

POLICY Research Brief

EXAMINING KEY EDUCATION ISSUES

A Membership Benefit of NSBA National Affiliates

The Value of Value-Added Analysis¹

By Darrel Drury and Harold Doran

As states and districts explore new ways to assess the quality of schools and support their continuous improvement, value-added analysis of standardized test-score data has emerged as an attractive complement to traditional means of measuring school effectiveness. By factoring out non-school-related influences on student achievement gains, the value-added approach lets state policy makers and local education authorities gauge the impact of education policies on student learning.

Modern accountability systems that incorporate value-added analysis provide an objective means for assessing the effectiveness of teachers, structuring individualized plans for professional development, evaluating new education programs and curricula, and driving instructional improvement in the classroom. This *Policy Research Brief* explores the emerging role of value-added analysis in shaping these accountability systems.

Historical Perspective

Before the standards-based reform movement that swept across the American education landscape in the past few decades, the quality of schools was measured almost exclusively in terms of inputs, such as the number of books in school libraries or the qualifications of teachers. Subsequently, demands for greater accountability in public education led states and districts to supplement these indicators with ones defined in terms of student performance, especially as measured by standardized test scores.

Today, in most states and districts, test-score data are reported as simple snapshots of student performance, commonly referred to as current-status indicators.

Such snapshots represent the average score for students enrolled in a district, school, grade level, or classroom, assessed using percentile ranks, the proportion of students meeting a state- or district-designated performance standard, or other means.

Although useful in describing performance for a given student population in a particular year, these indicators may actually provide less information about school quality than the traditional input measures they have largely replaced. Indeed, in

- Because of widespread student mobility in and out of different schools and districts, such indicators tend to be highly “contaminated” by the influence of multiple educational settings.

Value-Added Analysis

Concerns such as these have led researchers, policy makers, and practitioners to focus increasingly on “value-added” analysis, an approach to analyzing and reporting test-score data that addresses many of these pitfalls.

“

Value-added performance data can play an important role in aligning policies, resources, and instructional strategies.

”

the absence of other measures, current-status indicators are invalid and potentially misleading, for several reasons:²

- Current-status indicators reflect the combined influence of family, background, community, and years of prior schooling on student achievement and unfairly stack the deck against schools that serve disproportionately high numbers of disadvantaged students.

- Because they reflect the cumulative impact of school- and non-school-related influences on achievement over multiple years, current-status indicators can be skewed in any given year by prior influences on student learning.

- Current-status indicators fail to localize performance to a specific classroom or grade level, making them less useful as a means of evaluating reform initiatives, teacher effectiveness, and program innovations.

Whereas current-status measures report the performance of a group of students at a single point in time, value-added analysis focuses on the achievement gains of individual students over time (for example, from spring to spring). But unlike simple gain-score analysis—in which earlier test scores are subtracted from later scores—value-added analysis uses statistical methods that, in effect, separate out (or at least significantly adjust for) the influence of non-school-related variables, such as students’ socioeconomic background, on academic growth.³

Value-added test-score data, combined with other measures of performance, permit policy makers to hold teachers and administrators accountable for the value they add to students’ educational experience without penalizing (or rewarding) them for pre-existing differences in their

POLICY Research EXAMINING KEY EDUCATION ISSUES Brief

The *Policy Research Brief* series provides research and commentary on key education issues, informing decision making at the district level and supporting the broader public interest. The series is published biannually by the Advocacy and Issues Management Section of the National School Boards Association.
Copyright 2003, NSBA.

President **Mossi W. White**

Executive Director **Anne L. Bryant**

Deputy Executive Director **Joseph P. Villani**

Associate Executive Director,
Advocacy and Issues Management
Michael A. Resnick

Director of Publications **Sally Banks Zakariya**

Production Manager **Michele Sabatier**

About NSBA

The National School Boards Association is the nationwide advocacy organization for public school governance. NSBA's mission is to foster excellence and equity in public elementary and secondary education in the United States through local school board leadership. Founded in 1940, NSBA is a not-for-profit federation of state associations of school boards across the United States and the school boards of the District of Columbia, Guam, Hawaii, and the U.S. Virgin Islands.

About the National Affiliate Program

The National Affiliate Program extends NSBA's services directly to local school districts. School districts are eligible to join provided they are members in good standing of their state school boards associations.

About the Advocacy and Issues Management Section

The Advocacy and Issues Management Section implements NSBA's Action/Advocacy Agenda and carries out NSBA's lobbying efforts at the national level. By lobbying the Congress, the White House, and federal agencies, the section helps increase federal funding for local school districts and reduces costly federal mandates; helps improve federal education programs by making legislative and regulatory changes local board members support; protects the governance role of school boards from congressional attack; and promotes the role of school boards as a key democratic institution in our country's education system.

