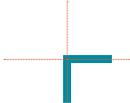


# Hopes, Fears, & Reality



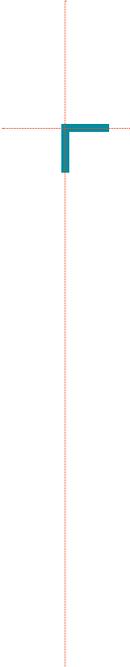
A BALANCED LOOK AT AMERICAN  
CHARTER SCHOOLS IN 2006

Robin J. Lake & Paul T. Hill, *Editors*

*National Charter School Research Project  
Center on Reinventing Public Education  
Daniel J. Evans School of Public Affairs  
University of Washington*

DECEMBER 2006





# About NCSR

The National Charter School Research Project (NCSR) brings rigor, evidence, and balance to the national charter school debate.

NCSR seeks to facilitate the fair assessment of the value-added effects of U.S. charter schools and to provide the charter school and broader public education communities with research and information for ongoing improvement.

NCSR:

- Identifies high-priority research questions.
- Conducts and commissions original research to fill gaps in current knowledge or to illuminate existing debates.
- Helps policymakers and the general public interpret charter school research.

The Project is an initiative of the Center on Reinventing Public Education.

We thank our current and past funders for their generous support:

- Achelis & Bodman Foundations
- Annie E. Casey Foundation
- Daniels Fund
- Thomas B. Fordham Foundation
- Bill & Melinda Gates Foundation
- The Heinz Endowments
- Jaquelin Hume Foundation
- Ewing Marion Kauffman Foundation
- Pisces Foundation
- Rodell Charitable Foundation
- Walton Family Foundation

Our advisory board guides the selection and methodology of NCSR research:

- Julian Betts, University of California, San Diego
- Susan Bodilly, RAND Education
- Anthony Bryk, Stanford University
- Lisa Coldwell O'Brien, Coldwell Communications; New York Charter School Association
- Abigail Cook, Public Policy Institute of California
- Jeffrey Henig, Columbia University
- Gisele Huff, Jaquelin Hume Foundation
- Christopher Nelson, Pisces Foundation
- Michael Nettles, ETS
- Greg Richmond, National Association of Charter School Authorizers
- Andrew Rotherham, Education Sector; Progressive Policy Institute
- Priscilla Wohlstetter, University of Southern California

## CHAPTER 6

# A Better Way: Measuring Charter School Success and Failure

*Laura S. Hamilton and Brian M. Stecher*

Whether charter schools are improving achievement is a subject of much debate among researchers, yet the criteria for measuring changes in academic achievement—namely, student scores on district or state tests—are rarely debated. Since state and district standardized tests typically provide the most readily available measures of student achievement, it is hardly surprising that they have become the near-universal metric of evaluation. In the current era of standards-based accountability and the No Child Left Behind (NCLB) law, achievement tests aligned with state content standards are arguably an appropriate way to measure school effectiveness at meeting generally agreed-upon goals.

At the same time, the reliance on standardized achievement tests provides at best an incomplete understanding of how any school is affecting the students it serves. One can both recognize the utility of tracking test scores while at the same time appreciate that test scores do a poor job of reflecting some important outcomes, such as preparing students to enter college and the job market. In the case of charter schools, there is even more reason to believe that test scores are a necessary but not sufficient measure of school effectiveness. Charter schools were originally envisioned by many reformers as laboratories for innovation. They were designed, that is, not only to buttress traditional goals but also to promote broader ones, reach underserved populations, and otherwise experiment with new pedagogical approaches.

In the pages that follow, we argue that relying exclusively on test scores as metrics of success provides an incomplete understanding of school performance. Indeed, the singular reliance on tracking test scores could paradoxically create incentives that will reduce

*Standardized achievement tests provide an incomplete understanding of how a school affects the students it serves.*

the validity of those scores for evaluating school performance. We propose instead a set of measures that could be incorporated into a more comprehensive system of indicators of school effectiveness for both charter and traditional public schools.

