Evaluating Value-Added:
Findings and Recommendations from the NASBE Study Group on Value-Added Assessment
Evaluating Value-Added

Findings and Recommendations from the NASBE Study Group on Value-Added Assessments

October 2005
The NASBE Study Group on Value-Added Assessments

Members
(all are state board of education members unless otherwise indicated)

Richard Farmer, Jr., Chairman-DE
Isis Castro-VA
Nancy Danhof-MI
Randy DeHoff-CO
Matthew Diethelm-AZ
Edward Donley-PA
Josephine Figueras-NJ
Priscilla M. Haden-WV
Kandy Imes-NE
Michael Kean-CA, Associate Member, McGraw-Hill Education
Mary Sue Murray-GA
Mike Nettles-NJ, Ex officio member, Educational Testing Service
Sharon M. Osborne-RI, State Board Executive
Karabelle Pizzigati-MD
Hilma Prather-KY
Clyde Reese-NJ, Ex officio member, Educational Testing Service
Debra Roberts-UT
G. R. “Sam” Schloemer-OH
Conrad (Wes) Snyder-MT, Associate Member, University of Montana-Missoula
John A. Tate III-NC
Richard Thompson-NC, Associate Member, University of North Carolina
Janet Tockman-ME

Presentations to the Study Group

Value-Added Assessments and Why They Matter to State Policymakers
Thomas Fisher, Assessment and Accountability Consultant

ABC’s of Value-Added Models
Henry Braun, Educational Testing Service
Laura Goe, Educational Testing Service

Value-Added/Growth Models as They Relate to State Accountability Systems
Brian Gong, National Center for the Improvement of Educational Assessment

State Actions in Value-Added Assessments
Lois Adams-Rogers, Council of Chief State School Officers
Rolf Blank, Council of Chief State School Officers

Value-Added Assessment in Tennessee
William Sanders, SAS Institute, Inc.

Lessons Learned from Districts Using Value-Added Models
Nancy Doorey, Brandywine (Delaware) Board of Education
Steve Iovino, Warwick school District, Pennsylvania

State Data Systems
Deborah Newby, Council of Chief State School Officers
Richard Kesner, Northeastern University

Value-Added and Formative Assessment: Tension, Synergy
Dylan William, Educational Testing Service
Siobhan Leahy, Educational Testing Service

Teacher Organization Viewpoints
Rob Weil, American Federation of Teachers
Tom Blanford, National Education Association

U.S. Department of Education Viewpoint on Value-Added Model and Its Possible Inclusion in AYP Calculations
Kerri Briggs, U.S. Department of Education

Policy Implications of Value-Added Assessments
Mitchell Chester, Ohio Department of Education

Study Group Staff
David Kysilko
Marsha M. Cullin
Weaver Rogers

NASBE gratefully acknowledges the Educational Testing Service (ETS) for its support of the Study Group meetings.

© Copyright 2005
National Association of State Boards of Education
All rights reserved
ISBN 1-58434-063-0

Additional copies of this report are available for $14.00 each + 10% for shipping and handling ($4.50 minimum) from the National Association of State Boards of Education at 277 South Washington Street, Suite 100, Alexandria, Virginia 22314; (800) 220-5183.
Table of Contents

Executive Summary ................................................. 4

Preface ........................................................................ 10

1 Overview of Value-Added Assessment ....................... 11

2 The Uses and Limitations of Value-Added Assessment ... 19
   Findings and Recommendations ................................. 27

3 Implementation Issues ............................................. 29
   Findings and Recommendations ................................. 34

Checklist for Policymakers Considering Value-Added
Assessment ................................................................. 37

Endnotes ........................................................................ 39
For decades, education policymakers have wanted to go beyond traditional ways of analyzing student test score data that only looks at the particular status or "proficiency level" that students have attained. It is very worthwhile, many have said, to also see how far students have progressed over a given year or number of years. This is particularly important when students have started far behind their peers. A student who goes from nonreading to reading has made great progress, even if the final test scores still shows that he or she is reading below grade level. It is also important, in this age of Adequate Yearly Progress (AYP), for schools that are educating large numbers of disadvantaged students to be able to show that their students have made significant strides, even if many have not yet achieved a level of "proficiency." Finally, educators have been intrigued with the possibility of accurately determining the specific impact that a school or teacher has had on students' learning.

To address these issues, many education policymakers, researchers, and practitioners have turned to value-added assessments (or more accurately, "value-added models for analyzing assessment data"). Value-added assessment models are statistical approaches that use multi-year student test score data, and student background characteristics in some models, to attribute student growth to schools, teachers, or both. In other words, value-added models attempt to determine how far a student has progressed compared to where the student started and to what degree that growth can be attributed to educational factors (as opposed to "external" factors such as socioeconomic status, race, parents' educational levels, or innate ability).

Because value-added models are now beginning to be used with increasing frequency in states and districts across the country—and because there are many questions surrounding value-added both in terms of just what the concept means and how the models should be used—NASBE established its Study Group on Value-Added Assessments to address and make recommendations on a number of issues critical to education policymakers, including: How does value-added fit with other aspects of the state's testing and accountability system? How accurate and valid are the results? What are the best uses for value-added data and analysis in terms of school improvement, accountability, or other uses? What doesn't value-added do well? What do states need to think about if they are planning to use value-added analysis as a component of their assessment and accountability system?

The Uses and Limitations of Value-Added Assessment

Through its investigation, the Study Group examined a number of possible uses for value-added models, including:

**Using value-added as a component of school accountability:** If, as research has borne out to some degree, value-added assessments are able to distinguish the effects of teachers and schools on student achievement, then using value-added data as a component of a state's accountability system would seem to make sense, for example, as one indicator in a school's report card. Researchers caution, however, that there are very significant policy and technical hurdles that must be overcome in order to successfully implement such a system.

**Using Value-Added for Teacher Accountability and Evaluation:** Since value-added analysis is often portrayed as a way to distinguish the effects of individual teachers and classrooms on student achievement, it is not surprising that many politicians and others are tempted to use them for teacher accountability. As the Study Group heard over and over, however, for a host of statistical and other reasons this area must be approached with considerable caution, especially in terms of high-stakes decisions such as ranking teachers, merit pay, and promotion or dismissal.
This doesn’t mean that value-added might not play some role in teacher evaluation, however. The Study Group’s conclusion is perhaps best summed up by the most extensive evaluation of value-added models to date, by RAND Education, which concluded that while “the research base is currently insufficient to support the use of value-added models for high-stakes decisions,” they do show “promise for lower-stakes diagnostic purposes” such as initially identifying possibly low- or high-performing teachers who can then be further evaluated to confirm results. In this way, value-added functions as a filter for detecting those teachers who would be subject to additional study through classroom observations, diagnostic tests, and portfolios. If further study confirms that a teacher is struggling, he or she can be counseled, provided professional development, or matched with a teacher who has been confirmed as highly effective.

**Using Value-Added as a Tool for School Improvement:** The Study Group was enthusiastic about the use of value-added models as a data-driven component of efforts to improve instruction at the classroom, school, and district levels. Indeed, many believe that this is the most significant advantage of value-added models. Some of the key areas that can be informed by value-added analysis include: policy and program evaluation; identification of students in need; schoolwide and team planning; individualized professional development; and resource management.

**Using value-added data for improving teacher training:** Given the uneven quality of graduates from the nation’s teacher training institutions, states have long struggled with finding mechanisms to hold preparation programs accountable for results, or at least to get a clearer picture of the effectiveness of new teachers and to help shore up deficiencies in their training. While this use of value-added analysis is still in its infancy, two states (Louisiana and Ohio) have initiated projects that use value-added data to link the effectiveness of new teachers with their preparation programs.

**Using value-added or growth measures as a component of AYP calculations:** Many educators and policymakers feel that strict, yearly adherence to proficiency targets as used under the No Child Left Behind Act unfairly punishes those schools dealing with large numbers of disadvantaged students, because they are required to catch up, often very quickly, to schools that start with higher achieving students. Many states are now looking into ways to add growth or value-added measures to AYP calculations, and the Study Group strongly urges the U.S. Department of Education to provide this flexibility while still maintaining the framework of proficiency for all students that is the basis of the law.

In considering these uses for value-added assessment models, the Study Group made the recommendation below:

### Findings and Recommendations on the Uses and Limitations of Value-Added Models

**Recommendation 1.** Value-added assessment is not designed for high-stakes use in teacher evaluations. The Study Group recommends that value-added information not be used for high-stakes teacher evaluation involving either rewards or punishments. We believe that educators should recognize that value-added assessment is a “tool,” but it is not the “total”—and indeed that the data can only with certainty identify about the top 10 percent and bottom 10 percent of teachers. Researchers also note that the number of studies supporting the value-added methodology is relatively limited and that there is a serious lack of validity studies. Thus, users are strongly encouraged to exercise caution in any use of value-added assessments so that the proper use of the technique does not go beyond the capabilities of the tool.

**Recommendation 2.** Value-added assessment has significant potential—when used in conjunction with other measures and supports—as a tool to improve teaching. The potential of value-added data to 1) differentiate the very most and least effective teachers and 2) show individual teachers more precisely in which areas and with which students they are and aren’t being successful can be a valuable tool in helping all teachers improve.
Recommendation 3. Value-added assessment information can be a powerful school improvement tool in promoting more effective practices and resource allocation. If a state or district uses value-added information, the Study Group recommends that the information be used for diagnostic purposes geared toward strengthening school performance and teaching techniques. Indeed, perhaps the most potent potential of the value-added model is through its power to start the dialogue about improving teacher effectiveness. At the same time, principals and school teams can use value-added data to identify effective or ineffective programs and curricula, identify individual students or groups of students who need additional support to reach their potential, and target school resources to areas in need.

Recommendation 4. Once-a-year value-added analysis does not substitute for other ongoing testing information that helps teachers adjust and target instruction to meet the needs of their students. While the Study Group believes that value-added assessment models have significant potential, the Group also recommends that schools avoid reliance on information from just a yearly testing program. Teacher-designed tests, commercial diagnostic measures, and other interim assessments done throughout the school year are crucial to keeping schools on target and can allow for mid-course corrections as needs arise. Waiting for results from one statewide testing program is inadequate for improving classroom instruction.

Recommendation 5. Value-added models show potential for tracking and improving the effectiveness of teacher training institutions. Using value-added data with newly licensed teachers has the potential to provide policymakers with information on which preparation programs are producing the most and least effective young teachers. Preparation programs can also use this information, especially if used alongside interviews with the teachers, to pinpoint weaknesses in their curriculum and other aspects of their program. However, as with other areas, the Study Group does not believe that value-added assessment is appropriate at this time as a stand-alone, high-stakes measure for evaluating teacher training institutions.

Recommendation 6. The Study Group urges the U.S. Department of Education to allow growth indicators as a component of AYP calculations within the No Child Left Behind framework. While the rigorous targets for attainment required by NCLB serve an important purpose in keeping a focus on the ultimate goal of high achievement for all students, adding growth is needed not only out of fairness, but because together these indicators provide the most accurate picture of the effectiveness of schools. In addition, failure to use growth as one indicator of success could end up making it even more difficult to retain effective teachers in disadvantaged schools.

Implementation Issues

Because value-added assessments are so new to the education enterprise, because they tend to be extremely complex in their mechanics, and because there are potentially a number of high-stakes implications for their use, state boards of education and other decisionmakers will find there are a host of issues that must be considered before, during, and after implementation of value-added systems. While some of these are technical issues that must be “checked off” as the state proceeds, the Study Group found that other technical, system-related, and political issues have the potential to be “deal breakers” or to render the value-added models far less effective than they otherwise might be.
Key Testing and Data Issues Affecting Implementation

Value-added models use results from state or district testing programs, along with lots of student and other data, for the basis of their analysis. Therefore, value-added analyses are believable only when the underlying testing program has technical quality, produces valid scores, and maintains an accurate longitudinal database. Holding to these standards is likely to be a challenge for many states. The Study Group identified the following as some of the key areas of concern:

- **Testing Issues:** Policymakers must first ensure that state tests have a high degree of validity and reliability. For value-added, it is also important that tests have “stretch,” that is, that they measure the full range of the material that is being covered. There must also be multiple test scores over several years for individual students. Finally, states must ensure that random measurement errors are kept to a minimum, especially if the value-added results are to be used as part of an accountability system.

- **Individual Student Identifier:** Value-added models require that each student has a unique student identifier (ID). About half of all students currently have or are working on establishing individual student IDs.

