

Research Guidance to State Affiliates on Value-Added Teacher Evaluation Systems

Justin Baer

Michaela Gulemetova

Cynthia D. Prince

**National Education Association
Research Department
Ronald D. Henderson, Director**

Table of Contents

Introduction	4
How to Use this Guide	5
Accuracy and Reliability	7
1. Does the state/district or contractor propose to implement a value-added model, a growth model, or an attainment model?	7
2. What student, family, and school characteristics (e.g., race/ethnicity, free/reduced price lunch eligibility, English language learner status, etc.) does the state/district or contractor propose to include when generating teacher VAM scores?	8
3. What is the minimum number of years of student achievement data that the VAM system requires in order to produce teacher-level estimates?	9
4. What is the minimum number of students per teacher that the state/district or contractor requires in order for VAM scores to be generated for a teacher?	10
5. How does the contractor propose to deal with the basic technical problems of matching students to teachers to subjects taught?	11
6. How does the contractor propose to address missing student data issues?	12
7. How does the contractor propose to address the effects of high student absenteeism on teacher VAM scores?	12
8. Which test(s) will be used to calculate student achievement gains and estimate teacher performance? Are these tests aligned with state/district curricula?	14
Fairness	15
1. How will the results from the value-added analyses be reported to teachers?	15
2. Following implementation, how will the value-added teacher evaluation system be monitored to evaluate intended and unintended consequences for teachers, students, and schools?	16
3. How will teachers in untested grades and subject areas be evaluated as part of the VAM system?	17
4. How does the contractor propose to evaluate the contributions of education support professionals (ESPs), administrators, school nurses, counselor to student learning?	18
5. What steps has the state/district or contractor taken to make the calculation of VAM scores transparent?	19

6. What materials has the contractor produced to explain how the value-added model works and how teacher effectiveness is determined? Are materials available for both teachers and parents? 20

7. How will teacher value-added scores be used by the state/district? 21

Feasibility 22

1. How much time will the contractor require to analyze the data and produce value-added scores? 22

2. How knowledgeable is the contractor about value-added evaluation systems that have been designed and implemented in other states and school districts? 23

3. How much experience with VAMs does the contractor have? How much experience does the contractor have working with states and/or school districts?..... 24

Appendix A: Technical Checklist..... 25

Introduction

State affiliates are asking NEA for assistance in making some very important decisions regarding the implementation of value-added models (VAMs) as part of teacher evaluation systems. The introduction of VAMs has the potential to change how teachers are assessed, and, possibly, how they are compensated.

The purpose of this document is not to argue for or against value-added models, but rather to provide practical guidance to state and local affiliates on judging the merits of a VAM system proposed by their state or school district (or the contractor selected by the state or district).

Top 5 Things You Need to Know about Value-Added Models (VAMs)

1. Value-added measurement is a statistical technique for measuring growth in student learning that has occurred between two test administrations. It estimates the amount of growth, or gain, that an individual student would ordinarily be expected to make in one school year, based on that same student's performance in previous years. Teachers whose students make on average *greater than expected gains* would receive higher value-added scores than teachers whose students made *expected gains* or *less than expected gains*.

The word *estimates* is key. The scores generated from VAMs are statistical approximations of a teacher's contributions to student learning. They will always be estimated – not absolute – measures of teacher effectiveness.

2. There is more than one kind of value-added model. Different models make different assumptions and control for different sets of student characteristics and out-of-school factors that affect student achievement gains (such as prior achievement, race/ethnicity, poverty, parental education, language learner status, and special education status). These factors are controlled, or held constant, in order to isolate the contribution that a particular teacher makes to student learning gains – the “value” that he or she adds. Different value-added models may produce different estimates of teacher effectiveness.
3. Student achievement data can be used to produce value-added estimates at two levels:
 - a. school level; and
 - b. individual teacher or classroom level.
4. Value-added estimates cannot be produced for all teachers, such as those who teach non-tested subjects and grades or those who have been teaching for fewer than two years.
5. A value-added assessment system that proposes to evaluate teacher performance on the basis of only one year of student achievement data or the scores of small numbers of students is highly suspect. Increasing the size of the sample (by including more years of student test data and/or more students per teacher) will increase both stability and precision of the value-added estimates.

How to Use this Guide



We have adopted a “traffic sign” format as a user-friendly way of understanding the technical merits of value-added systems. However, in determining the **overall** fairness of value-added systems, a multitude of other policy, political, and practical factors must be considered. This guide cannot speak to these non-technical issues and consequently cannot be used to make determinations about whether an affiliate should support or oppose a value-added system.