## THE GOALS OF PUBLIC SCHOOLS

---

Charter schools, like all public schools, are charged with promoting a variety of outcomes deemed important to society. Specifically, public education is supposed to produce well-informed, productive, and civic-minded adults. Broadly speaking, these three adjectives represent the key capabilities that public schools are entrusted to develop and that people use to judge schools' effectiveness.

A universal definition of our first attribute—the “well-informed” student—will always prove elusive. Yet most Americans would agree that well-informed students have learned how to read well and compute efficiently, possess basic knowledge about science, history, and government, and are informed about music and art. The term “achievement” is widely used to describe this broad class of outcomes—that is, what students know and are able to do in school subjects.

For many people, achievement is the most important outcome of charter schools, and “achievement” is often considered synonymous with “test scores.” In fact, an alternative way to measure student progress is in terms of “attainment.” As students mature, they pass various milestones that provide indirect indications of their achievement, and information about these milestones can be used as alternatives or complements to test-based data. Students who are promoted from one grade to the next on schedule, complete enough years of high school mathematics courses to fulfill the state requirements, and graduate from high school all demonstrate indirect evidence of meeting educational goals. When schools' attainment criteria for coursetaking, promotion, and graduation are combined with mastery of academic standards, measures of attainment successfully supplement test scores as indicators of achievement.

Schools are also supposed to help students become “productive” adults who can develop worthwhile careers and become contributing members of society. In addition to academic skills, productivity requires the development of career-related skills and less easy-to-measure attributes, such as the ability to communicate effectively and work in teams. There is no simple term to describe the productive student, analogous, say, to achieve-

ment for the well-informed student. But most descriptions of the broad class of productive outcomes fall under the heading of “preparation for postsecondary education and employment.”

Finally, schools are charged with enabling students to become “civic-minded” adults. Civic-minded students are familiar with the history of the United States, know and endorse the principles embodied in our founding documents, and respect public institutions. The concept of civic-mindedness is difficult to define and even harder to measure. Despite these difficulties, researchers should not overlook civic-mindedness when evaluating the impact of charter schools—indeed, promoting a sense of civic obligation in youth is one of the original justifications for public education in the United States.

The remainder of this chapter examines these three broad outcomes in greater detail and describes criteria that could be part of a more comprehensive system of indicators for charter schools as well as for traditional public schools. We close with a brief discussion of other features of charter and traditional schools that might be considered “leading indicators” of achievement, either because they provide necessary conditions for promoting achievement or because they are strongly predictive of academic performance.

## THE LIMITED UTILITY OF STANDARDIZED TESTS

---

There are many advantages to using scores from national, state, and district-wide standardized tests for comparing the academic performance of large samples of students over time. Less well known are the disadvantages of relying exclusively on these tests. The disadvantages include:

**IN MOST STATES AND DISTRICTS, ONLY A SUBSET OF GRADES AND SUBJECTS ARE TESTED.** Financial, administrative, and legal constraints on testing often preclude school officials and researchers from obtaining useful information on student achievement in the earliest elementary grades or attainment in social studies or the arts. These omissions are particularly problematic for secondary schools, which emphasize a wide variety of subjects other than reading and mathematics. Another limitation associated with the range of grades tested under NCLB is that in many cases the testing fails to provide information about student growth during the entire time students are enrolled in a

school. For instance, at the elementary level, the typical NCLB testing schedule provides no information until the end of third grade.

**MOST TESTS EMPHASIZE THE LOWER-LEVEL SKILLS THAT ARE EASIER TO MEASURE USING MULTIPLE-CHOICE OR SHORT ANSWER ITEMS.**

Even when states claim that they have verified the alignment between tests and standards, these tests are capturing only a subset of the content contained in the standards.<sup>1</sup> Moreover, the match between curriculum and tests is often weak. For example, a rise in scores on a general mathematics achievement test in high school is unlikely to reflect the full extent of what was learned by students enrolled in geometry or other higher-level mathematics courses.