- **Statewide Data Collection Systems:** For value-added methodologies to work well, states need a robust data collection and management system that functions seamlessly from the school building to the state level. Such systems must be able to link program, course, and student data and enable users to efficiently exchange data electronically. As part of this system, states need reasonable processes to identify data that are in error, to spot check certain information randomly, and to conduct site visits to audit the accuracy of data at the local level.

**State Education System Considerations**

Value-added assessment models will make a number of demands on state education agencies that policymakers should attend to for the program to be effective. These include:

- **Staff Training:** Many of the experts consulted by the Study Group emphasized that value-added approaches are ill advised without a simultaneous commitment to a significant amount of training. Teacher and principal training on value-added assessment will need to fully inform staff about the basic concepts behind value-added assessment and what the data mean. Training should also provide principals and teachers with opportunities to practice analyzing real value-added assessment data, as well as putting the analysis to work in improving instruction.

- **Measurement Expertise:** States and districts will need to have a significant measurement and statistical capacity either in-house or contracted on a consulting basis. States should also consider using expert advisory panels both for evaluating the initial development plans for the value-added system and for periodically reviewing the quality and effectiveness of the system once it is up and running.

- **Costs:** While some value-added developers note that the “big cost” of testing is in the purchase, administration, and scoring of tests, which states already do, purchasing or developing a value-added system on top of the existing testing program can add significantly to an education budget. This is especially true when other needs such as developing a robust data management system, massive school personnel training, and ensuring statistical and psychometric expertise are included.

**Public Engagement and Political Considerations**

Many Study Group presenters emphasized the need to maintain trust and involvement with affected stakeholders—particularly teachers—throughout any effort to implement value-added analysis. It is clear from many educational innovations over many years that any number of players can cause a program to fail due to lack of understanding, ignorance of the process, or mistrust. Value-added assessment, rightly or wrongly, has developed the reputation as a high-stakes policy instrument. Thus, extra care must be taken to develop trust and buy-in. Following are some of the particular political and public relations areas that must be considered.

- **The Overall Impact of Changes to the Testing or Accountability System:** Adding yet another significant change to a testing or accountability system can be demoralizing to teachers and confusing to the public. Because value-added systems may require significant change and enhancements to existing accountability systems, states should carefully consider the purposes of the existing system and think through the adjustments to current efforts that will be needed to incorporate a value-added approach. In addition, potential users should recognize that implementing a value-added system cannot be done quickly.
• **State board of education planning and communications:** State boards will need to act as a voice of reason to help the public, school personnel, and lawmakers understand both the benefits and challenges of value-added models. Perhaps most importantly, state boards will need to exercise leadership and insist that decisions about value-added assessment should be made using reason and good judgment and not in response to persuasive sales approaches by vendors who are marketing commercial value-added testing systems.

• **Relations with Teachers:** Teachers, in particular, may be cynical about seeing yet another education reform heading their way. They may be suspicious that the “real” reason behind value-added analysis is punishing or rewarding teachers, and they may be wary of the amount of time it will take to understand and use value-added results. On the other hand, teachers may see the growth orientation of value-added assessment as an improvement over current accountability programs and they may be very interested in the potential of the system to improve instruction for classes and individuals. All in all, the Study Group believes that teachers (and principals) will respond much more positively to value-added models if the system is not seen as a stick, but primarily as a way to assist educators in improving instruction and promoting more growth in student learning—in short, that it can be an important tool in helping teachers do their jobs.

---

**Findings and Recommendations on Implementation Issues for Value-Added Assessments**

**Recommendation 7.** States considering a value-added system should develop and complete an “Implementation Checklist” as part of the initial planning process. Due to the complexity of value-added models, the demands they place on state testing and data systems, and the effects they can have on the overall education system, it is imperative that policymakers conduct a complete and honest evaluation of their current structures to ensure they meet the requirements for a value-added assessments. (See the Study Group’s checklist on page 37 as a starting point.)

**Recommendation 8.** Value-added assessment requires massive training for state, district, and school personnel. The Study Group believes that any state or local district that wishes to use value-added assessment needs to prepare for massive training of teachers and others in how to use the information generated. Teacher training on value-added assessment will need to inform teachers about the basic theory and methodology of value-added assessment. Training should also provide teachers with opportunities to practice analyzing real value-added assessment data, and develop skills regarding how to use the results with faculty and parents.

**Recommendation 9.** Users of value-added assessment must consider confidentiality and transparency issues. The Study Group recommends that states and districts considering value-added assessment develop appropriate policies to keep teacher ratings confidential, and limit their availability to the teachers themselves and authorized local school officials. At the same time, confidentiality must be balanced with useful reporting to educators and the public in order to attain sufficient transparency. Legal assistance may be needed to determine wording for policies in this area.

**Recommendation 10.** Teacher education programs should incorporate more training in educational measurement and the use of data into their curricula. The Study Group believes it is crucial that training regarding the basic concepts of status and growth measures, value-added assessment models, and the development and use of formative assessments be provided to prospective teachers in higher education. Those preparing to teach should also have opportunities to practice using data for decisions prior to classroom employment.
Recommendation 11. State policymakers and a wide range of stakeholders, including teachers, should be involved early in planning for value-added assessment. Legislative leaders and legislators serving on education committees are two groups that need to understand the general rationale of value-added assessment and understand what the model can and cannot do. In addition, educators need to know what concerns and complaints might be anticipated from parents or teachers so that they will be prepared to respond if they are contacted by constituents.

Recommendation 12. Educators should exercise due diligence in evaluating commercial companies offering value-added assessment products. The Study Group recommends that all policymakers who are considering using a value-added system be fully aware that they are facing an entrepreneurial environment. It may be true that many, if not most, of these vendors are trustworthy, but they are still vendors interested in selling something for a profit. It is also true that each vendor has limitations and expertise that need to be evaluated for appropriateness to the educational environment and the timeline of a given state or local school district. Consumers of value-added information and services need to do thorough due diligence before signing agreements to work with firms or individuals.

Recommendation 13. Financial costs must be considered carefully when considering value-added assessment. The Study Group recommends that state boards and other potential value-added assessment consumers be alert to the potential costs of this approach. While some experts note that the “big cost” of assessment is in the purchase, administration, and scoring of tests (which states already do), including a value-added system on top of a testing program can add significantly to an education budget. Users should do a cost/benefit analysis to determine whether the information they will receive justifies the price.

Recommendation 14. Value-added assessment needs continued pilot testing, research, and validation work. The Study Group is convinced that value-added assessment is a highly promising—although still immature—approach. While preliminary results show significant potential to improve education, continued pilot testing and research is needed. The Study Group believes that only with appropriate validation studies and other research will educators and policymakers get maximum benefit from the methodology.

Concluding Thoughts

In a sense, educators currently face a “measurement emergency.” They face increasing pressure from NCLB and other state or local accountability programs to demonstrate increased student learning. This pressure encourages the use of testing information to judge the performance—the value being added—of teachers and schools. With educators, policymakers, and the public becoming more measurement and outcome-oriented, it is natural that public school personnel and policymakers would seek additional ways to judge progress and use testing data.

The Study Group believes that value-added models—with their promise of attributing student growth to schools, teachers, or both—have the potential of offering a way to analyze student achievement in a more individual, robust, and understandable way than previous methods. As it is refined as a model, and if it can be adequately validated, it can lead us toward new information about the teacher skills, curriculum components, or program initiatives that are particularly effective in improving student learning. But, perhaps the most important feature of value-added assessment is that it serves to keep everyone’s focus on student growth and learning momentum, which is the essence of the schooling experience. After all, is it not the main role of educators to take children from where they find them and then “add value?”
ASBE’s Board of Directors first proposed the organization’s examination of value-added assessments in the summer of 2004. The Board’s action was precipitated by a number of factors:

- First, general interest in value-added assessment had been growing since it was first implemented in the early 1990s in Tennessee, based on the model’s promise of being able to distinguish the effects of a school and classroom on student learning.

- Second, the No Child Left Behind Act (NCLB), with its single focus on identifying student achievement levels, had led many educators to vigorously point out that student growth, especially for schools with many disadvantaged children, was also an important indicator of school success for accountability purposes; coincidentally, the testing regime required by NCLB meant that the yearly sequence of assessments needed to make value-added and other growth measurements work was now available to all schools covering grades three through eight.

- Third, some policymakers felt that the ability of value-added models to distinguish the amount of growth students gained from being in a particular classroom held great promise as an added tool for holding teachers accountable for how much learning was taking place in their classes—in short, for rewarding good teachers and helping (or eventually dismissing) poor ones.

- Finally, many policymakers and educators were also enthusiastic about the promise of value-added models to help principals and teachers better pinpoint where improvements in curriculum and instruction were needed.

The interest generated by these and other factors led NASBE to ask a number of important questions about value-added assessment models that were critical to education policymakers, including: How does value-added fit with other aspects of the state’s testing and accountability system? How accurate are the results? What are the best uses for value-added data and analysis in terms of school improvement, accountability, or other uses? What doesn’t value-added do well? What do states need to think about if they are planning to use value-added as a component of their assessment and accountability system?

In order to answer these questions and help education policymakers make informed decisions on these assessment models, NASBE established a special Study Group on Value-Added Assessment. Made up of state board members from 22 states as well as other selected educators and researchers, the Study Group began meeting in January 2005, with additional sessions held in April and June. In the course of their work, participants heard presentations from fifteen individuals who had expertise and special knowledge about this topic. Speakers represented a broad spectrum of interests, including researchers and psychometricians, testing companies, professional associations, state education agencies, local school boards, administrators, and the U.S. Department of Education. Based on these presentations and discussions with the speakers, and on various readings and internal discussions, the Study Group has developed this report. We hope that the report and its recommendations will be valuable to state board of education members, local boards, educators, lawmakers, and citizens as we all work toward the common goal of having better data upon which to make important educational decisions.

A special note of thanks goes to the Educational Testing Service (ETS). ETS provided financial assistance to the project to support costs associated with the meetings, and also lent the Study Group its considerable experience and intellectual capital related to large-scale assessments. However, it should be noted that the recommendations and substance of this report reflect the collective thinking of Study Group members and do not necessarily represent the views of ETS.
Chapter 1
Overview of Value-Added Assessment

A. Value-Added Assessment in Public Education

For decades, educators have been intrigued with the possibility of accurately determining the specific impact that schools or teachers have on student learning. To begin to measure this kind of impact, it is first necessary to calculate student progress during a school year. But even though it sounds simple, just accurately measuring a student's progress from the beginning of the school year to the end of the year is quite a challenge. A much more difficult task is measuring a student's long-term progress over many years of schooling. And an even more complex puzzle is attempting to determine whether a particular school or teacher has made an impact on a student's growth that is above or below what would be predicted or typical.

An approach called “value-added assessment” represents one model for measuring school performance that tries to address some of these issues. This concept entails estimating “the effectiveness of schools and teachers based on the amount of academic progress their students make from one year to the next” in an attempt to track “the value that schools add to individual students’ learning over time.” Stated another way, value-added assessment approaches do more than just report on how much a child has grown academically (as measured by test scores) from one year to the next. They typically use statistical models to attribute student progress to the school, to the teacher, or both. The value-added models look at students’ progress in a given year in the context of performance in previous years and, for some models, the students' background characteristics.

Instead of just comparing a student with other students (norm-referenced testing) or against a prescribed standard (criterion-referenced testing), value-added approaches take the added step of comparing a student with his/her past performance on tests. Value-added models use statistical procedures to calculate significant variances in a group of students’ test scores to determine how far the student has progressed compared to where the student started and to what degree that growth can be attributed to schooling (as opposed to “external” factors such as socio-economic status, race,

Origins of the Value-Added Concept

While the term “value-added” came into use in education in the early 1990s, it originated as an economic concept used in manufacturing and business. According to one industry group, the business definition of value-added is “making something into something better.” As an example, taking a bushel of grain and grinding it and packaging it so that someone can use it for baking increases the worth or “value” of the bushel of grain. Or taking logs and processing them into lumber again adds value. In the commercial world, the amount of value created is typically judged in terms of the potential financial profit added by a particular action, after accounting for the cost of the processing operation. Travelers will certainly recognize value-added as the basic component of the value-added taxes (VAT) used in Europe.
parents’ educational levels, or innate ability).

