

Goals and Uses of Value Added Models

Discussion by Brian Stecher

Committee on Value-Added Methodology for Instructional Improvement, Program
Evaluation and Accountability

November 13, 2008

The theme of this session is the goals and uses of value added models for a variety of purposes, including (1) school and teacher improvement, (2) school and teacher accountability, (3) program evaluation, and (4) research. It seems more logical to me to turn the question around and ask what data are needed for each of these purposes and whether there is likely to be any value-added from using value added models. Before delving deeply into technical and psychometric issues (in later sessions), it is useful to begin the workshop by looking carefully at these potential purposes and elaborating the basic questions that are being asked, the types of information that are relevant, and the instances where value added methods might play the greatest role. I also think it is important to differentiate among uses related to students, teachers, and schools/districts, and to consider another purpose—monitoring—as the starting point. I have less to say about data for program evaluation or research, and will concentrate on the following three questions:

- What outcome information will help to monitor accumulated performance?
- What outcome information will help to improve performance in the future?
- What outcome information can serve as the basis for incentives to motivate better performance?

I hope the following discussions will prompt further thinking about the purposes of outcome measures and the relationship between purposes, measures and analytic methods. This preliminary analysis is not a reaction to another paper but a guess about the sorts of questions that might be relevant to the committee but do not appear to be addressed in later sessions. I suspect it contains as many errors of omission as of commission.

Monitoring

Before thinking about the use of data for improvement or accountability it is helpful to examine the use of data for description, i.e., what are the outcomes of schooling at a particular point in time. Table 1 lists some of the features of information for monitoring performance at different levels of the system. There are two key questions: how well has a student mastered a body of material, and how does one student's performance compare to a set of expectation or to the performance of other students. Criterion referenced measures serve the first purpose well and both

criterion and norm referenced measures serve the other . In most cases, student monitoring occurs infrequently, but it is important that the information be comparable and standardized.

Table 1
Outcome Information Useful for Monitoring Performance

Level	Key Questions	Useful Measures	Key Features
Student	<p>What accumulated knowledge and skills does the student possess?</p> <p>How does one student’s accumulated performance compare to expectations/other students’?</p>	<p>Standards-based, CRT mastery</p> <p>Norm-referenced performance</p>	<p>Infrequent, standardized, cumulative</p>
Teacher	<p>What is the distribution of the accumulated performances of students in the class?</p> <p>How do the accumulated performances of the class compare to expectations for the performances of similar classes?</p>	<p>Distribution of standards-based, CRT mastery</p> <p>Distribution of norm-referenced performance</p> <p><i>Controlling for student background factors</i></p>	<p>Infrequent, standardized, cumulative</p>
School/ District	<p>What is the distribution of the accumulated performances of students in the school?</p> <p>How does the distribution of accumulated performances differ across classes?</p> <p>How do the accumulated performances of the school compare to expectations for the performances of similar schools?</p>	<p>Distribution of standards-based, CRT mastery</p> <p>Distribution of norm-referenced performance</p> <p>Variation in distributions of classroom performance</p> <p><i>Controlling for student background factors</i></p>	<p>Infrequent, standardized, cumulative</p>

For monitoring teacher performance, measures of aggregate student outcomes are needed, and it is important that they be comparable across teachers. In addition,

when describing the performance of teachers (or school and districts) it is often desirable to attempt to control for differences in student characteristics. At the district or school level, it is also important to examine variation among teachers. Thus, the analytic demands become more complex as the level of aggregation increases.

Improvement

When the goal is to improve performance, different types of data may be more relevant. Table 2 shows some of the issues related to the use of outcome data for improvement. For example, improving student performance is an ongoing task that benefits from a continuous flow of information. Teachers can re-teach content, revise lesson plans, reallocate time across students, direct students to supplemental resources, engage parents, and take other actions to promote student learning. In almost all of these cases, their efforts will be enhanced by detailed information about individual students, information that is easy to obtain, related to specific content, locally scored, etc.

In contrast, school and teacher improvement does not rely on such frequent outcome data. For example, teacher improvement occurs episodically, through professional development workshops, summer institutes or periodic collaborative study of student outcomes. Similarly, principals have limited choices when it comes to improving school performance, and the changes tend to occur infrequently. They can change curriculum materials or programs, attempt to change teachers' practices through professional development, reallocate resources to specialists or mentors, or change staffing through performance reviews and hiring practices. Some of these actions are fostered by content specific information (e.g., targeted professional development); other actions do not require this level of detail (e.g., hiring and firing staff).