**HIGH-STAKES TEST SCORES CAN BECOME INFLATED OVER TIME.** A large body of research suggests that attaching high stakes to test scores can lead to a phenomenon known as “score inflation,” whereby apparent gains in test scores overstate actual improvement in achievement.<sup>2</sup> This problem occurs if teachers shift their instruction to focus only on tested material in the format used by the test rather than the full domain of knowledge the test is supposed to represent, or when teachers devote excessive time to test preparation. In addition, NCLB’s penalties for schools with significant numbers of students who test below the proficient level may encourage reallocation of teachers’ attention to students who are close to proficient to nudge them over the threshold, potentially distorting the meaning of proficiency and judgments based on it.

**TEST SCORES CANNOT EASILY BE COMPARED ACROSS JURISDICTIONS.** There is currently no measure of achievement that can provide good national estimates of charter school effectiveness. Any effort to combine information across jurisdictions using different tests will need to address differences in content, format, difficulty, stakes, and other characteristics of the tests and state and local accountability systems.

At best, relying solely on test scores to measure achievement provides an incomplete understanding of a school’s impact. At worst, the singular reliance on test scores can provide a severely distorted view of school effectiveness.

One way to address the limitations of existing standardized tests is to combine information from these tests with information from other available measures of student achievement. These might include district-administered assessments that are not part of the state or district accountability system, interim or benchmark assessments, or stu-

dent work samples that are gathered in a systematic way. Other tests, such as college admissions tests or Advanced Placement Exams, are typically taken by only a subset of the student population but might be useful for assessing some aspects of charter school achievement. However, all of these additional measures have limitations, including the selective nature of the population of students who take some of these tests, the lack of consistent measures over time for some tests, and the lack of standardized administration conditions, particularly for interim tests and work samples.

## MEASURING OUTCOMES OTHER THAN ACHIEVEMENT ON TESTS

Although the primacy of achievement test scores in most charter school studies is understandable and generally appropriate, researchers can compile a fuller picture of the educational effects of charter schools by examining other indicators. The indicators listed here do not cover the full scope of the three broad goals listed earlier, but are limited to attributes that show the most promise for being measured feasibly and accurately. The relevance of these outcomes to charter school effectiveness may vary in elementary and secondary schools, but most charter school families are likely to consider them important measures of success.<sup>3</sup> It would also be important to collect the same information from traditional public schools, both to ensure the availability of appropriate comparison data and to hold traditional schools accountable for the same broad set of outcomes that are being measured in charter schools. Alternate, supplementary measures for evaluating the effectiveness of charter schools include the following<sup>4</sup>:

### ATTAINMENT

- **Graduation rates.** The likelihood that a student will receive a high school diploma is arguably one of the most important academic outcomes to consider when examining charter school impacts. Although graduation is clearly most relevant for high school students, it might also become a long-term indicator of success in elementary and middle school.
- **Retention/promotion rates.** Examining student retention and the characteristics of students who are held back is helpful for understanding how charter schools affect educational attainment—and could be important for interpreting test-score trends. Promotion rates are likely to vary across states and districts, in part as a function of policy differences surrounding promotion criteria.
- **Transfers to other schools.** Although transfers might not be considered an outcome of interest for most schools, the numbers of students who transfer out of a

*The likelihood that a student will receive a high school diploma is one of the most important outcomes to consider when examining charter school impacts.*

school, and the types of schools into which they transfer (for example, alternative schools), are relevant for understanding how charter schools affect their students.