To use a specific classroom example, take an elementary school student who regularly scores at about the 60th percentile (nationally normed) on standardized tests. One would expect this student to learn enough to score at about the same level year after year, and that would be the expected growth. But suppose at the end of the 5th grade year the student suddenly jumps to the 75th percentile on the state test. In other words, for some reason, the student’s yearly growth is much greater than expected. For any one individual, there could be a range of explanations for this increase: the parents could have hired a tutor or used incentives to encourage the student to work harder, perhaps the student had always had slightly impaired hearing or eyesight or some other learning disability, and this problem was finally fixed. However, if in aggregate all the students in this child’s classroom also showed better than expected growth, then very likely there was something about this class—a very effective teacher or an improved curriculum—that led to the increased “value.” Value-added models are simply an attempt to systematize such reporting so that educators (and the public) can compare the growth that is expected from their students to the amount of growth they are actually getting—and ultimately to see if some schools and teachers regularly get better than expected growth, about what would be expected, or less than expected.

Regarding the results from value-added models, research analyst Karen Helland has written, “Because value-added assessment focuses on students’ rate of advancement, it provides a means of objective feedback on how well a particular teacher, school, or district is doing. This is an unprecedented development in assessment. The resulting feedback allows identification of high-achieving teachers, schools, or districts to more closely analyze the reasons for success.”

Or as Theodore Hershberg, professor of public policy and history and director of the Center for Greater Philadelphia has noted, the value-added concept centers on a disarmingly simple but crucial notion—schools cannot solve all of society’s problems, but “they can and should ensure that every child received a year’s worth of growth in a year.”

B. Value-added Assessment Viewed in the Context of Status and Growth Models

One way to better understand value-added assessment is to examine it in relation to other ways of viewing test scores. When policymakers and administrators make judgments about student performance, they generally use one of two basic approaches: status models and growth models. This section of the report provides a brief introduction to these models, and discusses value-added’s place as a “growth model-plus.”

Status Models

Status models (sometimes referred to as “attainment indicators”) have been the most common way of using assessment data. A status model simply asks where a student’s test score is in relation to a set of achievement standards (criterion-referenced tests) or in relation to other students (norm-referenced tests). For example, in one state a score of 65 or better on the 3rd grade reading test would mark the student as being proficient. A student with a lower score might be in the “basic” or “below basic” category. In other words, the score gives the student’s achievement status at a given point in time. These scores are generally aggregated for a whole class, school, district, or state, so one can say that a school’s status is that 75 percent of all 3rd graders had achieved the proficient level or better in reading. The current use of Adequate Yearly Progress (AYP) under NCLB is a classic example of a status model; it continually asks what percentage of students (both as a whole and disaggregated by subgroups) reach a certain level of proficiency; it does not ask questions about how much students grew over time. Status models, thus, are like a snapshot of a student’s performance at a given moment.

Various status models have been used by state education agencies for many years, and a number of states have gone beyond the most basic status model that simply reports performance in relation to an established proficiency standard. They have developed targets for how much improvement they hoped to see from year to year and compared the actual change in scores to this goal. For example, if 75 percent of this year’s 3rd graders achieved a level of “proficient” or better, the target for next year might be 80 percent of 3rd graders achieving proficiency. These types of status models provide more information than earlier reporting
A major challenge for the Study Group was to determine the definition of value-added assessment. Although there are many models of so-called “value-added” approaches, and while much has been written about the subject, it is not clear that everyone is talking about the same concept when they use the words “value-added.” To provide consistency throughout this report, unless otherwise noted, we are using the following working definition of value-added assessment models. While this is not a universally accepted definition, it represents, generally, the approach to which people are referring when they write or speak about value-added assessment.

Value-added assessment models are statistical approaches that use multi-year student test score data, and student background characteristics in some models, to attribute student growth to schools, teachers, or both.

We would note, however, that by using this definition, the Study Group does not imply that this is all that value-added models should or must be. Value-added assessment really should be viewed as an integrated system to support teaching and learning. The system must be built on an infrastructure that allows for the tracking of individual students over time; a data management capability that allows for the collection of a wide variety of educational variables—both student and teacher—that can be linked at the individual student level. The system must also administer high-quality assessments aligned to the state standards that reflect the breadth and depth of the curricula being taught and able to measure growth in student achievement over time. Finally, the system must employ a sophisticated model for calculating student growth over time and examining teacher and school contributions.

If we assume that growth models allow for the measurement of student gains in achievement over time, then value-added models are an added layer that allows for aggregation of student growth scores at the school or district level and for estimating the contributions of teachers and schools. Therefore, we can use results from value-added assessments at three levels: (1) monitoring the progress of each student; (2) supporting and evaluating individual teachers; and (3) reporting on the effectiveness of schools and districts.

Like other measurement models, value-added assessment is a “work in progress.” In future years the model will be tested, modified, and refined. Its methodology and assumptions will be tested by researchers. The model’s utility for providing useful information to parents and the public will be judged in the court of public opinion. Its efficacy for giving classroom teachers information that can improve student performance will be put to the test by teachers and principals. Local and state education officials will determine whether the costs associated with doing value-added assessment justify the information they obtain from the analysis. For-profit and non-profit organizations will continue to market value-added programs and some states and some school districts will purchase their products. Some states will go it alone and develop their own “home grown” value-added approaches. Through this convoluted but democratic process, we will develop more evidence regarding the merits of value-added assessment. Ultimately, the value-added assessment approach will compete in the education marketplace with other methods for analyzing test data for a place at the accountability and school improvement table.
Evaluating Value-Added systems, which simply announced test results with comparisons to national norms or some level of knowledge.

However, improvement is not the same as growth. As some researchers have noted, “AYP raises the bar continuously until the benchmark becomes all students proficient at the end of the 2013-14 school year. NCLB’s accountability system requires growth, but does not measure it.” The key point here is that these status models look for improvements from one year’s 7th grade class to the next year’s 7th grade class—they don’t follow the same 7th graders into the 8th grade to see how much these particular students, as a group or as individuals, have learned.

Advantages of status models are that they can keep educators focused on the high standards for student achievement. This approach avoids complacency that could come from being satisfied with moving forward in lieu of meeting a goal. In a sense, status models emphasize that education is serious business with serious expectations; they provide a sense of pressure, urgency and, perhaps, emergency. In addition, status models, if standards are applied universally, do not accept different achievement for different groups of students. Status models have appeal to the public and many policymakers because of the familiar concept of hitting a target or not.

Another advantage for education systems is that status models are fairly simple. They do not require yearly testing or tests over different grade levels that are well aligned. They do not need a lot of statistical machinery or data-processing power to make complex calculations, and so they are also transparent and easily understood by educators and the public alike.

However, status models also face various criticisms. Some who point out the limitations of status models note that despite the apparent specificity and clarity of an average test score, the number is actually not purely indicative of performance by the school system for a given year; instead, as Robert Meyer from the University of Wisconsin’s Center for Education Research writes, the score reflects “contributions to student achievement of schools and other non-school inputs to the learning process from multiple grades and multiple points in time.” Meyer also points out several specific reasons that status models using average test scores are highly flawed as school performance measures: (1) the score is contaminated by factors other than school quality—an example would be private tutoring; (2) basic average test score data is out-of-date, particularly in high school, where a 10th grade score reflects the cumulative data of K-10, some of which is 10 years old; and (3) and students move often and this mobility contaminates average test scores. Thus, a high school score may reflect the performance of three or four or more schools and several school districts.

**Growth Models**

As the name implies, growth models differ from status models in that they do not provide a snapshot of one point in time, but measure progress by comparing student achievement over two or more points in time. Growth models in general, according to Rolf Blank of the Council of Chief State School Officers, “refer to models of school accountability that measure progress by tracking the achievement scores or proficiency level of a student cohort from one year to the next with the intent of determining progress made.” An essential ingredient for using growth models is the availability of test data in consecutive grades. According to Study Group presentations, by 2006 all states should have assessment data for consecutive grades 3-8 in math and language arts. Much of this data has been collected in response to the No Child Left Behind legislation.

One of the important advantages of growth models is that they measure and compare exactly the same students as they move through schooling, rather than comparing this year’s 5th graders to last year’s 5th graders—two different groups of individuals. Another advantage of growth models is that they keep...
educators attuned to the changes in student learning over time and to progress made by individual students. This approach avoids complacency that could come from a single goal (such as reaching the proficiency level) indiscriminately applied to all students. Also, the concept of growth as an indicator appears to be well accepted by the general public. The 2003 Phi Delta Kappan/Gallup Poll, which surveys Americans’ views toward schools, found that 84 percent of the public stated that the best way to gauge the performance of a school was to measure individual student progress from where the child started.10

An additional reason for the strong interest in growth models is related to the No Child Left Behind Act and other state and district accountability systems that use a status model in requiring schools to meet specific student performance targets. This level of “high-stakes” testing prescribes various consequences if targets are not met, as well as making school scores known to the general public and policymakers. Such accountability requirements have caused educators to focus very specifically on ensuring that all students meet the performance targets (e.g., attain proficiency). While no one disputes the merits of this goal, there have also been several unintended consequences—problems that could be alleviated with the addition of a growth component.

The first unintended consequence rises from the fact that educators (or schools) do not receive official credit for any amount of student growth by itself unless the growth has led to meeting the performance goal. For example, “A 5th grader who moves from not reading to reading has made major progress, even if she remains far behind her age peers and is ranked not proficient on the 5th grade state test. However, under NCLB reporting rules, her meaningful growth is invisible.” 11 Thus, many argue that schools with large numbers of disadvantaged students who are starting well behind their more advantaged peers are being unfairly punished even when their students are showing substantial growth.

As this example shows, measuring growth can provide important information not explained simply by whether a student achieves a particular level of attainment. Thus, educators are seeking approaches that can give them credit for all the growth students make, whether or not it has led to meeting the performance objectives of NCLB.

While growth does not officially and technically equate to meeting the NCLB standard, many educators believe that demonstrating growth (using value-added or other growth models) will balance public criticism for schools that, while making progress, are not reaching the absolute performance bar of Adequate Yearly Progress under NCLB. They also believe that ignoring growth will demoralize teachers and will overlook excellent progress made in many classrooms that have students who started far behind their grade level.

A second unintended consequence noted by some educators and parents is that a focus only on achieving proficiency hurts those students who have already achieved proficiency—but could be learning much more. In this case, including a growth component as well as status in an accountability system would help ensure that schools pay attention to all students—to those who have already achieved beyond basic levels as well as to those who need help catching up.

Growth models in general also have weaknesses that need to be considered. While these models do highlight academic growth, they do not necessarily tell us whether proficiency standards are actually being met. Indeed, focusing on growth can distract educators and the public from the fact that standards may not be achieved soon or even—even though having students meet those standards (not growth in itself) is the ultimate goal of education. Some contend that without keeping this goal of proficiency for all clearly in sight, there will be large numbers of students who—even while showing growth—will eventually be left behind.

Others critics contend that growth models are more complex and less reliable than simply reporting on the status, or snapshot, of student

Schools cannot solve all of society’s problems, but “they can and should ensure that every child received a year’s worth of growth in a year.”

Theodore Hershberg, Director, Center for Greater Philadelphia
Evaluating Value-Added

Achievement. Still others maintain that growth models require additional costs that schools cannot afford, and point to the extra infrastructure requirements needed (including such things as accurate, reliable longitudinal data, computing power, and highly trained personnel to manage and explain the system) to determine student achievement growth accurately. Finally, other skeptics worry about the tendency for growth models to move the country toward using test scores to evaluate teacher performance, which they see as a dangerous slippery slope due to the inherent imprecision of testing data and data analysis for this purpose.

Value-Added as a Growth Model

It bears repeating here that value-added assessment models are a particular type of growth model that, in addition to the points above, has its own set of advantages and disadvantages. Value-added models have several features that set them apart from other ways of measuring growth. First, value-added reporting, as defined in this document, always starts at the student level. Scores may also be aggregated at the classroom and school levels, but the focus is on individuals. The second enhanced feature is that value-added models attempt to move beyond just reporting student achievement growth to actually isolating the “cause” of the growth.

According to some (including William Sanders, the principal architect of value-added assessments as used in Tennessee and other states, and the its best-known proponent), the greatest advantage of his value-added model is its ability to provide significant information about individual students. One weakness of a status-based accountability system such as NCLB is its focus on group or cohort achievement—no particular child has to make his or her target and teachers are not provided with rich information that shows individuals’ academic progress over the years. Sanders is convinced that value-added models represent a paradigm shift in their potential to focus on individual students. Value-added data, he says, allow users to drill down to see individual scores and run projections of what students are most likely to do in future years given current circumstances. Middle school and high school counselors, for instance, can be alerted to 5th graders with significant potential. The data can be a real boost to efforts to close achievement gaps in individual schools, and can be used to aid student placements in community colleges and higher education in ways that are more accurate than current methods. In all, Sanders claims that we have “not yet scratched the surface” in the use of value-added data for making projections and differentiating instruction.