The assumption underlying this guide is that state or local affiliates may be presented with proposals by state education offices or local districts to implement teacher evaluation systems that incorporate value-added models. VAMs vary both in their technical details (i.e., how scores for teachers are calculated) as well as how the scores are reported and used. This guide uses a question-and-answer format to help readers evaluate the design and technical merits of a proposed teacher evaluation system as they relate to the following areas:

- Accuracy and Reliability
- Fairness
- Feasibility

Within each area, we identify a series of questions for state and local affiliates to ask when reviewing VAM proposals submitted by a state, district, or contractor. Each question is followed by examples of potential approaches that a state or district (or a contractor selected by the state or district) may choose to take. The answers to the questions are presented using the following format:



- The answer proposed by the state/district or contractor is consistent with best practices for implementing VAMs;

OR

- In the absence of a single best practice, this answer is most likely to result in favorable treatment of teachers.



- This answer does not conform with best practices but still acknowledges the research on VAMs;

OR

- In the absence of a single best practice, this answer is less likely to result in favorable treatment of teachers.



- This answer ignores best practices and the research on VAMs;

OR

- In the absence of a single best practice, this answer will result in unfavorable treatment of teachers.

While many of the technical aspects related to VAMs are covered in this document, it does not include the full list of questions that may arise at the bargaining table. Please see Appendix A for a Technical Checklist that includes additional questions related to the design and implementation of value-added models.

Accuracy and Reliability

1. Does the state/district or contractor propose to implement a value-added model, a growth model, or an attainment model?

A state/district or contractor may choose from a variety of statistical approaches for calculating a teacher's effect on student learning. *Value-added models* make use of prior information about student achievement and student, classroom, and school characteristics. *Growth models* compare changes in test scores in the same students over time, but do not adjust for factors outside a teacher's control that can contribute to differences in student outcomes, such as poverty. *Attainment models* compare changes in the performance of different cohorts of students over time (e.g., the percentage of this year's 4th graders who scored at the advanced or proficient levels on the state test, compared to the percentage of last year's 4th graders who reached these targets). An attainment model measures the percentage of students who achieve a certain level of performance, or proficiency, at the end of the school year without accounting for differences in beginning levels of achievement or differences in student, classroom, and school characteristics.



The state/district or contractor proposes to implement a value-added model, which makes use of prior information about student achievement and student, classroom, and school characteristics. Of the three choices, value-added models are least likely to result in biased estimates of teacher effectiveness. **This does not mean that value-added models should be used for high-stakes decisions; to the contrary, they have significant limitations that make them inappropriate for such uses.** However, if faced with a decision between a value-added, growth model, or attainment model, a value-added model is preferred.



The state/district or contractor proposes to implement a growth model, which does not capture any information about student, classroom, and school characteristics.



The state/district or contractor proposes to implement an attainment model. These models are particularly problematic because they will likely underestimate teacher performance in schools that serve low-performing students.

2. What student, family, and school characteristics (e.g., race/ethnicity, free/reduced price lunch eligibility, English language learner status, etc.) does the state/district or contractor propose to include when generating teacher VAM scores?

When value-added models fail to account for preexisting differences between students, these differences will be incorrectly attributed to the teacher and included in his or her VAM score. To the extent possible, it is preferable to include background characteristics representing differences in student motivation, parental support at home, and peer influences when generating VAM scores because evidence suggests that these kinds of non-school factors are important influences on student achievement. However, few districts collect this type of information and instead have only basic measures such as gender, race/ethnicity, and free/reduced-price lunch status.

Districts may want to consider collecting additional information about out-of-school influences, such as the educational involvement of parents at home and resources available to students at home to assist their learning. These kinds of measures would likely allow for more precise estimates of what the teacher and school – aside from home and community – contribute to student learning. In any event, states and districts should include a range of student, family, and school characteristics in the value-added model.



The state/district or contractor proposes to include a range of individual student, family, class-, and school-level characteristics when calculating individual teacher VAM scores.



The state/district or contractor proposes a limited number of measures capturing the background characteristics of students and/or classes and schools.



The state/district or contractor does not intend to incorporate any background information about students, classes, and/or schools.

3. What is the minimum number of years of student achievement data that the VAM system requires in order to produce teacher-level estimates?

