

November 2003

November 2003 | Volume **61** | Number **3**  
**The Challenges of Accountability** Pages 55-59

### Adding Value to Accountability

How much is a school contributing to student learning? Value-added analysis offers precise, accurate measurements of student growth and a school's impact.

Harold C. Doran

Can the data gathered for accountability purposes provide helpful information to both schools and the public? Questions about how to analyze and use data effectively have become urgent as states and districts throughout the United States have developed high-stakes accountability plans. The driving force is No Child Left Behind.

Accountability, a dominant theme of this federal legislation, requires all states to develop accountability plans that measure the effectiveness of each public school, primarily through student achievement test score data. The accountability plans must also include other indicators of achievement, but high achievement on the other indicators cannot make up for poor performance as measured by test scores.

The cornerstone of these accountability provisions is *adequate yearly progress*, a term that is familiar to those working in Title I-eligible schools. Under the new federal law, all schools within a state, not just those receiving federal financial assistance, demonstrate adequate yearly progress when the percentage of students scoring at or above *proficient* on achievement tests increases by a certain amount each year. The legislation specifies a methodology for setting targets for all students. For those schools that receive federal assistance, which most U.S. schools do, failure to meet these targets will trigger a series of sanctions that could lead to school reconstitution.

The high-stakes nature of accountability necessitates accurate and valid methods for measuring adequate yearly progress, but the challenge of developing such methods currently confounds most state and district accountability plans. Accountability data should make it possible for policymakers and educators to determine the quality of schooling and to take actions to improve it, but so far, many of the methods used to analyze test score data are both unreliable and uninformative. They have remained popular because they are relatively easy to compute and explain. Simplicity at the expense of teachers and students, however, should not be a justification for any measurement method.

Instead, accountability plans desperately need sound analytic methods that more fairly depict the impact of a school on student learning and provide results that can support schoolwide improvement and planning. One such method, value-added analysis, is a potent technique that overcomes many of the problems associated with typical school evaluations.

### Status Versus Growth

Before the reauthorization of the Elementary and Secondary Education Act (ESEA) in 1989, *inputs*—such as per-pupil funding, student-teacher ratios, and teacher qualifications—were the exclusive measure of school quality. Since then, standards-based reform initiatives have led to a greater demand for *outcomes*,

specifically student achievement test scores, to serve as the primary indicator of school quality.

Unfortunately, these outcomes often provide only snapshots of school performance, such as the average percentile rank or percentage of students who score at or above proficient on one test. Typically, administrators and policymakers use these results to rank-order schools and to label those below a specified cut point as “underperforming” or “failing.”

Accountability in this form focuses on the current *status* of a school—its performance at a single point in time. The need to maintain or achieve a certain status creates perverse incentives for schools to manipulate test scores rather than foster real growth for their students. It encourages schools to target instruction for those students nearest the *proficiency* cut point and to ignore the instructional needs of high-achieving and low-achieving students. Schools that focus on the middle-performing students are likely to show an increase in the percentage of students at *proficient* but at the expense of providing meaningful instruction for students at high and low levels of achievement.

Many accountability plans rely on simple statistics, such as group averages, which are descriptive but not informative. Educators cannot use aggregated data to support instructional improvement at the student level. Doing so would be tantamount to a physician learning that the average blood pressure of her patients is 120 over 80; she cannot use this information to diagnose and treat an individual patient.