National School Boards Association

1680 Duke Street
Alexandria, VA 22314
Telephone: (703) 838-6722
Fax: (703) 548-5613
E-mail: research@nsba.org

students' background and ability. Value-added performance data can play an important role in more effectively aligning district-level policies, resources, and instructional strategies with the needs of individual schools.

To implement a value-added accountability system, three key data requirements must be met. First, students must be tested at least annually.⁴ Second, test scores must be reported on a common scale so students' test scores can be compared from one year to the next.⁵ Finally, to track student achievement gains over time, students must be assigned individual identification codes (IDs) that remain consistent over all school years, regardless of the school attended.

The first of these requirements will soon be met in every state. The No Child Left Behind (NCLB) Act of 2002 requires states to test at least 95 percent of all regular education students in language arts/reading and math each year in grades three through eight. However, the new law does not compel states to adopt tests for which scores are reported on a common scale; nor does it require the assignment of unique student IDs. Presently, only 17 states give some type of language arts and math test in each of the grades from third through eighth—just nine administer standards-based tests in both subjects—and no more than 16 states have the ability to link student records from one year to the next.⁶

Concerns about Value-Added Analysis

Critics of value-added analysis argue that, to be effective, an accountability system must be easily understood, pointing to the fact that the advanced statistical methods underlying value-added analysis are well beyond most educators' grasp. It is worth noting, however, that few people understand the statistical manipulations underlying the Consumer Price Index and other key economic indicators, yet most make effective use of such information in their day-to-day lives. Trading rigor and accuracy for simplicity is an indefensible strategy—the stakes are simply too high.

Others claim that value-added analysis inflates the degree of random error involved in measuring student performance, thereby increasing the risk that teachers or schools may be rewarded or punished for over- or under-estimates of their students' learning.⁷ But confidence in value-added assessment results can be increased by a variety of methods, includ-

ing averaging data over several school years or incorporating multiple achievement measures in the analysis.⁸

While value-added estimates of student performance are not perfect, it is clear that they provide more accurate estimates of student performance than current-status measures provide. Moreover, it is widely acknowledged that even an evaluation system involving substantial random error can motivate improvement. What cannot be tolerated is one that is inherently invalid and creates false or misdirected incentives.⁹

Finally, because value-added indicators adjust for student background characteristics associated with academic growth, some critics argue that they create lower performance expectations for disadvantaged students and the schools that serve them. This concern is misguided for two reasons:

- First, the value-added approach in no way precludes a state or district from setting high achievement standards for all its students—in fact, this is a requirement under NCLB.

- Second, schools serving more disadvantaged students will typically have *higher*, not lower, value-added performance goals. Because these schools generally start at a lower level of performance, they must cover more ground than other schools in order to achieve the same end. As Rob Meyer, a leading expert on value-added analysis, accurately observes, "This is a strength, not a weakness of the value-added approach."¹⁰

Driving Continuous Improvement

Accountability systems often emphasize their external function (reporting performance data to the state and community) at the expense of their internal function (providing information to drive continuous improvement in schools). Even when data *are* used for internal purposes, they often are poorly suited for the task. Value-added performance indicators give board members, central office administrators, principals, and teachers the opportunity to base critical decisions on measures that have greater precision and fewer flaws than average test scores and other traditional indicators of performance.

In combination with other means of gauging progress, value-added measures provide a basis for guiding decision making in several key areas:

- **Policy and program evaluation.** In contrast with current-status indicators, which potentially provide incorrect infor-

mation about the effectiveness of education policies and reforms, value-added performance measures provide a more accurate picture of what works and what does not.

This is clearly demonstrated in a convincing simulation study that examined average achievement (a current-status measure) and school performance (based on a value-added approach) over time, before and after the introduction of hypothetical reforms.¹¹ The study's conclusion is striking. Average test scores *decline* for five years after the introduction of successful reforms and *increase* for a decade after the introduction of reforms that have no real impact on performance.

Although counterintuitive, these results can be attributed to the fact that average achievement reflects information that is grossly out of date—it is, in effect, the product of achievement gains for the current year and all previous years of schooling combined. Thus, a spike or decline in average math achievement at the 10th-grade level may reflect programmatic or curricular changes implemented several years before in an earlier grade. Faced with such circumstances, a district relying on average test scores, rather than value-added indicators, might eliminate reforms that were beneficial or ramp up reforms that were of no value.