Value-added scores for teachers may vary from year to year for one of two reasons: 1) because a teacher’s true effectiveness in the classroom varies over time (e.g., a new teacher becomes more experienced in the classroom); or 2) because of the imprecision in calculating VAM estimates. The research on VAMs clearly indicates that it is not possible to get an exact measure of what a teacher contributes to student learning and that VAM scores are estimates that may change from year to year.

In order to allow for both variation in actual teacher effectiveness and the fact that VAMs, by their nature, are imprecise, the current best practice is to use a three-year rolling average when calculating a teacher’s VAM score for a particular year. That is, a VAM score for each teacher in a particular grade and subject area should be calculated for 3 consecutive years, and then averaged to generate an “effectiveness” score. This means that a VAM score for a teacher in a given year is, in fact, the average of his/her scores over the last 3 years in that grade and that subject.



The state/district or contractor proposes to use at least a 3-year rolling average to calculate VAM scores.



The state/district or contractor proposes to use a 2-year rolling average to calculate VAM scores.



The state/district or contractor proposes to use a single year of test score data to calculate VAM scores, resulting in possibly misleading conclusions about teacher effectiveness.

4. What is the minimum number of students per teacher that the state/district or contractor requires in order for VAM scores to be generated for a teacher?

Generally, the more students in a classroom, the greater the precision of the VAM score for a teacher. VAM scores are generated by averaging differences in student test score performance across all students in a class. Smaller class sizes will introduce more imprecision in isolating the relative contribution of a particular teacher.

Suppose, for example, in a five-student class that the grandparent of one student passed away shortly before the administration of the second test used to calculate VAM scores, adversely affecting the student's performance on the test. In a larger class (say 20 students), the effect of this student's poor test score – which was unrelated to the teacher – would be spread out across the average of 19 other students. While some of the teacher's VAM score will incorrectly be tied to the grieving student's poor performance in both the large and the small classes, the student's low score will carry a much higher weight in the small class, unfairly pulling down the teacher's evaluation.



While there is no consensus in the research literature, the following is a general rule of thumb: The state/district or contractor proposes a minimum class size of 15 students with test scores per teacher for each year in which VAM scores will be calculated. While 15 is the minimum, it is preferable for the threshold to be as high as possible (e.g., 20 is better).



The state/district or contractor proposes a minimum class size of 12 – 14 students with test scores per teacher for each year in which VAM scores will be calculated.



The state/district or contractor proposes to include classes with fewer than 12 students with test scores per teacher for each year in which VAM scores will be calculated.

5. How does the contractor propose to deal with the basic technical problems of matching students to teachers to subjects taught?

In order for VAM scores to be generated for teachers, a district must be able to link teachers, their students, and results from subject-specific student test scores. Moreover, students must be tracked over time (i.e., from grade to grade) and, ideally, from school to school within a district. If the links cannot be made, then differences in student performance on standardized tests cannot be attributed to particular teachers. The degree to which districts can uniquely link teachers, students, and tests depends in large part on the data systems they use, which should be considered before implementing VAM.



The district has data systems in place that uniquely link teachers, students, and subject-specific test results from year to year, and makes the data available to the contractor.



The district has data systems with incomplete linkages (e.g., information on student test scores is kept separately from class rosters). The district and/or contractor have a plan for integrating the systems and will verify the processes before generating VAM scores for teachers. If the district believes that there will be inaccuracies in the links between teachers, their students, and test scores (e.g., it is unclear whether Ms. Jones or Mr. Ramirez taught 6th grade math in Lincoln Elementary), then this uncertainty must be noted when calculating VAM scores. For example, teachers for which linkages cannot be made may be treated the same as teachers in untested grades or subject areas or teachers could be asked to verify their students using class rosters.



The district has limited or no ability to link teachers, students, and test scores, or the systems used to make the links are flawed. This will result in highly inaccurate VAM scores for teachers.

6. How does the contractor propose to address missing student data issues?

There are two kinds of missing data, recoverable and non-recoverable. Recoverable data are missing due to technical issues with the data systems maintained by a state or district. These data exist but may be missing from VAM analyses unless proper procedures are put in place to ensure their inclusion. Assuming the district and/or contractor implements appropriate data systems, these data should be recoverable. For example, if a student test score is missing because the student has changed schools, the use of unique student identifiers will allow districts to track students who move from school to school.