Status-based accountability systems that rely on such simple approaches provide misleading and invalid results for the following reasons:

- The test scores are subject to external variables beyond the control of schools. Students do not get assigned to schools randomly. If randomization were possible, we could mitigate the influence of external variables, such as economic status, on student performance, but the nonrandom assignment of students to schools and classrooms guarantees an element of bias among the test scores. Because randomization is not possible, we must take into account the influence of external variables when evaluating the quality of schools.
- Failure to recognize growth toward proficiency unfairly punishes schools serving disadvantaged populations. A school may be making tremendous progress but may still be some distance away from having all students at or above proficiency.
- A test score for an 8th grade student is invalid for evaluating 8th grade instruction because it reflects the cumulative impact of schooling over all previous school years. In that sense, current-status measures are cumulative (Meyer, 1997) and cannot evaluate instruction.
- Cut-score categories, such as *basic* and *proficient*, are gross measures of academic performance. Using these categories to measure growth is the equivalent of measuring a child’s height with a yardstick but only acknowledging growth when he or she has exceeded 36 inches. These gross measurements fail to recognize the significant academic progress that an individual may make within a cut-score category.

## **How to Determine the School’s Impact**

In recent years, value-added analysis has emerged as a method for depicting the impact of a school on student achievement. When incorporated appropriately with modern high-stakes accountability plans, value-added analysis can supply the nourishment needed to support student learning.

Value-added analysis is so named because it seeks to answer one fundamental question: How much *value* has a school *added* to a student’s learning? Value-added analysis makes two important assumptions. First, it assumes that one can measure an individual’s growth in learning from one measured occasion to the next.

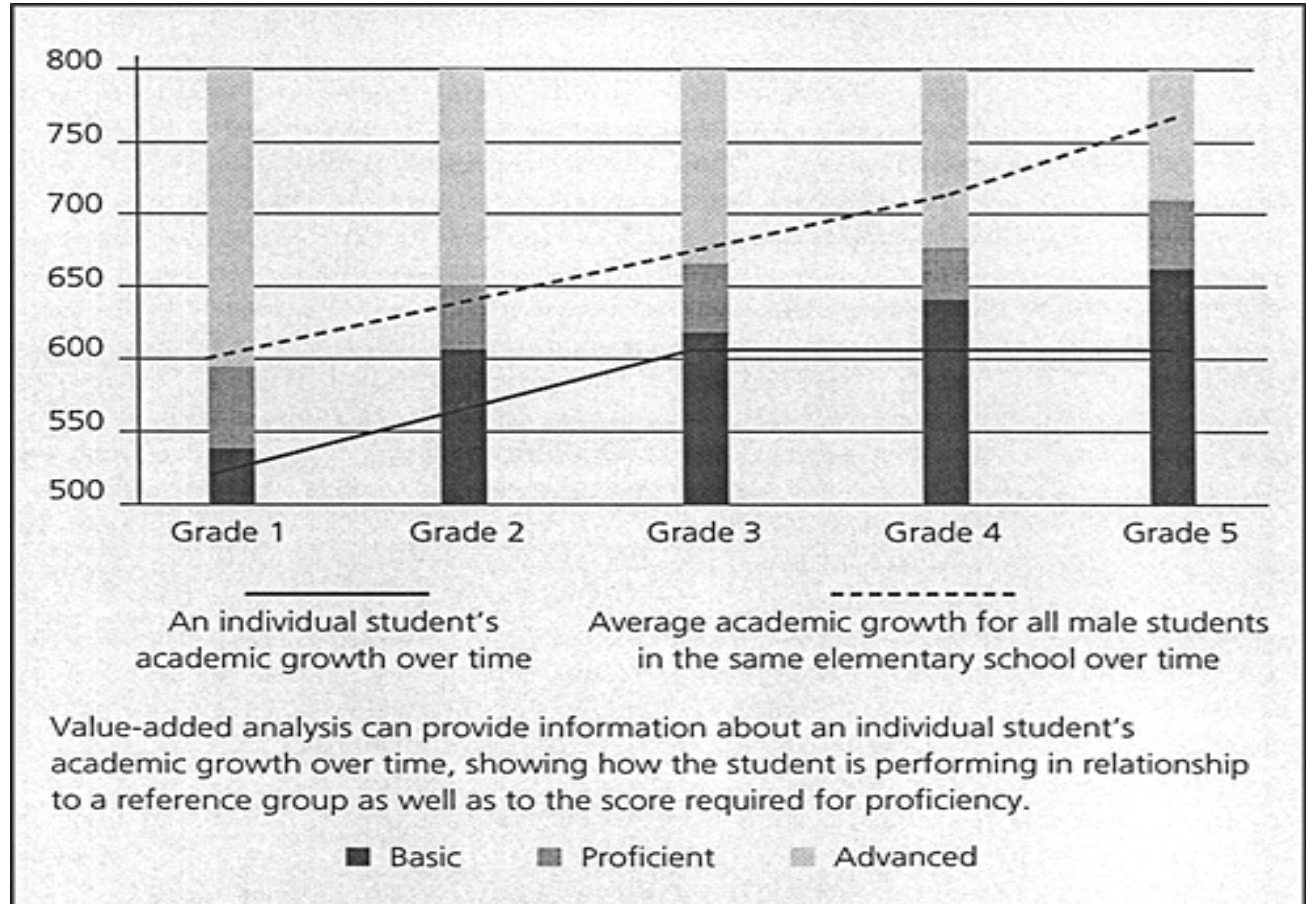
Therefore, it tracks the progress of individual students over time. These longitudinal measures differ from conventional measures because they promote growth for all students rather than focusing on status.