• **Instructional change.** Although teachers report that, to some extent, they fashion their instructional plans on the basis of students' test performance, at least one study suggests that, overall, teachers have serious doubts about the relevance of standardized testing to educational improvement.¹² Teachers in this study questioned the role of testing in improving schools, providing useful feedback about how well they are teaching, and clarifying learning goals. In light of the inability of most state and district testing programs to separate out achievement results associated with non-school-related factors, such

concerns are not surprising. Without the benefit of value-added analysis, teachers cannot discern, for example, whether achievement gains are the result of instructional practices or out-of-school influences.

• **Teacher effectiveness.** The potential of the value-added approach to affect student achievement is perhaps greatest when applied to the evaluation of teacher effectiveness. A number of recent studies using value-added analysis have found teacher effectiveness to be the single greatest predictor of school-related student learning. Taking note of this, some

to engage in a program of professional development aimed at strengthening their skills in areas of weakness.¹⁵

Value-Added Analysis and NCLB

With the passage of No Child Left Behind, the external accountability role of value-added analysis will be determined, in large part, by the degree to which states embrace it as a means of defining and measuring Adequate Yearly Progress (AYP). Currently, most states define AYP solely in terms of performance categories, and some would suggest that this practice must be continued under the new law.



Trading rigor and accuracy for simplicity is an indefensible strategy—the stakes are simply too high.



observers suggest that value-added analysis could be instrumental in reducing the achievement gap between advantaged and disadvantaged students. They argue that the most effective teachers, identified through value-added analysis of test-score data, should be given incentives to teach low-achieving students, thus boosting their achievement in relation to that of their more advantaged peers.¹³ Others advocate using value-added indices of teacher effectiveness, along with supplementary measures of teacher quality, as the basis for performance-based compensation systems.¹⁴

• **Individualized professional development.** Professional development needs to move away from the one-size-fits-all approach, in which all teachers—regardless of experience, competence, or success rate—receive the same type of training. Teachers have individual strengths and weaknesses, and staff development should reflect this reality. Value-added analysis can potentially play an important role in this regard. For example, the Dallas Independent School District (DISD) uses value-added assessment data to structure in-service training for struggling teachers. The DISD accountability system produces class “profiles” that identify skill areas in which students need particular help. Using these profiles, along with teacher portfolios and structured interviews, administrators develop individualized blueprints for staff development at the beginning of each school year. Over the course of the year, teachers are expected

However, the U.S. Department of Education has recently issued clarifications that seem to signal approval of measurement models recognizing both growth and proficiency.¹⁶ Some states, such as Colorado, have begun to experiment with value-added accountability, and researchers are aggressively exploring value-added approaches that measure both AYP and the degree of progress made by schools. One such model, developed by Yeow Meng Thum of UCLA, aims to meet NCLB statutory provisions by reporting test-score data in terms of performance levels while using value-added analysis to capture the rate of improvement required of a school to achieve 100 percent proficiency by the targeted year, 2014.¹⁷

It makes little sense to continue to define AYP solely in terms of the percentage of students crossing an arbitrary bar of “proficiency,” while ignoring the growth that occurs within broad performance categories. This is tantamount to measuring a child's height with a yardstick but acknowledging growth only when his or her height exceeds 36 inches. Relying exclusively on average test scores and proficiency cut-offs to measure AYP can lead to unintended consequences, such as leading teachers and administrators to target instruction to those students most likely to cross the proficiency cut point, while leaving others without focused instruction.

The present approach to measuring AYP also tends to place schools that serve large numbers of disadvantaged students at an unfair disadvantage, since they may

THE SCHOOL BOARD ROLE

Value-added analysis can guide school board decision making in the following areas:

- Policies and programs
- School reforms
- Curricula
- Teacher compensation
- Professional development
- School intervention

make great strides and still be labeled deficient. The most defensible position, it seems, is to develop accountability systems that report not only the extent to which schools are successful in achieving state and district standards, but also their rates of improvement.

NCLB's annual testing requirement, combined with new flexibility in defining and measuring AYP, has laid the groundwork for potentially revolutionary changes in how states and districts measure school performance. Certainly, compliance with the law can be achieved without restruc-

turing state accountability systems to include value-added analysis. But to follow the path of least resistance, rather than exploit this opportunity to provide schools with meaningful performance measures that can drive educational improvement, seems ill advised.

Looking to the Future

Although all value-added methods share the same underlying logic, there are some differences yet to be reconciled. For example, researchers differ in the way they handle missing data, their inclusion or exclusion of demographic data, their approach to dealing with test error, and even the analytical units (such as schools, grade levels, or classrooms) on which they base analyses. New American Schools (NAS) has recently assembled a team of leading experts in value-added analysis to begin to grapple with these issues.¹⁸ As a result of this work, NAS and its partners will be better able to assist states and districts in selecting a value-added approach that best meets local requirements.