The second type of missing data is when actual data cannot be recovered, either because of faulty or incomplete data systems or because the data were never collected. For example, if a student did not take a test or if information about a student’s background characteristics (e.g., number of absences during a year, free or reduced-price lunch status) were not collected, these data are missing. Depending upon the type of missing data, certain statistical techniques can be employed to impute data, but the accuracy and reliability of these techniques is somewhat uncertain.

In the case of non-recoverable missing data, a contractor should assess the magnitude of the problem and whether steps can be taken to fill data gaps. If missing data are sporadic and minimal, dropping students who do not have complete sets of scores may be an acceptable alternative. However, if data gaps are systematic and affect large numbers of students, then dropping students from VAM calculations introduces bias because the sample of students is no longer representative of the actual class of students that the teacher taught.



The contractor carefully assesses both recoverable and non-recoverable missing data and presents a plan for handling both. In the case of recoverable data, the contractor proposes to modify data systems to ensure data are included in the VAM calculations. In the case of non-recoverable missing data, the contractor proposes a solution that will minimize bias into the calculation of teacher VAM scores.



The contractor assesses both recoverable and non-recoverable missing data, but the solutions for handling these data are less clear or may introduce bias. For example, a contractor may propose to simply drop all students with missing data from the calculation of teacher VAM scores and has no intention to assess possible biases that arise.



The contractor does not check whether missing data is an issue and/or does not have a plan for addressing it.

7. How does the contractor propose to address the effects of high student absenteeism on

teacher VAM scores?

A teacher cannot add value if students do not attend class. A contractor should propose reasonable procedures to determine how much time each student was present over the course of the year in order to “count” toward the teacher’s value-added score. A student who is absent for half of the school year will likely perform poorly on an end-of-year test, through no fault of his or her teachers. Unless the contractor takes student absences into account, a teacher’s VAM score will likely underestimate his or her actual performance.

Students who are highly mobile may also be counted as absent. Data systems should be designed to track students across schools within a district and to indicate when a student is no longer enrolled in any of the district’s schools.



The contractor checks whether student mobility and absenteeism are issues and has a plan to address them. For example, the contractor may collect daily student attendance records and use this information to adjust teacher VAM scores. Additionally, data systems employed by the district allow the contractor to track students across schools.



The contractor acknowledges the impact of student mobility and absenteeism on teachers’ value-added scores, but has incomplete or unclear plans to address these issues.



The contractor is unaware of the impact of student mobility and absenteeism on teachers’ value-added scores and has no plans to address these issues.

8. Which test(s) will be used to calculate student achievement gains and estimate teacher performance? Are these tests aligned with state/district curricula?

Value-added scores are based on standardized student achievement tests and attempt to measure the amount an individual teacher contributes to student learning during a school year. When tests and curricula are not aligned, value-added measures will provide inaccurate information about a teacher's performance and may even penalize teachers who follow the curriculum rather than teach to the test.

As an example, suppose that a state has adopted math standards that call for every student to have well-rounded, broad problem-solving skills. However, the state-mandated achievement tests in use are designed to measure very basic minimum skills. While the intent of the teacher evaluation system may be to assess teachers' ability to cultivate student problem-solving skills, the VAM scores based on these tests would instead reflect the teacher's ability to instruct basic minimum skills.



The state/district or contractor proposes to base teachers' value-added scores only on tests that have been determined to be in alignment with state/district curricula.



The state/district or contractor proposes to base teachers' value-added scores on tests that are not aligned with state/district curricula or the alignment has yet to be assessed.

Fairness

1. How will the results from the value-added analyses be reported to teachers?

The results from value-added analyses can generally be reported in two ways: *relative* to other teachers, or as an *absolute* score independent of other teachers. The most common *relative* way of reporting results is to use percentiles. In this case, a teacher receives a percentile score (e.g., “5th percentile” or “95th percentile”) that shows where the teacher’s VAM score ranks in comparison to other teachers. *Relative* scores are problematic because ranking teachers from highest to lowest masks teacher contributions to student learning. For example, in a district in which all teachers are high-performing, the bottom 5 percent of teachers may have still made significant contributions to student learning.

In contrast, *absolute* scores are not ranked relative to other teachers and allow for the possibility that all teachers in a district have reached a target level of effectiveness. An absolute score, for example, could be a score of 200 on a scale that runs from 100 to 250. To aid in interpretation, cut points are often assigned to absolute scales to indicate levels of performance. On a scale from 100 to 250, for example, 100 to 149 might be defined as “low-performing,” 150 to 199 is “satisfactory performance,” and 200 to 250 is “high-performing.”