Value-added analysis also assumes that schools have only partly contributed to changes in test scores. Learning is the sum of many factors, including both school and nonschool elements. School evaluations need statistical methods to separate the impact of the school from the impact of nonschool characteristics, such as students' socioeconomic background. For this reason, value-added analysis relies on advanced statistical methods to estimate the effects of the school.<sup>1</sup>

Because conventional status measures fail to consider the impact of non-school-related factors on student learning, they actually measure the impact of both school and nonschool factors. Therefore, they cannot provide valid measures of *school* performance. By contrast, value-added analysis statistically estimates the contribution of the school as separate from the non-school-related variables, such as economic status, that contaminate traditional analyses. As Thum and Bryk argue, "Anything other than a value-added-based approach is simply not defensible" (1997, p. 102).

Figure 1 illustrates the richness of information that value-added analysis offers. The solid line tracks the academic growth of an individual male student; the dotted line represents the average performance of all boys in the same elementary school. The colored bars correspond to the scale score ranges for the performance categories on the test, including basic (red), proficient (green), and advanced (yellow). For a proficiency score in 5th grade, for example, a student must score between 660 and 710, the green area of the bar. The figure shows that the individual student was scoring lower than the group but was growing at the same rate as the others until grade 3, at which point his performance flattened.

**Figure 1. What Value-Added Analysis Can do**



This figure illustrates how this student is performing not only in relationship to a reference group, but also in relationship to the score required to become proficient, thereby providing a more precise, multidimensional perspective. Value-added analysis can assess both the percentage of students at the proficient level and the progress schools are making toward the proficiency standard (Thum, in press).

In conventional analyses, a class-room teacher would receive a one-dimensional report describing this student as falling below proficiency, but it would not report information relating to an important dimension—growth. Only when educators examine longitudinal trends of individual students, relating them to reference groups and outcome targets, can they tailor instructional strategies to meet the needs of individual students.

Value-added analysis, combined with other valid indicators, can more reliably assess school quality without punishing or rewarding schools for preexisting differences related to student background characteristics or other non-school-related factors. Value-added data coupled with data acquired from a qualitative school review enable richer perspectives and more accurate judgments about the actual performance of systems and schools. Together, they provide powerful contextual information that practitioners and community members can use to identify needs and take action to improve teaching and student learning.

## **Good Data for Dialogue and Action**

Historically, we have used accountability data to impart judgments about a school's effectiveness to the public. These data rarely engage both the school and the community in a thoughtful, diagnostic, and formative process designed to improve the quality of schooling.