Value-added models also have significant potential to help in school-wide improvement efforts. The information can be used to track much more closely those specific areas where student growth is occurring and where it is not, as well as which programs or curricula are most successful.

Another advantage specific to value-added models is their ability to discern the most effective and least effective teachers. While value-added models may not be appropriate for high-stakes teacher accountability at this time, they can be used to flag teachers who should be evaluated further to see if they need help, for identifying the most effective teachers who can then share their practices, and to assist superintendents and principals with staff deployment. (More detailed information on this and other uses of value-added models is available in Chapter 2.)

Finally, value-added data offer a wealth of opportunity for education researchers, particularly in the area of teacher impact on student achievement. Research to date using value-added data has provided convincing evidence that teachers have a discernible and differential impact on student achievement, and that these effects persist over time. For example, one widely publicized study from the Dallas public school system demonstrated that 90 percent of low-achieving 4th graders who were lucky enough to get three effective teachers in a row were able to pass the 7th grade math test by the time they were in that grade; on the other hand, of the low-achievers who were assigned to three ineffective teachers in a row, only 42 percent were able to pass the 7th grade test. If value-added approaches can be shown to clearly discern such dramatic differences among teachers, and if researchers can determine what causes these differences (admittedly a significant hurdle), then the potential to benefit education would be significant.

C. Value-Added Assessment in Practice

One of the best ways to understand and begin to evaluate value-added assessment is to examine how the approach has actually been used in
schools. A number of states and school districts have adopted various versions of the value-added approach. With limited space, we cannot include all these initiatives here, but a more detailed description of the Tennessee system follows. Other state and district examples are included throughout the report.

The Tennessee Value-Added Assessment System

Certainly the most prominent use of value-added assessment in this country has occurred in Tennessee with the use of the Tennessee Value-Added Assessment System (TVAAS). This system has been in place for approximately 13 years.

Developed by William Sanders in the late 1980s at the University of Tennessee, the value-added system was the centerpiece of the 1992 Tennessee Education Improvement Act. This strong accountability statute was developed in response to litigation filed on behalf of small, rural school districts and in response to urgings from the business community that the Tennessee state education system needed significant improvement. At the same time, then-Governor Lamar Alexander was seeking ways to evaluate teachers and saw the Sanders system as a possible approach.

The general purposes of the Tennessee Value-Added Assessment System are: (1) to help schools improve their educational programs; (2) to provide bonus funding to schools and districts that meet adequate progress and demonstrate high value-added effects; (3) to determine what technical assistance is needed for those in school improvement status; and (4) to determine the number, kind, and level of interventions selected by the state.

The Sanders model includes a software system that enables the estimation of district, school, and teacher effects on school achievement, and the data analysis involves the solving of thousands of equations iteratively. Under the Tennessee accountability system, value-added assessment is described as “a statistical system for educational outcome assessment which uses measures of student learning to enable the estimation of teacher, school, and school district statistical distributions.”

According to Sanders, the software enables a massive multivariate, longitudinal analysis using all achievement data for each student no matter how sparse or complete. Sanders maintains that this software solves the missing data problem; increases the sensitivity of estimations by using each student’s previous history in all subjects simultaneously; provides protection against the spurious misclassification of teachers; accommodates self-contained classrooms, team teaching, and departmentalized instruction; and allows for the inclusion of non-vertically scaled test data.

The Tennessee law goes on to say that the “system will use available and appropriate data as input to account

William Sanders on Value-Added Results

- “The evidence is overwhelming that if any child catches two very weak teachers in a row, unless there is a major intervention, that kid never recovers from it.”
- “I feel very strongly that even [the least effective teachers] should be given some assistance and some time to become more effective, or they should be encouraged to seek employment elsewhere. Because they’re harming kids.”
- “We have not yet scratched the surface on the use of data for making projections and differentiating instruction.”
- “Value-added models represent a paradigm shift in their potential to focus on individual students.”
- “The falloff in gain at 6th grade for higher achieving students is very great... Our failure to sustain the growth of high-achieving poor and minority kids has major implications for attempts to close the achievement gap.”

Sources: 
** W. Sanders, Presentation to NASBE Study Group (18 April 2005).
for differences in prior student attainment, such that the impact which the teacher, school and school district have on the educational progress of students may be estimated on a student attainment constant basis.” As the model is implemented, estimates of average student achievement progress are calculated for every school and teacher in all the local school districts in Tennessee. Summary results are presented via reports that give estimates of the growth in student achievement. In addition, the reports contain the amount of growth that has been produced by each local school administrative unit, each school building, and each teacher. All these reports have been made public except for the teacher reports; these are only provided to the teacher and his/her supervisor.

Some Tennessee schools seem to have used the value-added program with great success. At West End Middle school in Nashville, the school was on the state’s low-performance list several years ago and the value-added scores were also very poor. The principal and staff reviewed all the test data and created a supportive learning environment in which teachers were not afraid to share information and learn from each other. They met in teams and focused faculty meetings on academics and instructional issues. Now, 84 percent of West End students are proficient in math and 85 percent in reading. Furthermore, the school scored all A’s on value-added.20

Parents say they are totally confused because some schools with high scoring students get low ratings because the schools do not show enough improvement under the TVAAS program. For example, some parents of students at Annie Berry Elementary School are angry because even though the school posted scores of 95 and 97 for the percent of its students who were proficient in math and reading, it got an ‘F’ for value-added. The measure for whether students are making enough growth each year. The principal of the school says that his staff just doesn’t understand how the value-added statistics are derived. He is also very frustrated that the value-added score can’t give him any suggestions about what his school should do differently to improve.23 Parents and staff at other seemingly successful schools that have gotten very low value-added grades have had similar complaints.

In response, supporters of the TVAAS maintain that almost all students should be showing significant growth, even if they are scoring in the top levels. Supporters go on to say that a low value-added score for high-achieving schools simply means the students are covering too little material too slowly. Thus, they aren’t learning anything new.24

As evidenced by the frustration and confusion of both parents and professionals, it seems obvious that parents and educators need better training on how the value-added system works and how to interpret the ratings the schools receive. Because of the frustration, Tennessee has now expanded its training significantly beyond what was originally provided.

In summary, the Tennessee model has met with mixed results. Moreover, the future of the program is uncertain. Just recently, because of criticism of the program, Tennessee’s $1.4 million value-added testing program has been put under the spotlight by state senators who have proposed taking money used to support value-added testing and using it to support the governor’s preschool program. Whereas Tennessee’s value-added assessment program cost approximately 60 cents per student in 1995, the cost 10 years later is closer to $1.50 per student.25

Lessons Learned from Users of Value-Added Assessment

From the various states and districts that have experimented with value-added assessment, there are a number of lessons already learned. It is evident that, to succeed, a value-added assessment program must provide for adequate personnel training. Adequate funds must also be budgeted. The value-added assessment program must be explained to the public and preferably pilot-tested prior to full implementation. Considerable attention must be given to involving and working with faculty for the program to be accepted. Parents and policymakers must understand what the program can do and what it cannot accomplish. Adequate data collection and robust tests must be available for success in implementing the value-added approach.

In the following sections of the report, we will examine more closely the specific uses and limitations of value-added models, as well as the many issues related to implementation that states should be aware of both before and as such a system is being put in place.
Chapter 2
The Uses and Limitations of Value-Added Assessment

Value-added assessment models hold significant promise to help education systems at all levels improve their performance—yet in practice there are also serious technical and policy issues that hinder or even make unwise some of the uses of value-added data. The Study Group gave lengthy consideration to these issues, which represent the heart of the Group's recommendations. This chapter looks in detail at some of the specific uses for which value-added assessments have been proposed.

1. Using value-added as a component of school accountability

If, as research has borne out to some degree, value-added assessments are able to distinguish the effects of teachers and schools on student achievement, then using value-added as a component of a state's accountability system would seem to make perfect sense. Indeed, that purpose is specifically written into the legislative language in several states—including Tennessee and Ohio—that are using value-added in holding schools accountable. The Ohio Revised Code specifies, for example, that the state “shall incorporate the value-added progress dimension into the report cards and performance ratings issued for districts and buildings.” It also specifies, indirectly, that it is the Sanders model that will be used.

While such legislative language appears straightforward, devising the details of such a system is quite complicated, both on policy and technical levels. Following are some of the policy issues that must be considered.

a. What is the unit of measurement? In other words, what or who is being held accountable, and does current research support the use of value-added models for that purpose? The choices for state policymakers are clearly school districts, school buildings, and teachers. As seen above, Ohio legislators opted for “districts and buildings,” and Tennessee has brought value-added accountability to the school level for over a decade. Researchers are currently mixed as to whether value-added models, at this time in their development, are wholly appropriate for this level of accountability. Brian Gong, director of the Center for Assessment and a frequent consultant to states on their assessment systems, told the Study Group that “simple growth models [are] most appropriate for school accountability,” while “more complex, ‘conditioned’ value-added models are less appropriate as the main models for school accountability,” though they “may be useful for supplemental accountability.”

b. How much growth is enough? Growth can be measured in several different ways, and policymakers must decide which method to use. Gong explains that there are two basic sources for developing growth targets:

1) Data-driven estimates for what has happened historically or with a normed group. In this case, a state might look at what the growth for reading has been over a three-year period in each grade. This rate of growth becomes the baseline or expected growth rate in the future. The key advantage of a data-driven estimate is that it is based on reality. Expected growth rates are also the same for everyone, making the system more easily understood. The key disadvantage is that this method addresses what has been, not what should be. It is
unlikely, for example, that the average growth rate of a state over the past 10 years will bring all students to proficiency by 2014.

2) Policy-driven growth targets. These methods of deciding growth targets are generally anchored in long-term goals (such as all students reaching proficiency) that have been developed and approved through a social process (such as by a state board of education). There are a number of methods that can be used—including growth through achievement levels, using scaled scores, or ensuring that individual students are “on track” to reach a proficiency target—each method requiring a number of judgment calls by policymakers. For example, should schools or students who start at a lower point be required to show more growth (in order to “catch up”) than schools or students starting at a higher level?

While the key advantage of policy-driven expected growth is that it solves the problem in data-driven methods of getting all students to proficiency, there are a number of disadvantages as well. First, not being based on “reality,” it is often unknown whether the targets set through the policy-driven method can actually be reached. Second, because these methods are just now emerging (most growth or value-added systems currently in use are data-driven), there is not a long history of evaluation through which the statistical packages can be validated.

It should be noted that some researchers recommend combining these two ways of defining growth so that, as Gong explains, “policy-driven growth targets are informed by data.” Such a method would be subject to ongoing monitoring and modification to accommodate checks on reality. 27

c. Treatment of growth versus achievement. A third policy area is deciding how much to value student or school growth versus actual levels of achievement. Mitchell Chester, Ohio’s Assistant Superintendent for Policy and accountability, uses the diagram below to describe the basic relationships that can exist between growth and achievement. In an accountability system, policymakers must decide how much “weight” to give each component and answer such questions as, should a school be sanctioned if there is not sufficient growth, even if achievement levels are relatively high? Conversely, should a lower achieving school be rewarded for substantial growth even if large numbers of students are below proficient?28 (See diagram below left.)

2. Using Value-Added for Teacher Accountability and Evaluation

Since value-added assessment models are frequently portrayed as a way to distinguish the effects of individual teachers and classrooms on student achievement, it is not surprising that many politicians and others are tempted to use them as a vehicle for teacher accountability. As the Study Group heard over and over, however, for a host of statistical and other reasons this area must be approached with considerable caution, especially in terms of high-stakes decisions such as ranking teachers, merit pay, and promotion or dismissal.

The first problem is a statistical one. Given margin of error considerations that are built into testing systems, by the time one gets to judging classroom effects, it is very difficult to accurately identify any teacher but those in roughly the top or bottom 10 to 15 percent. Even William Sanders, in his appearance before the Study Group, noted that he is most confident of his data at the extremes; in other words, he feels fairly
confident that using his value-added model he can identify the very best and very worst performing schools and classrooms. In fact, regarding identifying cases at these extreme ends of the continuum, he likened himself to an actuary, saying that he may not be totally accurate, but “don’t bet against me.” On the other hand, he said he is much less confident that his model can accurately detect significant performance differences among the 70 to 80 percent of teachers who fall somewhere in the middle range. See graph on page 23 for a depiction of how standard error makes it virtually impossible to distinguish teachers in the middle range.