All reports to teachers should include full, complete and understandable explanations that value-added scores reflect estimates not precise measures of teacher performance.



The district and/or contractor propose to report *absolute* VAM estimates using a scale with interpretable and fair cut points. The design of the reporting scale is transparent and includes input from a variety of stakeholders, including the union.



The district and/or contractor propose to report *relative* VAM estimates using percentile scores or some other format that forces teachers to be ranked from highest to lowest.



The district and/or contractor provide no information on how VAM estimates will be reported or have not considered the issue.

2. Following implementation, how will the value-added teacher evaluation system be monitored to evaluate intended and unintended consequences for teachers, students, and schools?

Teacher evaluation systems that incorporate VAMs depart from traditional evaluation systems and rely on a highly complex, technical methodology that is still evolving. Given the unique features of schools and districts and the many ways in which VAMs can be used, it is important for districts to monitor the *intended* and *unintended* consequences of their VAM-based evaluation systems.

In examining *intended* consequences, districts should assess the degree to which the VAM system is meeting the goals for which it was designed. For example, if one district goal was to increase teacher retention by allocating bonuses to “top”-performing teachers identified by VAM, the district should analyze whether the number of “top” performers who remain in the district has increased following implementation of the VAM system within a reasonable timeframe. Evaluation systems using VAM are typically linked to multiple policy goals, so it is important for the district to clearly articulate these goals and to have a plan for assessing whether they have been met.

Unintended consequences are more difficult to detect but equally important to assess. In San Diego, for example, the value-added model implemented by the district was unintentionally structured so that low-performing students were able to show greater gains than high-performing students.¹ This could have resulted in an incentive for teachers to move from high-performing schools to low-performing schools in order to achieve higher VAM scores. Regardless of whether the reassignment of teachers to low-performing schools is an explicit policy goal, it is critical for the district to understand how the VAM is operating and possibly influencing the decisions made by teachers.

¹ Koedel, C., and Betts, J. (2009). Value-added to what? How a ceiling in the testing instrument influences value-added estimation. Available: http://economics.missouri.edu/working-papers/2008/WP0807_koedel.pdf



The district clearly articulates the goals motivating the implementation of the VAM system, intends to conduct evaluation studies examining the intended and unintended consequences of the system, and intends to make mid-course corrections to the system on the basis of these findings.



The alignment between the district’s goals and the VAM system is unclear; the district has no plans to evaluate the intended or unintended consequences of the VAM system following implementation; or the district does not plan to make any mid-course corrections.

3. How will teachers in untested grades and subject areas be evaluated as part of the VAM system?

Individual teacher VAM scores can only be generated for teachers in grades and subjects with at least 2 years of standardized test scores (typically math and English/language arts). If individual VAM scores will count for a proportion of a teacher’s overall evaluation score, then the district or contractor must propose an alternative for teachers in untested grades and subject areas. One common practice is to include school-level VAM scores as part of the evaluation score for teachers without individual VAM estimates. That is, an average value-added score is calculated for all students in a school (still based only on those students in grades and subjects with standardized tests) and the school average is incorporated into a teacher’s evaluation score.

While this approach is frequently used, there is no research base to support it. Two problems with this strategy are that 1) it assumes that the effects of teachers of untested grades and subjects on student learning are, to some extent, captured in the combined VAM estimates of other teachers across a school, and 2) it holds teachers of untested grades and subjects accountable for changes in student learning for which they have no direct control.



The district/contractor has considered the issue of teachers of untested grades and subjects and proposes a fair and feasible solution. Be aware that a common practice is the use of school-level VAM scores as part of the evaluation of teachers of untested grades and subjects, though the reliability and validity of this approach has yet to be assessed.



The district/contractor has not considered the issue of teachers in untested grades and subjects and has no solution for how the evaluation scores for these teachers will be handled.

4. How does the contractor propose to evaluate the contributions of education support professionals (ESPs), administrators, school nurses, counselor to student learning?

Like teachers in untested grades and subjects, the direct individual contributions of ESPs and other non-teaching personnel to student learning cannot be captured using value-added methods. Consequently, a state/district or contractor will need to propose an alternative evaluation method. One alternative is to use a school-wide VAM average, but, as noted in the previous question, research has yet to demonstrate the validity and reliability of this approach.