We should not mistake the historic function of accountability, however, for its future potential for school and community engagement. Accountability data can, and should, link internal members of schools and systems (school board members, administrators, teachers, and principals) with external stakeholders—the state and the community.

Within schools, appropriately analyzed data can serve the internal, professional purposes of accountability systems (O'Day, 2002), challenging educators to reflect on their teaching practices and to consider whether students have full opportunities to learn. Accurate data can help school administrators modify school policies and practices and reallocate resources to fully support areas in need. Appropriately analyzed data can also meet the external, public purposes of accountability, inspiring public actions to support the improvement of education for children and fostering community engagement, community leadership, and community resource allocation.

Combining high-quality data with a systematic process for change breathes life into accountability systems. Engaging all stakeholders in reflective dialogue around multiple sources of high-quality data can lead to better decisions in the community and classroom. To ensure that accountability meets their needs, stakeholders—in small groups, separately and together—should first initiate discussions about what they want to learn from the data. They should then disaggregate the data so that everyone can see it from a number of perspectives and generate hypotheses about the data results. They must next create action plans for change on the basis of the most credible hypotheses and reevaluate the effectiveness of the new strategies after reviewing the new data. Such a framework for interpreting data can lead to better decisions for improving instruction and student performance.

## Components of Value-Added Analysis

To make value-added analysis possible, the following components must be in place.

*An annual testing system.* The testing requirements of No Child Left Behind indicate that an annual testing system will soon be in place. The value-added analysis can also use the specific measurement systems that a school uses, such as a computer adaptive test or tailored test.

*Student data in electronic format.* Although many school systems receive individual student data on paper, few have access to these data in usable electronic format. Electronic data usually arrive as ASCII text or tab-delimited files and require a process called Extraction, Transfer, and Loading to become organized into a suitable format. Schools using computer adaptive tests already have access to electronic data.

*Student and teacher identification numbers.* Student names are often misspelled or changed, and student mobility in many schools is high. In such situations, merging individual yearly test score files to create a longitudinal database becomes difficult. State education agencies should assign and maintain unique student identification numbers that remain consistent over all years, regardless of which school the student attends in the state. So far, 16 states have developed unique student identification numbers to link student records over time (Olson, 2002).

To track teacher performance over time, unique teacher identification numbers should remain constant from year to year. Value-added results can help identify an individual teacher's strengths and weaknesses and provide strong support for differentiated professional development.

*Sufficient sample sizes.* School districts generally have sample sizes large enough to produce credible estimates. Individual charter or contract schools, on the other hand, should consider developing a consortium with other schools to enhance the sample size of the data set.

*A consistent test score metric.* Being able to measure students over an extended period of time and make comparisons requires a test metric that can report development. A process known as *vertical equating* connects all forms and levels of the test and places them on a single, continuous scale. Test scores developed in this fashion use the same ruler to measure student progress over time, a necessary component of value-added analysis.

## Criticisms of Value-Added Analysis

Critics have argued that accountability systems can be effective only if everyone can understand them easily (Ballou, 2002). They claim that the statistical techniques of value-added methodologies are too complex and therefore too difficult to understand. Should a physician, then, use less scientific surgical procedures simply because they are easier for the patient to understand? Rather, the physician should explain the procedures as clearly as possible to the patient and fully disclose them to other knowledgeable physicians for their review. The statistical models of value-added analysis are complex, but analysts can make the results clear and accessible to all stakeholders and explain its processes to professional colleagues for their review.

Other critics have argued that value-added analysis sets lower expectations for different groups of students. This assertion is untrue for two reasons. First, value-added analysis is a technique designed to measure how much students have learned as measured by test scores. In no way does it preclude schools or parents from setting high expectations for academic learning. Second, schools serving disadvantaged students will likely have higher, not lower, goals for improvement. Because these schools typically start at a lower level of performance, they must cover more ground than other schools to achieve the same end goal: 100 percent of

students at or above proficient by 2013–2014.