Researchers have noted a number of other challenges in trying to use value-added data for teacher evaluations. For example:

- Students move from place to place (sometimes with multiple moves in one year), a condition that particularly affects schools in low-income areas.
- Students can have multiple teachers in one year, even in elementary schools, and if team teaching is used, this only adds to the problem.
- Many subjects are not tested as part of the state assessment system.
- Special education teachers (or other specialists) may not work with a specific class of students, or students with disabilities may be excluded from the data.
- Positive teacher effects as evidenced by higher value-added scores is not the same as teacher effectiveness. For example, is one teacher who is good at narrowing the curriculum and teaching to the standardized test really better than another teacher who gets a lower score but delivers a broader, richer curriculum that includes material that is not tested?
- Missing data, errors in the statewide database, testing and scoring errors, and test cheating all add to problems of validity in the system.

In sum, most of the concerns about using value-added models to evaluate teachers center around how certain we can be that the “cause” of the effects seen in the data (i.e., the differential achievement gains shown by the students of different teachers) is really the varying effectiveness of the teachers. Braun of ETS points out that the only way we can “unequivocally interpret the result of a statistical analysis as a causal effect” (e.g., a particular teacher caused a certain amount of growth in her class) is through conducting a “large well-designed randomized experiment.” But because of how schools operate, he says, because “teachers are not randomly assigned to schools and students are not randomly assigned to teachers... you can’t be sure that the observed results are caused by the [teacher]. They might be related to initial differences in students and the class as a whole.” Indeed, Sanders'
research from Tennessee found that high-achieving students are much more likely to have highly effective teachers than low-achieving students, and that the test scores of students who have either a very good or bad teacher can linger for years. Such confounding factors mean, as Kupermintz notes, that “it is unclear whether these results reflect systematic inequalities in the allocation of teachers to students or a possible misattribution of teacher effects.”

This doesn’t mean that value-added might not play some role in teacher evaluation, however. What the Study Group heard again and again is perhaps best summed up by the most extensive evaluation of value-added models to date, by RAND Education, which concluded that while “the research base is currently insufficient to support the use of value-added models for high-stakes decisions,” they do show “promise for lower-stakes diagnostic purposes” such as initially identifying possibly low- or high-performing teachers who can then be further evaluated to confirm results. In this way, value-added functions as a filter for detecting those teachers who would be subject to additional study through classroom observations, diagnostic tests, and portfolios. If further study confirms that a teacher is struggling, he or she can be counseled, provided professional development, or matched with a teacher who has been confirmed as highly effective.

C. Using Value-Added as a Tool for School Improvement

The Study Group was enthusiastic about the use of value-added models as a data-driven component of efforts to improve instruction at the classroom, school, and district levels. Indeed, many believe that this is the most significant advantage of value-added models. Following are some of the key areas that can be informed by using value-added data in conjunction with other measures. In addition, see the box below for a description of how Pennsylvania is implementing a value-added system focusing on school improvement.

- **Policy and program evaluation.** Value-added provides policymakers, administrators, and teachers with a “more accurate picture of what works and what does not” when looking at the effectiveness of reform efforts.

- **Identification of students in need.** Value-added data allows teachers to more precisely pinpoint those students or groups of students who are struggling, as well as identifying higher achieving students who may not be progressing as fast or as far as they could.

- **Schoolwide and team planning.** Many users of value-added point to its significance in this area. As one district superintendent appearing before the Study Group said, “[Value-added] provides important data about student progress that can be used to engage teachers in meaningful conversations about needed changes to the curriculum, instructional practices, and assessment techniques to increase student achievement. It stimulates reflection on the best practices to enhance student learning.”

---

Horizontal line represents national norm gain. “Fuzzy” line represents margin of error around the actual 3-year average student gain. Results are from 1996. From W. Sanders, presentation to the NABE Study (14 April 2005).
Testing to Inform Instruction

Schools and districts face the practical timing issues of providing value-added assessment information to support teachers in their daily instruction. The most common practice with value-added assessment is to have data provided to local schools/districts once per year, usually based on statewide testing. While this process can provide useful information for teachers and curriculum supervisors, it does little to assist teachers on a day-to-day basis. Thus, teachers could proceed through many months of an entire school year using an ineffective practice or materials with no value-added feedback except once a year.

The Northwest Evaluation Association (NWEA), a nonprofit group, tries to remedy this problem by providing an enhancement to traditional value-added assessment methods. Formed in 1974, NWEA is a partnership between Portland-area school districts and the Seattle Public Schools. It works on contract with 1,300 school districts in around 40 states. NWEA develops tests called Measures of Academic Progress to track academic progress on a quarter to quarter basis. Thus schools and districts know every quarter how much growth has occurred, without having to wait until the end of the year. With computerized results NWEA reporting provides results the day after the test is given, versus some state results that may not come out for 4 to 6 months or later after tests are given. Thus, schools/districts can respond in a more business-like fashion and give students the benefit of a system that quickly responds to their needs. In addition to the prompt results, NWEA tests also provide districts with more detailed results than typical standardized tests.

Individualized professional development. Moving away from one-size-fits-all inservice training, value-added data can help schools pinpoint the areas where individual teachers need more assistance. The Dallas, Texas school system, for example, uses value-added analysis to produce “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 to engage in a program of professional development aimed at strengthening their skills in areas of weakness.”

Resource Management. Value-added data can help district and school-level administrators better allocate resources to areas of need and to those efforts that have been shown to produce the greatest influence on student achievement.

While the assessment data provided by value-added models can be an important component of school and teacher improvement efforts, the Study Group would like to emphasize one key limitation: Value-added and other growth models, while they can help pinpoint what students are and are not learning and how much the students are progressing, cannot tell educators why growth is not occurring in a particular area or what to do about the problem. Additional analysis and the use of ongoing formative assessments are needed to actually alter instruction to help individual or groups of students learn more. Dylan Wiliam, an ETS research analyst, used a baseball analogy to illustrate this point. Basic monitoring of learning (discerning whether learning is taking place), is like telling a pitcher more batters are getting hits off you, he told the Study Group. Testing at the next more-diagnostic level (what is not being learnt) is like telling the pitcher “it’s your curveball that’s the problem.” But true formative feedback actually says what to do about the problem, like telling the pitcher, “you’re dropping your shoulder when you throw your curveball.” Wiliam emphasized that “to be formative, the assessment must include a recipe for future action” in order to adapt instruction to meet learning needs.

This last point is key if policymakers expect a value-added model to truly lead to improvements in achievement. The whole enterprise must be grounded in support and other mechanisms that lead to effective use of the data. As Sanders himself noted to the Study Group, after 12 years of use in Tennessee, only “about 40 percent of schools are aggressively using the data” the system supplies.
In Pennsylvania: Value-Added for School Improvement

In 2002, the Pennsylvania State Board of Education required that value-added assessment be implemented statewide. A pilot program began in the same year with 32 school districts, and other pilot districts are being phased in gradually. Beginning in the fall of 2006, all 501 districts will receive value-added reports (comprehensive reports for the pilot districts, limited reports for all other districts). In the fall of 2007, all districts will receive more comprehensive reporting. The state has contracted with William Sanders and SAS for using the EVAAS version of value-added assessment.

PVAAS is intended as a tool for data-informed decision-making, using score results in reading and math for grades 3-8. Analysis is available for individual students, and is also aggregated by grade, by school, and by subgroups. The reports provide detailed information about the progress rates of an individual school relative to a pool of districts with similar off-grade testing. They also present a table of available student scores in the subject tested, followed by the student’s probability of achieving a particular level of proficiency on subsequent Pennsylvania System of School Assessment (PSSA) tests; a graph also accompanies each table. Pennsylvania has avoided some of the sticky politics of value-added assessment by pledging not to use the model to evaluate teachers. Indeed, the state is very clear that “PVAAS is not a tool for teacher accountability.”

The state encourages superintendents, principals, and teachers to use the PVAAS in a wide range of school improvement efforts, including:

- Tracking student growth to ensure that all students (high and low achieving) and all subgroups of students are making appropriate levels of progress;
- Analyzing the influence of specific programs and initiatives on students’ progress;
- Identifying areas of strength and weakness in grade levels and in specific subject areas;
- Identifying groups of students and individuals who are at risk for not meeting school, district, and state academic expectations;
- Using value-added data as a part of the school-planning process, for teacher team meetings, and for student level planning;
- Using value-added projection data to allocate district and school-level resources for student supports (such as tutoring) or teacher professional development; and
- Working with grade-level and subject-area teams to reflect upon current practices in curriculum, assessment, and instruction.

Source: Pennsylvania Department of Education. Online at www.pde.state.pa.us/a_and_t/cwp/view.asp?a=108&Q=108916&a_and_tNav=6429]&a_and_tNav=].

D. Using value-added data for improving teacher training

Given the uneven quality of graduates from the nation’s teacher training institutions, states have long struggled with finding mechanisms to hold preparation programs accountable for results, or at least to get a clearer picture of the effectiveness of new teachers and to help shore up deficiencies in their training. Last year, Louisiana became the first state to use value-added assessments as a vehicle for linking teacher quality back to teacher preparation programs.42

Initiated by the Board of Regents in Louisiana, the state’s focus on using value-added measures to monitor teacher education programs exemplifies the realization that teacher quality begins prior to the teacher’s first day on the job and that policymakers cannot just deal with improving teachers currently in the classroom. Part of the impetus, according to the education committee chairman of the Louisiana House of Representatives, is making “the colleges of education... part of our overall accountability.
system.” Or as Kati Haycock, Director of the Education Trust, pointed out regarding the program, “What I think Louisiana has said is: Why fiddle around with more proxies [such as licensing exams]; why not look at what matters most, which is when they get into the classrooms are they actually prepared to help students learn more.”

George Noell, a researcher in the Department of Psychology at Louisiana State University and A&M College, is testing three value-added models using achievement data for students in grades 4 through 9 from 10 school districts. Dr. Noell’s preliminary findings indicate that it will be possible to use student achievement databases in Louisiana to assess the effectiveness of teacher preparation programs. To date, he has found that greater growth was demonstrated by children taught by teachers from certain teacher training institutions. Further, this preliminary analysis also identified one particular teacher preparation program whose teachers seemed to even surpass the mathematics learning growth produced by experienced teachers.

Meanwhile, in Ohio, there is also an effort to use a value-added approach as a way to gain further information about characteristics of effective teaching. The University of Cincinnati, The Ohio State University and the University of Dayton are heading up a related study which examines the preparation, in-school support, and effectiveness of Ohio Teachers. The project involves all 50 of the state’s colleges and schools of education. Perhaps because the data from the project will not report on individual districts, teachers or programs, there is support from the Ohio Education Association and the Ohio Federation of Teachers.

In summary, if schools of education can use value-added assessment identify the “value” a particular teacher is adding to the student’s learning and to determine what teacher behaviors and training are responsible for the “added value,” then there is great promise of using this information in training teachers to use these effective techniques.

E. Using value-added or growth measures as a component of AYP calculations

As discussed earlier, one of the reasons for interest in value-added and growth measures of achievement is as a counter to the NCLB emphasis on attainment or status. Many educators and policymakers feel that strict, yearly adherence to proficiency targets unfairly punishes those schools dealing with large numbers of disadvantaged students, because they are required to catch up, often very quickly, to schools that start with higher achieving students. As stated by the Council of Chief State School Officers, the NCLB accountability model does not always value progress with regard to subgroups or schools starting significantly below proficiency or with improved student performance at other levels of achievement (e.g., movement from proficient to advanced and below basic to basic). Furthermore, the NCLB’s safe harbor provision is hampered by trying to measure small changes in student performance, for small numbers of students, over a small amount of time. States are committed to meeting AYP for all students, but the current system neither sufficiently values progress for lower-performing subgroups nor accounts for the degrees by which groups do not make AYP.