The district/contractor has considered the issue of ESPs and other non-teaching personnel and proposes a fair feasible solution. Be aware that a common practice is the use of school-level VAM scores, though the reliability and validity of this approach has yet to be assessed.



The district/contractor has not considered the issue of ESPs and other non-teaching personnel and has no solution for how evaluations would be conducted.

5. What steps has the state/district or contractor taken to make the calculation of VAM scores transparent?

Value-added scores are generated from complex statistical methods and can vary significantly depending on the data used and the technical details of how the VAM is constructed. While making VAMs simpler may, on its face, seem appealing, simplicity often comes at the cost of accuracy and reliability. Rather than calling for the simplification of VAM scores, affiliates might consider negotiating with states/districts or contractors to make the procedures used to generate the scores transparent.

Contractors should provide detailed technical documentation about how the VAM was constructed, the rationale for the inclusion or exclusion of data, issues the contractor encountered when calculating the scores, and how these issues were resolved. Additionally, the contractor should include statistics showing the distribution of VAM scores across all teachers or subgroups of teachers (e.g., the mean and standard error). The technical documentation should allow other researchers to replicate the VAM scores, which may be especially important if a district intends to use the scores to make high-stakes decisions (e.g., dismissal).

Finally, the state/district or contractor should engage a third party external evaluator to assess the VAM system or establish a technical review panel that includes an external evaluator, a union representative, and other stakeholders.



The state/district or contractor proposes to develop detailed technical documentation that will allow the VAM scores to be replicated. The contractor can provide examples of the content of the technical documentation and agrees that this documentation will be released at the same time as the individual teacher VAM scores. The state/district or contractor will also allow for external evaluation.



The state/district or contractor proposes to develop technical documentation, but the content and timeline for release is unclear. The state/district or contractor will not commit to an external evaluation.



The state/district or contractor does not intend to release detailed technical documentation and rejects external evaluation.

6. What materials has the contractor produced to explain how the value-added model works and how teacher effectiveness is determined? Are materials available for both teachers and parents?

Value-added models are complex and evolving, and they diverge from the usual ways in which student learning is reported and teachers are evaluated. In order to make certain that key stakeholders – especially parents and teachers – understand VAMs, the state/district should produce accessible materials that explain how and why a VAM system is being implemented and the implications for students, parents, and teachers. Preparing these materials also serves as another step to make certain that the state/district has thought through the multiple issues involved with VAMs.



The state/district or contractor proposes to produce separate, user-friendly materials for teachers and parents describing the VAM system, how it departs from the previous evaluation system, how it will be implemented, and the implications for teachers and parents. The materials are prepared with input and feedback from key stakeholders, including the union.



The state/district or contractor intends to prepare materials for teachers (and possibly parents), but with limited or no input from key stakeholders, including the union.



The state/district or contractor has no plans to produce materials for teachers describing the VAM system.

7. How will teacher value-added scores be used by the state/district?

The individual teacher scores generated as part of a VAM analysis can be used in multiple ways. Before implementing a new evaluation system that incorporates VAM, it is important that the state/district clearly articulate the purpose behind using VAM and explain how it will inform the teacher evaluation (and/or compensation) process.

Clarity in the intended use of VAMs is especially important because VAMs only provide *estimates* of teacher effectiveness in the classroom. As noted in the introduction to this document, research shows that individual teacher VAM scores will vary from year to year for a variety of different reasons. Consequently, researchers advise against the use of VAMs for high stakes decisions – e.g., dismissals, promotions, compensation – that could be based on poor data or poorly constructed VAM systems, rather than a teacher’s true effectiveness.



Value-added scores will be used as part of formative assessments, to guide professional development, and to provide feedback to teachers. For example, a teacher will receive a value-added score as part of his or her yearly evaluation, which also includes other measures of effectiveness such as peer or principal observations.



Value-added scores will be used as one component in a teacher evaluation and/or compensation system with some high-stakes consequences. For example, value-added scores will account for a percentage of a teacher’s yearly evaluation score, which will be used to allocate bonuses.



Value-added scores will be used as the **sole** component in a teacher evaluation and/or compensation system with high-stakes consequences. For example, a teacher’s value-added score will be released publicly or it will be the sole factor used in determining bonuses.

