is a 501(c)(3) nonprofit operating foundation dedicated to the promotion of evidence-based education policies and practices. It was founded in 2004 and named after Ernie Wing, an outstanding special education advocate who was an early champion of evidence-based education and quality services for children.

The Wing Institute is a "catalyst" organization, designed to bridge the research-to-practice gap by supporting and accelerating collaboration among researchers, educators, policy makers and consumers across disciplines. Its goal is to accumulate and disseminate existing knowledge, help create new knowledge, and facilitate the effective application of that knowledge to real-world settings.

It achieves this through the following strategies:

Evidence-Based Education Knowledge Network. The Institute operates an interactive, Internet-based Knowledge Network designed to link education stakeholders interested in promoting evidence-based education. It facilitates the open exchange of ideas, resources, expertise, and support.

Information Clearinghouse. The Institute's Web site (www.winginstitute.org) houses an online information clearinghouse that reviews research, analyzes current issues and policies, provides links to related organizations and activities, and includes resources for education stakeholders.

Professional Forums. The Institute organizes, sponsors, and facilitates professional summits, seminars, and conferences to encourage dialogue, collaboration, and communication among researchers, decision makers, and educators across disciplines and organizations.

Publications. The Institute researches and disseminates journal articles, policy analyses, research reviews, books, newsletters, position papers, and tools related to application of evidence-based education to assist education decision makers.

Research. The Institute supports, promotes, reviews, and completes research in education policies and practices, including funding for graduate research.

Public Policy. The Institute engages in general public policy activities such as analysis, research, and development.

The Wing Institute is a "catalyst" organization, designed to bridge the research-to-practice gap by supporting and accelerating collaboration among researchers, educators, policymakers, and consumers across disciplines. Its goal is to accumulate and disseminate existing knowledge, help create new knowledge, and facilitate the effective application of that knowledge to real-world settings.

It achieves this through the following activities:

Evidence-Based Education Knowledge Network
Information Clearinghouse
Professional Forums
Publications
Research
Public Policy

The Wing Institute is a 501(c)(3) nonprofit operating foundation dedicated to the promotion of evidence-based education policies and practices. It was founded in 2004 and named after Ernie Wing, an outstanding special education advocate who was an early champion of evidence-based education and quality services for children.

The papers in this book are derived from the Wing Institute 2011 summit, "Performance Feedback: Using Data to Improve Educator Performance." Our annual summits are organized around topics critical to the Wing Institute roadmap for establishing evidence-based education. The four points of the roadmap are efficacy research (What works?), effectiveness research (When does it work?), implementation of evidence-based interventions (How do we make it work?), and monitoring (Is it working?). Each paper in this book has been selected to address fundamental issues relevant to understanding the vital importance of feedback in the development of quality education.

Advances in Evidence-Based Education Volume 3
ISSN: 1941-1693 ISBN: 978-0-9800177-2-4
Published by The Wing Institute | Fall 2013
www.winginstitute.org