Some organizations are advocating other, even broader and bolder uses of the value-added assessment approach for program evaluation in higher education. The John William Pope Center for Higher Education Policy has called on higher education to institute a comprehensive focus on “academic value-added.” Further, it maintains that willingness to implement such an approach should be a major evaluation criterion in the selection of college presidents. Whether the value-added movement will reach this far into higher education is unclear, but it is evident that the concept of value-added has an appeal beyond K-12 education policies. Source: Online at www.popecenter.org/issues/article.html?id=1575.
Evaluating Value-Added

Nancy Doorey, a former teacher and former member of the Delaware State Board of Education who is now president of a local school board, made the case to the Study Group that not only are some schools unfairly penalized, but that these penalties may actually be counter-productive to the goal of having all students reach proficiency. Noting that high-poverty schools regularly lose 20 percent or more of their teachers every year (and that most teachers who leave a school move to lower poverty schools with fewer minorities), Doorey said that the current system further “stacks the odds against diverse and poor schools, creating an incentive for teachers to avoid or leave such schools... Using status, teachers making superior gains with more challenging students and a more transient student body say they are humiliated rather than praised, and their school could be reconstituted. Why would they stay?”

In addition, this problem is likely to get worse the closer we get to 2014, as schools will be required to meet higher and higher proficiency targets. Many states are looking into ways to add growth or value-added measures to AYP calculations, and some have already applied to the U.S. Department of Education for permission to use these systems. A few states, such as Massachusetts, early on received permission to use a simple “growth” model (see box), though this is still based on looking at percentages of students within successive cohorts reaching proficiency rather than a system that encompasses individual student growth. Early in her tenure, Secretary of Education Margaret Spellings signaled a willingness to consider growth as an AYP component, provided it is still based on all students reaching proficiency. The Department is currently (during the summer and autumn of 2005) convening a task force of researchers, policymakers, and practitioners to examine the issues. But given the

---

**Growth and AYP in Massachusetts**

In Massachusetts, making adequate yearly progress is based on meeting targets for participation rate in the testing system, as well as attendance or graduation rate targets, plus meeting the targets for school and district scores on the state Composite Performance Index (CPI). The CPI is based on a point system where schools are awarded points for each student scoring in the advanced/proficient range, needs improvement-high, needs improvement-low, etc. The state has calculated yearly targets based on steady increases until all students would score in the advanced/proficient range by 2014. The growth safety net is that each school and district also has its own trajectory to 100 percent proficiency based on where it started. Thus, if a school misses the state’s target score in a given year, but its score shows that it is still making enough improvement that it will reach 100 percent proficiency by 2014, then that school’s CPI score is considered sufficient for AYP purposes. The formula is as follows:

\[
AYP = A + (B \text{ or } C) + D, \text{ where}
\]

\[
A = \text{ Participation Rate: Are all (or almost all) students taking part?}
\]

\[
B = \text{ Performance: Have we met the state's target Composite Performance Index (CPI) for the current review period?}
\]

\[
C = \text{ Improvement: Is our rate of improvement such that all students will reach proficiency by 2014?}
\]

\[
D = \text{ Performance or Improvement on Additional Indicator:}
\]

- K-8: Does our attendance meet the state target or represent a one percent improvement over 2003?
- High School: Did our Class of 2004 meet the state graduation rate target?

Source: Massachusetts Department of Education.
many technical considerations and the number of different models being developed, progress is likely to be slow—and meanwhile all decisions on state growth plans have been put on hold. One of the key sticking points is whether any growth model used as part of AYP must remain within the letter of the NCLB Act or whether it could exist within the spirit of the law.

Having heard from and considered a wide range of evidence and positions, the Study Group holds that growth indicators must be allowed as a component of AYP calculations within the No Child Left Behind framework.

While the rigorous targets for attainment required by NCLB serve an important purpose in keeping a focus on the ultimate goal of high achievement for all students, adding growth is needed not only out of fairness, but because together these indicators provide the most accurate picture of the effectiveness of schools.

Summary

Overall, the Study Group was enthusiastic about the potential of value-added assessment data to be a significant tool in school improvement efforts. The Group also felt that it could have a place as one component of school or district accountability systems, or even as one piece of a larger teacher evaluation process. But the Study Group cautioned that value-added is not at this time—and perhaps never will be—ready for sole use in high-stakes decisions.

Finally, the Study Group found that value-added assessments are generally highly complex and come with a large set of other issues that must be dealt with before, during, and after implementation of these models. These implementation issues, which policymakers must be aware of before going down the value-added path, are explored in the next chapter.

Findings and Recommendations on the Uses and Limitations of Value-added Models

**Recommendation 1.** Value-added assessment is not designed for high-stakes use in teacher evaluations. The Study Group recommends that value-added information not be used for high-stakes teacher evaluation involving either rewards or punishments. We believe that educators should recognize that value-added assessment is a “tool” but it is not the “total.” Because some research shows that teacher effectiveness can be 1 to 20 times more powerful than income, class size, race, or family background, policymakers may feel an urgency to implement approaches such as value-added in an attempt to identify and reward their best teachers. However, it must be remembered that other research shows that these value-added teacher effects are not so easy to identify, and indeed that the data can only with certainty identify about the top 10 percent and bottom 10 percent of teachers. Researchers also note that the number of studies supporting the value-added methodology is relatively limited and that there is a serious lack of validity studies. Thus, users are strongly encouraged to exercise caution in any use of value-added assessments so that the proper use of the technique does not go beyond the capabilities of the tool.

**Recommendation 2.** Value-added assessment has significant potential—when used in conjunction with other measures and supports—as a tool to improve teaching. The potential of value-added data to 1) differentiate the very most and least effective teachers and 2) show individual teachers more precisely in which areas and with which students they are and aren’t being successful can be a valuable tool in helping all teachers improve. One principal explained how value-added can work in practice this way: “Every year I sit down with the data, looking at which teachers are the most effective in each subject, and where teachers have a chance to improve. Say teacher A is getting 150 percent of the expected growth for 4th graders in reading, while Teacher B is only getting 85 percent of the expected growth. I’ll have those two teachers sit down together and talk about their teaching practices and what material they’re covering, so Teacher B can learn from Teacher A and find out why she’s so successful.”


Recommendation 3. Value-added assessment information can be a powerful school improvement tool in promoting more effective practices and resource allocation. If a state or district uses value-added information, the Study Group recommends that the information be used for diagnostic purposes geared toward strengthening school performance and teaching techniques. Indeed, perhaps the most potent potential of the value-added model is through its power to start the dialogue about improving teacher effectiveness. At the same time, principals and school teams can use value-added data to identify effective or ineffective programs and curricula, identify individual students or groups of students who need additional support to reach their potential, and target school resources to areas in need. As one Pennsylvania principal said, "We are using [value-added] to look at individual students and in developing strategies to improve student performance through differentiated instruction and remediation programs. It also allows us to look at student subgroups and school entities to measure how each group or school is performing" and then target resources to "provide the supports necessary to help the subgroups, as well as each school entity." 52

Recommendation 4. Once-a-year value-added analysis does not substitute for other ongoing testing information that helps teachers adjust and target instruction to meet the needs of their students. While the Study Group believes that value-added assessment models have significant potential, the Group also recommends that schools avoid reliance on information from just a yearly testing program. Teacher-designed tests, commercial diagnostic measures, and other interim assessments done throughout the school year are crucial to keeping schools on target and can allow for mid-course corrections as needs arise. Waiting for results from one statewide testing program is inadequate for improving classroom instruction.

Recommendation 5. Value-added models show potential for tracking and improving the effectiveness of teacher training institutions. Using value-added data with newly licensed teachers has the potential to provide policymakers with information on which preparation programs are producing the most and least effective young teachers. Preparation programs can also use this information, especially if used alongside interviews with the teachers, to pinpoint weaknesses in their curriculum and other aspects of their program. However, as with other areas, the Study Group does not believe that value-added assessment is appropriate at this time as a stand-alone, high-stakes measure for evaluating teacher training institutions.

Recommendation 6. The Study Group urges the U.S. Department of Education to allow growth indicators as a component of AYP calculations within the No Child Left Behind framework. While the rigorous targets for attainment required by NCLB serve an important purpose in keeping a focus on the ultimate goal of high achievement for all students, adding growth is needed not only out of fairness, but because together these indicators provide the most accurate picture of the effectiveness of schools. In addition, failure to use growth as one indicator of success could end up making it even more difficult to retain effective teachers in disadvantaged schools.
Chapter 3
Implementation Issues for Value-Added Assessment

Because value-added assessment models are so new to the education enterprise, because they tend to be extremely complex in their mechanics, and because there are several potential high-stakes implications for their use, state boards of education and other decision-makers will find there are a host of issues that must be considered before, during, and after implementation of value-added systems. While some of these are technical issues that must be “checked off” as the state proceeds, the Study Group found that other technical, system-related, and political issues have the potential to be “deal breakers” or to render the value-added models far less effective than they otherwise might be. This chapter examines the three main implementation areas (testing and data issues, issues related to the education system, and political considerations), and concludes with a checklist of issues states and districts can use as they contemplate or begin planning the use of value-added assessment models.

1. Key Testing and Data Issues Affecting Implementation

As noted by Kupermintz, “Advances in testing practices, psychometric and statistical modeling, as well as in longitudinal data collection, management, and maintenance have ushered a new generation of value-added models, specifically designed to support accountability systems by providing information on educational productivity.”

Despite this sophistication, however, Thomas Fisher, a former director of assessment for Florida, told the Study Group that value-added analyses are believable only when the underlying testing program has technical quality, produces valid scores, and maintains an accurate longitudinal database. He cautioned that very likely no state has yet achieved all these standards, and that no amount of sophistication of value-added assessment models can make up for these problems.

a. Testing issues. The Study Group recognizes that a variety of testing issues are relevant to whether value-added assessment is a viable model. Listed below are several important educational measurement issues that must be considered when analyzing value-added models. We have attempted to present these issues in brief and simple form; those desiring more information are referred to the Standards for Educational and Psychological Testing (developed jointly by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education), which is the generally accepted reference regarding these issues.

- **Test Validity:** Tests used for value-added assessment should withstand validation scrutiny. Validation refers, in the most simple terms, to whether the test measures what it purports to measure and whether the test is appropriate for the purpose for which it is intended to be used. Establishing test validity is a complex process and will always involve professional judgment as well as the collection of data and accumulation of evidence.

- **Test Reliability:** Tests used for value-added assessment purposes...
should be reliable. No test is a perfect measure; all contain some error, even if it is minor. Several sources of error exist: some of the major sources are imperfections in the test itself, factors in the students taking the test and scoring procedures. A test might be reliable for one group, but not another. One would hope that if test takers repeatedly took a test that the results would be consistent or dependable on each administration. Such results would generally be viewed as "reliable" with minimum error.

- **Test Stretch:** Value-added assessment systems require that the tests being used have adequate "stretch." The concept of stretch refers to the breadth or range that a test can measure. As stated by van der Ploeg and Thum, "Tests should measure using units that... span the full range of the construct being measured.... Small changes for those just gaining skill in the construct must be just as measurable as small changes among those highly skilled."

- **Longitudinal Data Requirements:** Value-added assessment requires that multiple test scores be available over time for individual students. Because of NCLB and other state testing programs, states will increasingly have longitudinal data available.

- **Measurement Error:** All measurements, including tests, are somewhat inaccurate due to random error. Even physical measures such as rulers and temperature gauges are not exact and perfect. In education, the goal is to develop tests that have as little error as possible or to use analysis procedures that minimize or neutralize the errors that do exist. Especially if value-added assessment is used to make high-stakes judgments, it is important to use tests with minimal random error.

**b. Unique statewide student identifier.** When each student has a unique student identifier (ID), it is then possible to match individual student records across various K-12 education databases, converting "snapshots" of students into a "motion picture" of progress over time. Longitudinal systems required to do value-added assessment type analysis require at least two consecutive years of data points (scores) for students, but more observations are definitely recommended. Where states have a unique identifier for each student in the state, students can be followed even if they move to new schools. Thus, this will significantly minimize the loss of data. Therefore, theoretically, the only students for whom test data is lacking will be those who came from out of state. The Council of Chief State School Officers reports that about half of all states either have or are now working on developing unique student identifiers.

**c. Statewide data systems.** For value-added methodologies to work well, states need a robust data collection and management system that functions seamlessly from the school building to the state level. Such systems must be able to link program, course, and student data and enable users to efficiently exchange data electronically. As part of this system, states need reasonable processes to identify data that are in error, to spot check certain information randomly, and to conduct site visits to audit the accuracy of data at the local level. Some states have found it necessary to impose consequences for submitting incomplete, untimely, or inaccurate information. Texas, for instance, in 1999, publicly judged two school districts as "unacceptable due to data quality" under the state accountability system.