Feasibility

1. How much time will the contractor require to analyze the data and produce value-added scores?

Capable contractors should be able to provide a reasonable estimate of the amount of time that this step will require, based on previous experience and the size and scope of the project. A critical implementation mistake is to fail to build in sufficient time for data verification and correction, especially during the first year of implementation. One of the chief lessons learned from early implementers of value-added teacher evaluation systems is that data errors can threaten the credibility of the entire evaluation system, particularly if they result in errors in bonus calculations or credit or discredit teachers on the basis of students they did not teach.¹

¹Milanowski, A., Witham, P., Schuermann, P., Kimball, S., & Pietryka, D. (2010). *Harvesting lessons on educator incentive plan design from technical assistance provided to Teacher Incentive Fund grants*. Working paper. Washington, DC: Center for Educator Compensation Reform, U.S. Department of Education. <http://www.cecr.ed.gov/pdfs/HarvestingLessons.pdf>



The contractor provides a reasonable estimate of the amount of time required to analyze and turn around the data, based on state/district testing schedules and the amount of data to be analyzed. The timeline allows sufficient time for data collection, verification, correction, analysis, and reporting; and meets state/district needs (e.g., allows sufficient time for the state/district to identify low performers for professional development or recruit high performers to work in priority schools).



The contractor's estimate is unrealistic, which suggests that the likelihood of errors could be high; the contractor and the state/district have no data verification and correction procedures planned or in place.

2. How knowledgeable is the contractor about value-added evaluation systems that have been designed and implemented in other states and school districts?

A contractor who is well-informed about the types and purposes of value-added methods in use elsewhere is likely to recommend designs that will meet state/district needs. A contractor who is aware of design and implementation challenges that other states/districts encountered will also be in a better position to help others avoid some of the same problems.



The contractor can identify numerous states and districts that are using value-added methods and how they are being used (e.g., for school accountability, teacher evaluation, teacher compensation). The contractor can identify the particular methods used in each instance, as well as strengths and limitations of the systems and how they differ.



The contractor has some knowledge about how value-added methods are being used at the state and local levels, but cannot identify some of the major methods in use or important distinctions among them.



The contractor has very limited knowledge of other states and districts that are using VAMs and the purposes for which they are being used.

3. How much experience with VAMs does the contractor have? How much experience does the contractor have working with states and/or school districts?

Because most states and districts are not likely to have the knowledge or capacity to design and implement a value-added teacher evaluation system by themselves, the knowledge, skills, and experience that external partners can contribute is of critical importance. Ideally, the contractor will have both strong technical capabilities as well as extensive experience working with schools and school systems. Beware of contractors who are knowledgeable about value-added methods but lack actual hands-on experience designing and implementing VAM systems for states and school districts:

“Having strong external partners with expertise in program design, data management, and developing and implementing performance metrics appears to contribute to getting a viable program up and running more quickly. Unless a state or district has substantial prior experience in... performance measurement, communications, and data system design, beginning with capable external partners lowers the odds of design and implementation problems.”²

²Milanowski, A., Witham, P., Schuermann, P., Kimball, S., & Pietryka, D. (2010). *Harvesting lessons on educator incentive plan design from technical assistance provided to Teacher Incentive Fund grants*. Working paper. Washington, DC: Center for Educator Compensation Reform, U.S. Department of Education. <http://www.cecr.ed.gov/pdfs/HarvestingLessons.pdf>



The contractor has extensive experience successfully developing and implementing value-added measurement systems in states and/or districts; has experience working in a labor-management environment; can describe data challenges, communication challenges, and political challenges that states and districts are likely to face, and can recommend strategies to avoid or minimize them.



The contractor may be quite knowledgeable about value-added measurement, but has had limited practical, hands-on experience designing, implementing, or scaling up a statewide or districtwide teacher evaluation system that uses value-added methods and has limited experience working in a labor-management environment.



The contractor has little or no experience working with states and/or school districts and in a labor-management environment; has no practical, hands-on experience designing, implementing, or scaling up a value-added teacher evaluation system for a state or school district.