As new systems of value-added accountability come on line, it is also important that districts manage information effectively and make it accessible to stakeholders at every level. Modern Web-based systems for warehousing data are, therefore, an essential element in a successful system of value-added accountability.

Finally, whatever accountability model a state or district ultimately chooses, it will be unlikely to motivate greater student learning if board members, central office administrators, principals, and teachers do not understand how to use this new information effectively. Therefore, training in data-driven decision making for stakeholders at every level must accompany the implementation of new value-added systems of accountability.

About the Authors

Darrel Drury, vice president for research and evaluation at New American Schools (Alexandria, Va.), oversees the delivery of services to states and districts in the areas of needs assessment, information management, strategic planning, accountability, evaluation, and training for data-driven decision making.

Harold Doran, director of research and evaluation and a former school principal, is the coordinator of New American Schools' comparative study of the leading methods used to conduct value-added analysis.

Please address comments and inquiries to ddrury@nasdc.org.

Endnotes

1 The authors wish to express their appreciation to Mike Resnick, associate executive director of NSBA, for his insightful comments on two earlier drafts of this Policy Research Brief. We also wish to thank Libia Gil, chief academic officer of New American Schools, and those who served as external reviewers—Steven Ross, Bill Sanders, and Yeow Meng Thum—for their suggestions for improving the initial draft of this Brief.

2 Meyer, R. (1997). Value-added indicators of school performance: A primer. *Economics of Education Review*, 16(3).

3 Simple gain score analysis produces unadjusted scores that are contaminated by the many non-school-related factors that have been shown to affect student learning. This approach is also confounded by a statistical artifact known as regression to the mean, where student cohorts with extremely high or low average initial scores tend to be unstable and move toward the middle of the score distribution in subsequent administrations of a test. In contrast, value-added analysis controls for factors outside of schools' control and accounts for the appropriate amount of shrinkage toward the middle.

4 While not technically a prerequisite for value-added analysis, annual testing is required for performance to be localized to a specific classroom or grade level. It is also recommended that tests be aligned with state and district content standards.

5 Scale scores must be vertically equated to enable comparisons over time.

6 Olson, L. (2002, January). Testing systems in most states not ESEA-ready. *Education Week* 21(16).

7 In Ballou, D. (2002, Summer). Sizing up test scores. *Education Next*, Dale Ballou argues that a student's performance on any given test will be due partly to true ability and partly to random error. When two tests are compared, as in value-added analysis, the problem of random error is compounded. The measurement errors associated with the two tests—taken several months apart—are unlikely to be related and, therefore, when the difference between earlier and later scores is computed, the measurement errors do not "cancel out." At the same time, however, true ability is relatively stable and, therefore, when the difference between scores is taken, a substantial proportion of the scores that represents true ability will cancel out. Thus, the proportion of a gain score that represents measurement error is magnified vis-a-vis the initial scores.

8 Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin*, 92(3), 726-748.

9 Greene, J. P. (2002, Summer). The business model. *Education Next*.

10 Meyer, R. (1997).

11 Meyer, R. (1997).

12 Herman, J. L., & Golan, S. (1993, Winter). The effects of standardized testing on teaching and schools. *Educational Measurement: Issues and Practice*.

13 Haycock, K. (1998, Summer). Good teaching matters: How well-qualified teachers can close the gap. *Thinking K-16*.

14 Critics charge that the statistical error associated with value-added analysis would result in rewarding some ineffective teachers and withholding rewards from other relatively effective teachers. However, if multiple factors are used for assessment, results are averaged over extended time periods, and comparison groups are carefully selected (e.g., rewarding the top 10 percent of all 5th grade teachers in the district), such concerns should not outweigh the potential advantage of more closely aligning incentives with achievement (see Summers, A. (2002, Summer). Expert measures. *Education Next*).

15 Sullivan, C., & Drury, D. (2002). *Building capacity: Professional development for new and experienced teachers*. Alexandria, Va.: National School Boards Association.

16 According to one recent statement from the U.S. Department of Education concerning definitions of AYP, "states and LEAs are strongly encouraged to develop systems to recognize...low-performing schools that are making...improvement." See U.S. Department of Education. (2002). Dear colleague letter. Online at www.ed.gov/News/Letters/020724.html.

17 Thum, Y. M. (2002). No child left behind: Methodological challenges and recommendations for measuring adequate yearly progress [Unpublished manuscript].

18 The New American Schools' team includes Damien Betebrenner (University of Colorado), Tony Bryk (University of Chicago), Jerome D'Agostino (University of Arizona), Robert Linn (University of Colorado), Robert Meyer (University of Chicago), Steve Raudenbush (University of Michigan), Steven Ross (University of Memphis), William Sanders (SAS, Inc.), and Yeow Meng Thum (UCLA).