### PRODUCTIVITY

- *Enrollment in college-preparatory or advanced coursework.* One measure of a high school's contribution to the development of productive adults is the percentage of students who complete the courses required to qualify for college admission. High schools can accelerate students' progress through college by offering advanced coursework, such as Advanced Placement or International Baccalaureate classes. Enrollment in advanced coursework at the secondary level can also be considered a useful proxy for tracking the development of productive students by elementary and middle schools.
- *Participation in college-admissions testing programs (SAT, PSAT, ACT).* The percentage of students who take admissions tests provides additional information about the extent to which schools are producing students who expect to pursue postsecondary education.
- *College readiness.* One simple measure for evaluating if students who graduate from a particular charter school are adequately prepared for postsecondary education is to track whether those students enroll in remedial coursework in college.
- *Postsecondary educational attainment.* A critical outcome of K-12 schooling for both parents and policymakers is where students go to college after completing high school. Although the data needs are daunting, several states are developing monitoring systems that will permit some tracking and analysis of the proportion of students who attend two- and four-year colleges, the percentage who eventually receive degrees, the quality of institutions attended, and the specific degree programs pursued.
- *Employment and earnings.* Roughly one third of high school graduates choose not to attend college immediately after graduating from high school.<sup>5</sup> For these students, researchers and others would benefit from having data on the types of careers they pursue and the amount of money they earn. Employment and earnings could also be examined for students who do attend postsecondary institutions. At present, a few states are able to link school attendance records with state unemployment insurance files to track employment status and earnings.
- *Enrollment in occupational/vocational programs.* Many students benefit from taking occupational and vocational courses while in high school. For example, among students who go directly into the labor market, those who have taken vocational courses achieve higher wages.<sup>6</sup> In addition, many of the students who enroll in college have taken vocational technical courses. Vocational coursetaking provides another indicator of a school's contribution to the eventual productivity of its students.

One simple measure for evaluating if charter school graduates are adequately prepared for postsecondary education is to track whether those students enroll in remedial coursework in college.

### CIVIC-MINDEDNESS

- *Civic values.* Some critics of school choice fear that public schools like charters that depart from the neighborhood school model will produce citizens who are less civic-minded and community oriented. Civic values and attitudes such as tolerance and patriotism have been measured in a variety of school choice studies.<sup>7</sup> These measures could be used as a source of information about civic outcomes of charter schools.
- *Civic actions.* Similarly, it might be possible to measure the extent to which charter school students or graduates engage in activities that demonstrate civic participation, such as voting or volunteering.

### “LEADING INDICATORS” OF CHARTER SCHOOL PERFORMANCE

In addition to developing some alternative criteria for assessing charter school outcomes, researchers could also create a system of “leading indicators” of charter school performance that contribute significantly to the success or failure of charter schools. These leading indicators are not measures of outcomes as such, but are germane nonetheless to evaluating charter school performance. Researchers should not revert wholesale to analyzing inputs and processes in charter schools. Yet selected aspects of school structure and process can shed light on differing outcomes among charter schools or between charter and traditional public schools. Researchers, for example, could consider the following:

#### STRUCTURAL ELEMENTS<sup>8</sup>

- *Safety.* Unsafe and dangerous schools threaten students’ well-being and interfere with their learning, so it is appropriate to measure whether charter schools offer safe havens for learning. Several surveys and other data collection techniques have been developed to assess the severity of threats to student safety, including the availability of alcohol and drugs and the presence of threats, bullying, and intimidation.<sup>9</sup>
- *Teacher quality.* Researchers cannot define with certainty the characteristics of effective teachers, but they do know that good teachers are critical to student achievement. At a minimum, studies of charter schools should determine whether teachers have knowledge in the subject(s) they teach. Research evidence suggests that subject matter knowledge is an important characteristic of effective teachers, particularly at the secondary level.
- *Class size.* There is strong experimental evidence that class size matters in student learning, particularly in the early grades.<sup>10</sup> Care needs to be taken that mea-

*Selected aspects of school structure and process can shed light on differing outcomes among charter schools or between charter and traditional public schools.*

asures of class size reflect the actual number of students in each classroom rather than the overall pupil-to-teacher ratio.