Finally, some have argued that No Child Left Behind permits only one form of measurement, the increase in the percentage of students who score at or above proficient. A surface reading of the legislation may suggest that only one kind of measurement is acceptable, but a memorandum from the U.S. Secretary of Education (2002) encouraged the development of accountability systems that recognize improvement over time. Further, the No Child Left Behind legislation repeatedly calls for scientifically based research. Reverting to nonscientific methods to define and measure adequate yearly progress would be antithetical to the goals of legislation that strives for increased scientific rigor.

## Scientific and Practical

Value-added analysis is a method likely to provide meaningful and relevant data to support appropriate classroom and public action. Its inclusion in high-stakes accountability plans makes sense, both scientifically and practically. Combined with conventional status measures, value-added techniques can provide a multidimensional view of each school. In this sense, the diagnostic spirit of the No Child Left Behind legislation combines with the law's goal of measuring school results.

Combined with a systematic process for interpreting and applying these data, value-added analysis can help all stakeholders provide high-quality education to students.

## Endnote

<sup>1</sup> Traditional statistical techniques, such as ordinary least squares regression or repeated-measures analysis of variance, are unlikely to provide unbiased estimates of school performance (Osgood, 2001). Instead, analyses conducted on the basis of the general linear mixed model (commonly referred to as hierarchical linear models, mixed statistical models, and multilevel models) are better suited to analyze test score data. See Doran (2003), Millman (1997), or Raudenbush and Bryk (2002) for a review of statistical methods for assessing accountability in education.

## References

- Ballou, D. (2002). Sizing up test scores. *Education Next*, 2, 10–15.
- Doran, H. C. (2003, April). *Value-added analysis: A review of related issues*. Paper presented at the annual conference of the American Educational Research Association, Chicago, Illinois.
- Meyer, R. H. (1997). Value-added indicators of school performance: A primer. *Economics of Education Review*, 16, 283–301.
- Millman, J. (1997). *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* Thousand Oaks, CA: Corwin Press.
- O'Day, J. A. (2002). Complexity, accountability, and school improvement. *Harvard Educational Review*, 72, 293–321.
- Olson, L. (2002). Testing systems in most states not ESEA-ready. *Education Week*, 21(16).
- Osgood, D. W. (2001). Advances in the application of multilevel models to the analysis of change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for analysis of change*. Washington, DC: American Psychological Association.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models, applications, and data analysis methods*. Thousand Oaks, CA: Sage.

Thum, Y. M., & Bryk, A. S. (1997). Value-added productivity indicators: The Dallas system. In J. Millman (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* (pp. 100–109). Thousand Oaks, CA: Corwin Press.

Thum, Y. M. (in press). Measuring progress towards a goal: Estimating teacher productivity using a multivariate multilevel model for value-added analysis. *Sociological Methods & Research*.

U.S. Secretary of Education. (2002, July 24). *Dear colleague* [Online]. Washington, DC: U.S. Department of Education. Available: [www.ed.gov/policy/elsec/guid/secletters/020724.html](http://www.ed.gov/policy/elsec/guid/secletters/020724.html)

*Author's note:* Richard Wenning, formerly of New American Schools; David Baker, Associate Superintendent of Flowing Wells School District, Tucson, Arizona; and Bill Sanders, Manager of Value-Added Assessment and Research for SAS in School, Cary, North Carolina, contributed to this article.

**Harold C. Doran** ([hdoran@nasdc.org](mailto:hdoran@nasdc.org)) is Director of Research and Evaluation at New American Schools, where he leads the New American Schools Value-Added Project.

Copyright © 2003 by Association for Supervision and Curriculum Development

**[Association for Supervision and Curriculum Development \(ASCD\)](#)**

1703 N. Beauregard Street, Alexandria, VA 22311 USA • 1-800-933-2723 • 1-703-578-9600

[Copyright © ASCD, All Rights Reserved](#) • [Privacy Statement](#)