**d. Transparency:** Many have criticized the value-added assessment approach (particularly Sanders' TVAAS model) for being a "black box" of mysterious statistical calculations that cannot be tested for verification by other researchers. Some have contended that to conduct a proper validity investigation, Sanders must make his TVAASS data available to other researchers. According to education researcher Kupermintz, at least as of late 2003, his numerous requests to Sanders for information have been met with "blank refusals, offering no other reason than a concern that 'the data may be misused.'" It has also been reported that researchers such as Robert L. Linn from the University of Colorado and researchers at the Carnegie Foundation have asked for data from Sanders only to be turned down or stalled. Since Sanders has the most well-known value-added assessment model, it is natural that other researchers want access to his data. As long as this data are not forthcoming, researchers will not have confidence in the TVAAS findings.

Transparency is not only important for research and validation, however. On a very practical level, those using the information (policymakers, administrators, teachers, parents) must also feel confident about the
value of the statistical methodology and the data they are using. Thus, how the system works must be explained as clearly as possible to lay people who are not statisticians or psychometricians.

e. **Missing data:** Problems of incomplete data represent one of the biggest challenges facing those who wish to use value-added assessment. Achievement data can be missing for individual students for many reasons. Also, information which pairs students with particular teachers can be incomplete. Whether value-added models can adequately deal with missing data is debatable. Sanders, in his study group presentation, maintained that his model can handle this issue.

f. **Statistical complexity:** One factor that makes value-added assessment appealing to some is the availability of advanced statistical techniques, powerful computing power, and large data sets of multi-year test results. In the U.S., complex statistical procedures are often employed to get value-added information. For example, according to Dr. George Nell from Louisiana State University, “...the strongest value-added approaches use a cross classified HLM or mixed model approach.” Improvements in computing speed and programming allow these analyses to be performed quickly and more cost effectively than in past years. Those who use more elaborate analysis, such as Sanders, believe that their results are superior.
to simpler methods. Depending on where and how value-added approaches are implemented, sophistication in data analysis can vary greatly. Speaking to the Study Group, Siobhan Leahy and Dylan Wiliam from ETS explained that Great Britain has been using a value-added approach for many years nationwide with very simple statistics, basically comparing percentages. Whether the accuracy of the findings from more elaborate methods of analysis justifies the cost is difficult to determine. More research on value-added analysis will be needed to answer this question.

2. State Education System Considerations

a. Training requirements for value-added assessment. The Study Group believes that any state or local district that wishes to use value-added assessment needs to prepare for massive training of teachers and others in how to use the information generated. Many of the presenters to the Study Group emphasized that value-added approaches are ill advised without a simultaneous commitment to training. Teacher and principal training on value-added assessment will need to fully inform staff about the basic concepts behind value-added assessment and what the data mean. Training should also provide principals and teachers with opportunities to practice analyzing real value-added assessment data, as well as putting the analysis to work in improving instruction. In addition, staff will need to develop skills regarding how to use the results with faculty and parents. If policymakers are not prepared for the financial and scheduling commitment to make this training available, there is little likelihood that value-added assessment can succeed.

b. Evaluation of commercial companies offering value-added assessment products. The Study Group urges all policymakers who are considering using a value-added system be fully aware that they are facing an entrepreneurial environment. Many companies and consultants are offering to assist educators with value-added issues and programs. It may be true that many, if not most, of these vendors are trustworthy, but they are still vendors interested in selling something for a profit. It is also true that each vendor has limitations and expertise that need to be evaluated for appropriateness to the educational environment and the timeline of a given state or local school district. Consumers of value-added information and services need to do thorough due diligence before signing agreements to work with firms or individuals. It is always wise to check with prior customers, examine appropriate company documents, and use appropriate legal assistance as needed when evaluating various vendors.

c. Measurement personnel at state education agencies and local school districts. As one education writer has said, “Like the makers of hot dogs...testing experts know too well what goes into the creation of achievement tests.” Their intimate knowledge of the technical difficulties involved in measuring student achievement makes a number of these testing experts some of the most vocal (and persuasive) opponents of testing. Such healthy criticism can be helpful in the implementation of testing programs, including those with a value-added component. For example, as noted by M cAdams, test data can be given more weight than it deserves and administra-
tors might use test data inappropriately to make personnel decisions. Thus, state and local educators rely on their research and testing staff to alert users to the caution that should be exercised in using testing data so that policymakers don’t misuse value-added assessment information.

In addition to the quality control that research and testing personnel can add to state and local education agencies, adequate personnel are also required for the many administrative tasks and teacher training associated with value-added assessment. Even if value-added assessment efforts are contracted, there are still numerous duties for testing staff related to implementation, oversight and coordination of contract matters. A serious problem in this area is that state and local districts often do not have adequate budgets to afford to employ adequate research and testing staff. In some downsizing efforts, these are the first positions to be cut. They are often slated for reduction or elimination with the argument that these positions do not provide “direct support in the classroom” or that they are just symptomatic of a “bloated school bureaucracy.”

Another problem is the lack of value-added assessment expertise in general. As Arie Van der Ploeg, a senior researcher at Learning Point Associates, notes, “the number of persons in the United States who understand value-added analysis, growth modeling, and schools...is far too small to support an industry. We badly need more trained and more experienced expertise in place to assist schools, districts, and state education agencies.” Thus, states and local districts must argue early and often for the substantial research and testing staff that they will need if they implement a new value-
added assessment effort. Otherwise, without adequate support for testing, contracting, training, and other necessities, the value-added experience is not likely to be successful.

If sophisticated in-house psychometric and statistical capacity is not available, states and districts will likely need to secure consultants or use expert advisory panels that are knowledgeable about measurement issues in general and value-added assessments in particular. It is also possible that the objectivity of outside assistance may increase the trust level within the state regarding the process. If vendors will be used, states should do careful due diligence before making selections.

d. Costs. The Study Group recommends that state board of education and other potential value-added assessment consumers be alert to the potential costs of this approach. While some Study Group presenters (such as Sanders) noted that the “big cost” of testing is in the purchase, administration and scoring of tests, purchasing or developing a value-added system can add significantly to an education budget. This is especially true when other needs such as developing a robust data management system, massive school personnel training, and ensuring statistical and psychometric expertise are factored in. Users should do a cost/benefit analysis to determine whether the information they will receive justifies the price.

3. Public Engagement and Political Considerations

Successfully implementing a value-added system requires policymakers to conduct a major outreach effort and to carefully consider the political ramifications of the new system in order to ensure buy-in from stakeholders. Many Study Group presenters emphasized the need to maintain trust and involvement throughout any effort to implement value-added measurement. Representatives from the major teacher’s organizations were particularly clear about the fact that teachers perceive themselves as having the most to lose professionally with a value-added approach if it is carried out improperly. Widespread involvement is also needed to gain the insights and suggestions from various parts of the education community. It is clear from many educational innovations over the years that any number of players can cause a program to fail due to lack of understanding, ignorance of the process, or mistrust. Value-added assessment, rightly or wrongly, has developed the reputation as a high-stakes policy instrument. Thus, extra care must be taken to develop trust among all stakeholders. Following are some of the political and public relations areas that must be considered.

a. The overall impact of changing the testing program.

Lurching from one testing or accountability system to another (or continuously piling on new components) can be demoralizing to teachers and confusing to the public. States should consider the purposes of the existing system and think about what adjustments might need to be made to current efforts to incorporate a value-added approach. Changing testing programs is somewhat like the proverbial turning of a battleship—it is a major effort involving many individuals. In addition, potential users should recognize that implementing a value-added system cannot be done quickly. Value-added systems may require significant change and enhancements to existing accountability systems. It is crucial that any developmental work be started early; all participants in the process must realize that they are likely looking at years—not months—to develop, field test, and implement a high-quality system.

b. State board of education planning and communications.

State boards of education will need to consider how to involve the governor, legislators, and other policymakers in the state as value-added models are being considered. State boards must be careful about “over-promising” what can be gained from value-added data so that lawmakers are not disappointed when results become available. The public and the press also need to understand the value-added concept so that public information about results can be properly understood.

In general, state boards will need to act as a voice of reason to help the public, school personnel, and lawmakers understand both the benefits and challenges of value-added models. Perhaps most importantly, state boards will need to exercise leadership and insist that decisions about value-added assessment should be made using reason and good judgment and not in response to persuasive sales approaches by vendors who are marketing commercial value-added testing systems.

c. Considerations involving teachers.

Teachers are likely to have several reactions to value-added assessment. Some may see value-added assessment as another in the long line of “education reforms” that they have seen come and go throughout their career. Many of these
educators have had less than satisfactory experiences with past reforms; they may have seen reforms that have been under-funded, thrust aside due to political considerations, poorly planned, or improperly implemented.

Some teachers are likely to be suspicious that the “real” reason for value-added assessment is that policymakers want to use the results to punish or reward teachers. These more experienced teachers have an understanding of how enormously complex teaching is, and they are skeptical that a valid way has finally been found to accurately measure what they do. They are also worried about how much of their time it will take to learn about value-added concept and how to use the results to improve instruction. They likely feel that policymakers do not have a full appreciation of how hectic their schedule is and how little time is left for “just one more thing” like value-added analysis.

On the other hand, teachers may see the growth orientation of value-added assessment as an improvement over current accountability programs. For an accountability system to be accepted by teachers it must, among other things, speak about individual student growth, in addition to student performance. Thus, teachers may be interested in the “potential” of value-added assessment, but desirous of an exceptionally careful implementation and roll out. Teachers would likely be more accepting of using value-added models for lower-stakes, non-punitive diagnostic purposes. In this regard, the Study Group reiterates its recommendation from the previous chapter, that value-added assessment should not be used as a single tool for teacher and school accountability. If value-added assessment is to be used in any sort of accountability fashion, it should only be one of several indicators. This position is consistent with the Standards for Educational and Psychological Testing, which specify that “Using a single test to measure success or to sanction students, educators, schools, or districts is an inappropriate use of a single instrument.”

It is also important to involve teachers in the process of developing and implementing the value-added system, both in its overall design and at particular points along the way. For example, it is critical to involve teachers in deciding the question of how learning growth will be considered within the system and how much growth will be seen as “enough” for individual students as well as groups of students and schools.

All in all, teachers (and principals) will respond much more positively to value-added models if the system is not seen as a stick, but primarily as a way to assist educators in improving instruction and promoting more growth in student learning—in short, that it can be an important tool in helping teachers do their jobs.

Concluding Thoughts on Student Growth and Value-Added Assessment

In a sense, educators currently face a “measurement emergency.” They face increasing pressure from NCLB and other state or local accountability programs to demonstrate increased student learning. This pressure encourages the use of testing information to judge the performance—the value being added—of teachers and schools. With educators, policymakers, and the public becoming more measurement and outcome-oriented, it is natural that public school personnel and policymakers would explore business systems (such as the value-added approach) used by successful industries to judge progress and to seek additional ways to use testing data that are available.

But finding better tests and better data analysis is not a simple matter. As some researchers have said, “Psychometrists and statisticians talk about reliability, validity, and inferential risk. They make use of highly sophisticated quantitative tools and obscure acronyms: item response theory, differential item functioning, Monte Carlo simulation, hierarchical linear models (HLM), ANOVA, MANOVA, LISREL, Type II errors, the list goes on.” However, in spite of all this sophistication, a former U.S. Secretary of Education has noted that, “Most of the data we need, we cannot get. Most of what we get, we cannot trust. Of that which we can trust, far too much is obsolete, unintelligible to laymen, or unsuited to crucial analyses and comparisons.”

The value-added assessment growth model has the potential of offering a way to measure student growth in a more individual, robust, and understandable way than previous methods. As it is refined as a model, and if it can be adequately validated, it can lead us toward new information about the teacher skills, curriculum components, or program initiatives that are particularly effective in improving student learning. But, perhaps the most important feature of value-added assessment is that it serves to keep everyone’s focus on student growth and learning momentum, which is the essence of the schooling experience.
After all, is it not the main role of educators to take children from where they find them and then “add value?” Teachers add value by teaching various subjects. They add value by modeling appropriate behavior. In fact, good teachers add value in a host of measurable and unmeasurable ways. As Jim Mahoney, Executive Director of Battelle for Kids, has said, “As human beings, we need opportunities to grow, succeed, and celebrate. I am not talking about minimum goals or low aspirations. I am talking about setting reasonable expectations and acknowledging progress the way we acknowledge achievement. You can’t get through the door of achievement without also going through the door of progress.”