- *Grade configuration.* Most public schools are divided into elementary schools (grades kindergarten through fifth), middle schools (grades six through eight), and high schools (grades nine through twelve). Alternative arrangements, such as K-8, are preferred by some educators and parents because they require students to go through fewer transitions and are thought to offer more positive environments for learning.<sup>11</sup> Different grade configurations are important distinguishing features of some charter schools.

### PROCESS MEASURES

- *Exposure to content.* Students do not learn course content that they have never seen, so tracking exposure to content can reveal telltale information about student outcomes. At the elementary level, exposure to content has been measured through teacher reports of content coverage and reviews of curriculum materials. At the secondary level, exposure can also be measured in terms of access to, and participation in, courses and course sequences that lead to mastery of advanced content.
- *Time on task.* The amount of learning time in the school day is a strong predictor of achievement. Time on task can be measured broadly in terms of the length of school day and year, but more sophisticated measures would track the time students spend engaged in learning activities.
- *Instructional support.* Learning is facilitated by a variety of supporting materials and equipment, including textbooks and supplemental learning materials, supplies and equipment for experimentation, libraries with current reference materials, access to the Internet and online resources, and supplemental staff with expertise in science, mathematics, or other complex subject matter. All these types of learning supports can be measured with relative ease.
- *Attendance.* Students who are absent from school are unlikely to learn, and sustained poor attendance is associated with poor academic performance. Large differences in attendance rates are good predictors of academic outcomes—and attendance data are easy enough to obtain from existing records.
- *Participation in athletic and arts programs.* Participation in athletic and artistic programs are considered intermediate outcomes because they may lead to higher achievement and mastery of skills that have career implications. Athletic and artistic performance opens the way to work and careers for some students; in addition, these activities foster other desirable attributes, such as perseverance, discipline, and the ability to work in teams.
- *Parent satisfaction.* Charter schools depend on parent satisfaction for their existence, and it seems sensible to include measures of satisfaction as an indicator of how well schools are meeting the needs of students and families. Monitoring the

existence and size of wait lists would provide one indication of how satisfied parents are with the school's offerings.

## DISCUSSION

It is unrealistic to expect that all or even most of the data highlighted in the preceding pages will be available at charter schools in the near future. Nonetheless, enriching and expanding the availability of high-quality data would increase educators, parents, and voters' access to indicators of charter school performance, and the promise of a better-informed future makes it worth thinking about what a comprehensive indicator system *should* include.

Policymakers confront a number of obvious obstacles to creating such a system. Perhaps the most transparent obstacle is a lack of data. Some of the outcomes and processes discussed here (like civic-mindedness) are rarely measured, and when they are measured, they may not be measured well. Other outcomes and processes might be measured—but we lack the data infrastructure to link these measures to other student information in a way that will allow us to interpret them accurately. To cite one example, developing data systems that track students from the K-12 system into college and the workplace is an especially challenging endeavor, though some states are beginning to tackle the problem.

A second concern stems from the well-known problem that performance measures are often corrupted, particularly when high stakes are attached to them. This problem was discussed earlier in the context of high-stakes tests, but it applies to other measures as well. In fact, some of the indicators proposed earlier might be even more subject to manipulation than test scores. One of the advantages of a system that uses multiple measures of school performance (as outlined here) is that it is more resistant to corruption than a system based on a single or a small set of measures.<sup>12</sup> Still, it is important to devise strategies for monitoring the validity of indicators over time—and in cases where corruption is evident or likely, to develop audit mechanisms to detect it.

A crucial advantage of assessing charter schools with more comprehensive criteria is that evaluations can also be customized to address the needs of different schools and groups active in the charter school movement. In the researcher's ideal world, parents, educators, and lawmakers assessing charter schools would review and assess all the information available about charter schools before reaching conclusions about their per-

*Performance measures are often corrupted, particularly when high stakes are attached to them.*

formance. However, the reality is that district administrators are likely to be interested in a somewhat different set of measures than, say, parents or state policymakers. In theory, it is possible to develop a comprehensive system of indicators that could meet the needs of all users. But in practice it is more likely that researchers will find themselves providing different sets of indicators to distinct interest groups to help policymakers make well-informed assessments and avoid information overload.