As before, the analytic questions grow more complex at higher levels of the system. Value added information about the effectiveness of specific teachers or schools would be helpful in targeting improvement efforts, but it is not very useful for designing specific interventions. Information about practice may be as important as information about outcomes when the focus is on improvement.

Table 2.
Outcome Information Useful for Improving Performance

Level	Key Questions	Useful Measures	Key Features
Student	<p>Which prerequisite skills does the student possess?</p> <p>Has the student mastered current skills?</p> <p>What misconceptions (errors) does the student hold (make)?</p>	<p>Mastery of pre-requisite skills</p> <p>Partial/complete mastery of current skills</p>	<p>Frequent, embedded, flexible, locally scored</p>
Teacher	<p>What knowledge and skills do students learn when taught by the teacher?</p> <p>How does the distribution of student learning compare to expectations for the performance of similar students with other teachers?</p>	<p>Distribution of gains in students' mastery of standards/skills</p> <p>Distribution of gains in students' norm-referenced performance</p> <p><i>Controlling for student background factors</i></p>	<p>Less frequent, content-related, comparable, value-added</p>
School/ District	<p>What knowledge and skills do students learn when attending the school?</p> <p>How does the distribution of student learning differ across classes?</p> <p>How does the distribution of student learning compare to expectations for the performances of similar schools?</p>	<p>Distribution of standards-based, CRT mastery</p> <p>Distribution of norm-referenced performance</p> <p>Variation in distributions of classroom performance</p> <p><i>Controlling for student background factors</i></p>	<p>Less frequent, content-related, comparable, value-added</p>

Accountability

The third purpose is accountability, which I take to mean the allocation of incentives on the basis of outcome measures. Incentives might be sanctions (as in the case of NCLBO, cash rewards (as in the case of pay for performance), non-monetary recognition (as in the case of many state accountability programs). Some accountability systems are based on achieving a particular level, others on demonstrating adequate gains. The latter is often preferred on the grounds of fairness, motivation, etc. Even in a gains based system, it may be desirable to make adjustments for prior performance and/or background. These might include adjustments to targets or adjustments to the measures themselves. Clearly value-added models are attractive in the context of gain-based accountability, particularly if they can produce estimates that are comparable across different conditions. To be effective such measures must be valid, fair and be accepted by stakeholders in the system.

In summary, this analysis illustrates the point that the goals of monitoring, improvement and accountability raise fundamentally different questions about school performance, which are best addressed by different kinds of data and methods of analysis. To oversimplify, information demands relating to students are often different and less complex than for teachers, schools or districts. Value-added models may be helpful in the context of improvement for identifying strong and weak performers, but their most appropriate use may be in the context of accountability.

Miscellany

Easton (2008) claims that “predominately African American schools will have uniformly lower value-added estimates than schools and classroom with students of other races/ethnicities because of the consistent achievement gaps across the school district.” This occurs because VAM techniques that do not adjust for race are negatively biased against teachers who teach high proportions of students from racial/ethnic groups that perform less well on the outcome measure.

Is this understanding of VAM correct? My limited knowledge of the research suggests that the answer is “yes,” “no,” or “maybe” depending on whom you ask, which models are used and which data are examined. This situation presents a problem for the use of value added models. If they cannot be made understandable to potential users and if they cannot be communicated clearly to the public, then they meet the same fate as the New Math, a sophisticated curriculum reform that was rejected because it was not understood by the key stakeholders.

**Table 3.
Outcome Information Useful for Accountability**

Level	Key Questions	Useful Measures	Features
Student	Has the student learned prescribed content?	CRT mastery or gains in mastery	Valid, fair, understood, value-added
	Has the student's learning improved an appropriate amount?	Norm-referenced performance or gains in performance <i>Adjusted measures or targets based on prior performance and/or background factors</i>	
Teacher	Have students learned the prescribed knowledge and skills?	Distribution of student learning	Valid, fair, understood, value-added, robust
	Have students made adequate growth in knowledge and skills?	Distribution of gains in student learning <i>Adjusted measures or targets based on prior performance and/or background factors</i>	
School/ District	Have students learned the prescribed knowledge and skills?	Distribution of student learning	Valid, fair, understood, value-added, robust
	Have students made adequate growth in knowledge and skills?	Distribution of gains in student learning	
	Are teachers effective in raising student performance?	Distribution of teacher improvement measures	
		<i>Adjusted measures or targets based on prior performance and/or background factors</i>	