**Study Group Recommendations: Implementation Issues for Value-Added Assessments**

The following recommendations focusing on value-added implementation continue the findings and recommendations from Chapter 2 (page 27).

**Recommendation 7.** States considering a value-added system should develop and complete an “Implementation Checklist” as part of the initial planning process. Due to the complexity of value-added models, the demands they place on state testing and data systems, and the effects they can have on the overall education system, it is imperative that policymakers conduct a complete and honest evaluation of their current structures to ensure they meet the requirements for a value-added assessments. (See the Study Group’s checklist on page 37 as a starting point.)

**Recommendation 8.** Value-added assessment requires massive training for state, district, and school personnel. The Study Group believes that any state or local district that wishes to use value-added assessment needs to prepare for massive training of teachers and others in how to use the information generated. Many of the presenters to the Study Group emphasized that value-added approaches are ill advised without a simultaneous commitment to training. Teacher training on value-added assessment will need to inform teachers about the basic theory and methodology of value-added assessment. Training should also provide teachers with opportunities to practice analyzing real value-added assessment data, and develop skills regarding how to use the results with faculty and parents.

**Recommendation 9.** Users of value-added assessment must consider confidentiality and transparency issues. The Study Group recommends that states and districts considering value-added assessment develop appropriate policies to keep teacher ratings confidential, and limit their availability to the teachers themselves and authorized local school officials. At the same time, confidentiality must be balanced with useful reporting to educators and the public in order to attain sufficient transparency. Legal assistance may be needed to determine wording for policies in this area.

**Recommendation 10.** Teacher education programs should incorporate more training in educational measurement and the use of data into their curricula. Colleges must improve teacher training by having teacher education graduates ready to use value-added and other assessment results. The use of data will only increase in future. Therefore, the Study Group believes it is crucial that training regarding the basic concepts of status and growth measures, value-added assessment models, and the development and use of formative assessments be provided to prospective teachers in higher education. Those preparing to teach should also have opportunities to practice using data for decisions prior to classroom employment.
Recommendation 1. State policymakers and a wide range of stakeholders, including teachers, should be involved early in planning for value-added assessment. The Study Group recommends that states planning to implement value-added assessment involve major policymakers early in discussions of their plans. Legislative leaders and legislators serving on education committees are two groups that need to understand the general rationale of value-added assessment and understand what the model can and cannot do. Legislators also need to be aware of the costs of using the value-added approach and how value-added assessment will mesh with the existing state education programs. In addition, educators need to know what concerns and complaints might be anticipated from parents or teachers so that they will be prepared to respond if they are contacted by constituents.

Recommendation 2. Educators should exercise due diligence in evaluating commercial companies offering value-added assessment products. The Study Group recommends that all policymakers who are considering using a value-added system be fully aware that they are facing an entrepreneurial environment. It may be true that many, if not most, of these vendors are trustworthy, but they are still vendors interested in selling something for a profit. It is also true that each vendor has limitations and expertise that need to be evaluated for appropriateness to the educational environment and the timeline of a given state or local school district. Consumers of value-added information and services need to do thorough due diligence before signing agreements to work with firms or individuals. It is always wise to check with prior customers, to examine appropriate company documents, and use appropriate legal assistance as needed when evaluating various vendors.

Recommendation 3. Financial costs must be considered carefully when considering value-added assessment. The Study Group recommends that state boards and other potential value-added assessment consumers be alert to the potential costs of this approach. While some experts note that the “big cost” of assessment is in the purchase, administration, and scoring of tests (which states already do), including a value-added system on top of a testing program can add significantly to an education budget. Users should do a cost/benefit analysis to determine whether the information they will receive justifies the price.

Recommendation 4. Value-added assessment needs continued pilot testing, research, and validation work. The Study Group is convinced that value-added assessment is a highly promising—albeit an immature—approach. While preliminary results show significant potential to improve education, continued pilot testing and research is needed. Many kinds of inquiries would be helpful. For example, it would be interesting to compare the utility of the complex TVAAS/EVAAS approach of William Sanders to simpler approaches such as the British model. Finally, the Study Group believes that only with appropriate validation studies and other research will educators and policymakers get maximum benefit from the methodology.
Checklist for Policymakers Considering Value-Added Assessment

Growing out of the recommendations and findings of this report, the Study Group provides the following list of issues states and local school districts should consider when thinking about using value-added assessment. While this checklist does not cover every item that might be considered, it does raise many of the major issues that should be reviewed prior to moving ahead with a value-added assessment effort.

1. **We know exactly why we are considering value-added assessment.** It is critical that policymakers think clearly about exactly what they want from a value-added system. The features desired and how the system will be used must be specified up front. States must know what they want the value-added assessment model to do and what questions you want answered; laying out these issues in a matrix may be helpful.

2. **The impact of changing our accountability system has been fully considered.** Changing from one system to another (or continuously piling on new components) only demoralizes teachers and confuses the public. States should make sure that value-added assessment is carefully blended with any existing accountability systems. States should consider the purposes of the existing system and think about what adjustments might need to be made to current efforts to incorporate a value-added approach. Potential users should recognize that implementing a value-added system cannot be done quickly. It is crucial that any developmental work be started early; all participants in the process must realize that they are likely looking at years—not months—to develop, field test, and implement a high-quality system. If value-added assessment is worth doing, it is worth doing well.

3. **Our data system is robust enough to support value-added assessment.** States should have comprehensive, reliable data systems that can track individual student information. The system should link subjects taken, teachers, classrooms, departmental priorities, and the school.

4. **Our students have unique student identifiers.** States should assign each student a unique statewide student identifier that can be used to match records accurately across databases and years. State must also ensure that when a student moves he/she keeps the unique student ID number.

5. **Our tests are valid and reliable.** States should be using well-evaluated, valid, and reliable tests that minimize errors of measurement.

6. **Our tests contain items that challenge a range of students.** States should use tests with a range of item difficulties and complexity to represent a wide spectrum of the curriculum. States may have to create statewide assessment tests that can be used for value-added assessment—current basic skills tests may only test for the lowest end of curriculum, and may not be adequate to measure the performance of top performing students.

7. **Our tests match our curriculum.** States should make certain that tests are sensitive to what teachers are asking students to learn. For value-added assessment to be meaningful, the tests used for value-added assessment must have a high overlap with the content of the curriculum being taught. It would make little sense to spend time testing and analyzing data that did not match what teachers were teaching.

8. **We can configure our test data in a variety of ways.** States should have carefully designed data structures that provide flexibility to handle varied sources of inputs, data disaggregation, and longitudinal multimeasure data.

9. **We understand the political and policy implications of reporting data.** When considering value-added assessment, policymakers must be aware of issues related to the unit of analysis. There are confidentiality and public relations implications for collecting and reporting data at the state, district, classroom, or school level. As in the Tennessee implementation of TVAAS, states may wish to make some results available only to teachers and their supervisors. States need to consider whether any teacher effectiveness information is considered to be personnel information and treated accordingly. Legal consultation is recommended in developing this communication plan.

10. **We know what level of student achievement growth is appropriate.** States should consider what represents appropriate growth standards by subject/grade. Just knowing that one school or teacher is associated with more growth is not adequate. States should determine what degree of growth is reasonable, compared to some standard. For example, states could determine what amount of growth...
Evaluating Value-Added

would be needed to keep students “on-track” for graduation or for meeting a certain level of performance such as an NCLB standard. In addition, the state should ensure that teachers have been involved in developing the growth standards.

11. **We have a well-designed plan for handling missing test data.** States should include in any planning a strategy for how to handle students who are not tested. States should also have plans to help schools and districts minimize missing data. As we have discussed in this report, missing data is one of the most crucial issues in determining whether value-added assessment data are accurate.

12. **We know how we will report both growth and achievement results so as to minimize confusion about these two outcome measures.** As states consider value-added approaches, they should plan for how they will report and use both growth and achievement results. These considerations are important for NCLB, and they are also major issues for reporting to legislators, parents, and the general public.

13. **We have developed a comprehensive implementation plan for our value-added assessment program.** States considering value-added assessment should develop a detailed implementation schedule and a list of all stakeholders with whom they plan to collaborate. States should clearly define in detail the precise tasks that need to be accomplished to implement a value-added approach and specify which individuals in the organization have responsibility for various tasks. Finally, states should have a well-developed public information plan for value-added assessment that includes how they will communicate the results of a value-added program.

14. **We have made extraordinary efforts to involve a variety of stakeholders.** Many Study Group presenters emphasized the need to have trust and involvement of the state’s education professional organizations and other stakeholders throughout any effort to implement value-added measurement. Widespread involvement is also needed to gain the insights and suggestions from various parts of the education community. Value-added assessment, rightly or wrongly, has developed the reputation as a high-stakes policy instrument. Thus, extra care must be taken to develop trust among all stakeholders.

15. **Our value-added assessment program will be pilot tested.** It is crucial that any value-added system be pilot tested before the full roll out of the program. Gathering lessons learned from other states currently using the value-added approach is also invaluable.

16. **We understand all the cost issues associated with implementing a value-added assessment program and funds are available to pay for the program.** Value-added assessment will cost money, perhaps a great deal of money. It is not unusual to see references to costs approaching $2 per student. Thus, careful financial planning is required. Value-added assessment programs should not be undertaken without adequate resources.

17. **We have carefully considered the various value-added models and vendors available.** While the William Sanders TVAAS/EVAAS model is the best known of the value-added models, this model requires complex computer analysis of data. However, there are various other approaches. In England schools have been using a type of value-added assessment for many years, but their system is based on very simple calculations. Policymakers must decide which of many value-added assessment models meets their criteria.

18. **We have access to high-quality technical expertise to ensure quality control.** Unless states have a very sophisticated in-house psychometric and statistical capacity, they will likely need to secure consultants or use expert advisory panels which are knowledgeable about measurement issues in general and value-added assessments in particular. Value-added issues can become quite complex, and expertise is needed. It is also possible that the objectivity of outside assistance may increase the trust level within the state regarding the process. If vendors will be used, states should do careful due diligence before making selections.

19. **We are not relying only on a “one-test-per-year” approach for school improvement information.** States should not expect significant improvement in teaching and learning if they are just using value-added information from a single yearly test. Schools also need to use classroom assessments and other approaches that give feedback to teachers and principals throughout the year.
Endnotes


5. K. Helland, “Value-Added Assessment.”


12. R. Blank, Growth Models Primer.


14. B. Gong, presentation to the N ASBE Study Group on Value-Added Assessment (Alexandria, VA, 10 June 2005).

15. Battelle for Kids, Project SOAR. Available online at www.battelleforkids.org/brk/rt/about/our_work/improve/SOAR.

16. M. Chester, presentation to the N ASBE Study Group.

17. W. Sanders, presentation to the N ASBE Study Group.


22. W. Sanders, presentation to the Study Group.


24. Ibid.


27. Ibid.

28. M. Chester, presentation to the N ASBE Study Group on Value-Added Assessment (Alexandria, VA, 10 June 2005).

29. Ibid.

30. M. Chester, presentation to the N ASBE Study Group.

31. W. Sanders, presentation to the N ASBE Study Group.


33. H. Braun and L. Goe, presentation to the N ASBE Study Group.


35. Battelle for Kids, Project SOAR. Available online at www.battelleforkids.org/brk/rt/about/our_work/improve/SOAR.

Evaluating Value-Added

available at www.nwea.org/assessments/


39 Drury and Doran, “The Value of Value-Added Analysis.”

40 D. Wiliam, presentation to the NASBE Study Group on Value-Added Assessment (Princeton, NJ, 18 April 2005).

41 W. Sanders, presentation to the NASBE Study Group.


44 Ibid.

45 Louisiana Board of Regents document. Available online at www.regents.state.la.us/school.htm.

46 T. Hershberg, “Value-Added Assessment.”


50 H. Kupermintz, “Teacher Effects and Teacher Effectiveness.”

51 K. Carey, “The Real Value of Teachers.”


53 H. Kupermintz, “Teacher Effects and Teacher Effectiveness.”

54 More information about the Standards for Educational and Psychological Testing is available online at www.apa.org/science/standards.html#overview.


56 W. Sanders, presentation to the Study Group.


58 C. Daugherty, Nine Essential Elements

59 H. Kupermintz, “Teacher Effects and Teacher Effectiveness.”

60 Ibid.


63 A. van der Ploeg, personal communication to authors.


66 T. Blanford, National Education Association representative, in presentation to the NASBE Study Group on Value-Added Assessment (Alexandria, VA, 10 June 2005).