Similarly, even once a comprehensive set of indicators is established, policymakers are unlikely to expect the same results at all charter schools. In what instances should educators and parents accept differences in outcomes that stem from variations in curriculum, instruction, or other school characteristics? Charter schools with a thematic focus, such as business, health, or technology, may reasonably be expected to achieve different outcomes than charter schools of a more traditional scope. Charter schools are rich and varied, and the reading and mathematics test scores currently used to assess charters provide at best an incomplete picture of their effectiveness. A comprehensive set of indicators that allows for customizing analysis might be a way of addressing the fact that a core set of outcomes should be of interest to all schools, and an additional set of criteria will be of primary interest to a subset of charter schools.

The more comprehensive, high-quality data that analysts can bring to the charter school debate, the better. But researchers, parents, and educators need not feel handcuffed by imperfect data. Given the narrowness of most current charter school assessments, broadening the evaluation agenda may yet demonstrate that we still have a lot to learn about the full impact of charter schools.

*Broadening the evaluation agenda may yet demonstrate that we still have a lot to learn about the full impact of charter schools.*

## NOTES

1. Robert Rothman, Jean B. Slattery, Jennifer L. Vranek, and Lauren B. Resnick, *Benchmarking and Alignment of Standards and Testing*, CSE Technical Report 566 (Los Angeles: Center for Research on Evaluation, Standards, and Student Testing, 2002).
2. Robert L. Linn, "Assessments and Accountability," *Educational Researcher* 29, no. 2 (2000): 4–16; for a review of the research on score inflation, see Laura S. Hamilton, "Assessment as a Policy Tool," *Review of Research in Education* 27 (2003): 25–68.
3. Although we do not discuss research methodology in this chapter, it is worth pointing out that some of the approaches used to evaluate achievement outcomes cannot be used for one-time events such as high school graduation. High-quality analyses will require sophisticated longitudinal modeling approaches, e.g., survival analysis for graduation rates, or comparisons of trajectories for earnings.

4. We make a distinction between productivity and attainment to create categories and simplify the presentation, but obviously attainment affects productivity, so there is overlap between these two categories, and many of the indicators fit under both.
5. See [http://nces.ed.gov/programs/digest/d05/tables/dt05\\_181.asp](http://nces.ed.gov/programs/digest/d05/tables/dt05_181.asp).
6. Ferran Mane, "Trends in the Payoff to Academic and Occupation-Specific Skills: The Short and Medium Run Returns to Academic and Vocational High School Courses for Non-College Bound Students," *Economics of Education Review* 18, no.4 (1999): 417-437.
7. Patrick Wolf, "School Choice and Civic Values," in *Getting Choice Right: Ensuring Equity and Efficiency in Education Policy*, ed. Julian R. Betts and Tom Loveless, 210-244 (Washington, DC: Brookings Institution Press, 2005).
8. Other structural features such as school size are likely to be of interest to some stakeholders, but we focus here on a small number of structural features that are most likely to be related to outcomes.
9. Jaana Juvonen, Vi-Nhuan Le, Tessa Kaganoff, Catherine H. Augustine, and Louay Constant, *Focus on the Wonder Years: Challenges Facing the American Middle School* (Santa Monica, CA: RAND, 2004).
10. The experimental evidence is strong. See Jeremy D. Finn and Charles M. Achilles, "Answers and Questions About Class Size: A Statewide Experiment," *American Educational Research Journal* 27, no. 3 (1990): 557-577. The real world evidence is limited because few places have done class-size reduction well. For example, in California there was a marked decline in the average preparation of teachers after implementation of a class-size reduction program, which may have reduced potential benefits.
11. Juvonen et al., *Focus on the Wonder Years*.
12. Daniel Koretz, "Using Multiple Measures to Address Perverse Incentives and Score Inflation," *Educational Measurement: Issues and Practice* 22, no. 2 (2003): 18-26.