Appendix A: Technical Checklist

	Yes	No	NA
Accuracy and Reliability			
1. Has the state/district clearly articulated the reasons for using a value-added model to assess teacher performance and explained how the information will be used (e.g., high-stakes vs. low-stakes decisions)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Has the state/district decided whether to produce value-added scores for individual teachers, groups of teachers, or both?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. If the plan is to assign value-added scores at the group level, has a decision been made about which groups to include (e.g., all teachers in a school, all math teachers in a school, all 4th-grade math teachers in a school)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Has a decision been made about which student, family, and school characteristics will be included in the value-added model (e.g., English language learner status, race/ethnicity, free/reduced price lunch eligibility)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Are the tests that are used to produce teachers' value-added scores aligned with state/district curricula?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Will at least 3 years of student test scores be required in order to calculate value-added scores for individual teachers?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Has a decision been made about how teacher performance will be assessed for those who have been teaching for less than two years?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Will test scores be required for class sizes of at least 15 students per teacher in order to calculate value-added scores for individual teachers?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Because value-added scores for teachers are based on the test scores of their students, they cannot be produced for teachers of untested subjects and grades. Value-added estimates may also be inadvisable measures of performance for teachers of English language learners and teachers of students with disabilities because traditional standardized tests may not be valid or reliable for these students. Have decisions been made about how the following groups of teachers will be assessed:			
a. Teachers of non-tested subjects (e.g., art, music, physical education, foreign languages)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b. Teachers of non-tested grades (e.g., pre-K to Grade 2, 5th-grade science)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c. Teachers of English language learners?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d. Teachers of students with disabilities?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Have decisions been made about how ESPs, administrators, school nurses, etc. will be assessed?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Have decisions been made about how the contractor will evaluate the contributions of education support professionals (ESPs), administrators, school nurses, counselor to student learning?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	Yes	No	NA
12. Mid-year testing substantially increases the difficulty of correctly attributing student learning gains to a specific teacher. If the plan is to use student achievement tests as part of a teacher evaluation system, are the tests administered close to the beginning and/or end of the school year?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. High rates of student mobility and absenteeism also increase the difficulty of determining how much an individual teacher contributes to student learning. Has a decision been made about how much time a student has to be present in order to “count” toward the teacher’s value-added score?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14. Have procedures been established to collect existing data such as student attendance (at the course level) with high frequency and accuracy?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15. Have procedures been established to recover missing student data, rather than simply dropping students with incomplete data from the sample? For example, does each student have a unique student identifier so that mobile students can be tracked across schools and districts?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16. Does the implementation timeline allow for reasonable delivery of data as well as rigorous quality control checks before assigning value-added scores?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Fairness

1. Will materials and training be provided so that teachers and principals know how to interpret value-added estimates of teacher performance?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Is the value-added model transparent? That is, will technical information about the value-added system be released for external review and validation, and will it explain what is included in the model and how the model works?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Will <i>absolute</i> value-added estimates be reported for teachers that use a scale with interpretable and fair cut points, as opposed to <i>relative</i> value-added estimates that use percentile scores or some other format which ranks teachers from highest to lowest?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Does the timeline for implementation include a pilot year to identify data gaps, correct data coding and data entry errors, try out data verification procedures, etc.?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Will lessons learned during the pilot year be used to make midcourse corrections before rolling out the entire value-added system?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Will an external evaluator be hired to monitor the value-added system for the purpose of assessing intended and unintended consequences for teachers, students, and schools?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Feasibility

1. Does the state/district have sufficient capacity in the data infrastructure, assessment, and testing systems to implement a value-added system?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Are data available at the appropriate level of specificity? For example, do existing data systems enable the state/district to link students to teachers and subjects taught so that the state/district can identify which teachers should be credited when student performance improves?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Does the system for linking students, teachers, and subjects work at both the elementary and secondary levels?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	Yes	No	NA
4. Has the state/district determined which persons and organizational units need information or data access and from where? Classroom value-added measures, for example, typically require teacher characteristics from the human resources system, student demographics from the student information system, and assessment results from the testing office.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Has the state/district planned or anticipated changes that might affect the teacher evaluation system or schedule, such as a new student information system, change of test date, or subjects covered?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. In the event that the state/district changes tests, is there a plan to assess their comparability and determine whether changes over time reflect real learning gains or differences in the tests themselves?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Assessing and maintaining data quality standards require financial resources (e.g., uncovering data quality errors and administering data quality checks at the school level). Does the state/district have a plan to provide the additional resources needed to implement and maintain data quality?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Has the state/district considered the financial costs of improving data quality and, if applicable, the possible addition of assessments in grades and subjects not covered under mandated state tests?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. If the state/district plans to hire a contractor to design and implement the value-added system, is the contractor knowledgeable about value-added systems in other states and school districts?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Does the proposed contractor have:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• strong technical skills?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• experience designing and implementing value-added models?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• experience working with states and/or school districts?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• experience working in a labor-management environment?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Although this checklist borrows some items from materials produced by the Center for Educator Compensation Reform, no endorsement of either the Center or its materials should be inferred.