



AMERICAN INSTITUTES FOR RESEARCH®

# Creating Summative Educator Effectiveness Scores

## Approaches to Combining Measures

MARCH 2012



# **Creating Summative Educator Effectiveness Scores: Approaches to Combining Measures**

**March 2012**

**Sheri Frost Leo and Lisa Lachlan-Haché**



AMERICAN INSTITUTES FOR RESEARCH®

1000 Thomas Jefferson Street NW  
Washington, DC 20007-3835  
202-403-5000 • TTY 877-334-3499  
[www.air.org](http://www.air.org)

This work was conducted and supported through AIR's *Performance Management Advantage* service—building evaluation systems that bridge the art and science of teaching.

Copyright © 2012 American Institutes for Research. All rights reserved.

1963\_3/2012

# Contents

	<b>Page</b>
Introduction.....	1
Types of Measures .....	1
Selecting Appropriate Measures .....	1
General Design Issues.....	2
Technical Properties.....	2
Local Policy and Stakeholder Context.....	2
Communication Considerations.....	2
Approaches to Combining Multiple Measures .....	3
Combining Measures Using the Numerical Approach .....	3
Advantages of the Numerical Approach.....	4
Drawbacks of the Numerical Approach.....	4
The Profile Approach.....	5
Illustration of the Profile Approach .....	5
Advantages of the Profile Approach.....	7
Drawbacks of the Profile Approach.....	8
The Holistic Approach.....	8
Advantages of the Holistic Approach .....	8
Drawbacks of the Holistic Approach.....	9
Summary of the Numerical, Profile, and Holistic Approaches .....	9
Real-World Application: A Hybrid Model .....	10
Conclusion .....	13
References and Additional Resources .....	14



## Introduction

Recognizing that human capital is the most important asset to teaching and learning, states and districts are rethinking their approaches to evaluating teachers' performance. The goal is to better differentiate effective and ineffective performance in order to inform a host of human capital decisions. In particular, the focus on teacher evaluation has spurred action by states and districts to improve the processes and tools for assessing teachers' performance, including the use of student achievement and growth as a significant criterion among multiple measures of teacher effectiveness.

Among the many decisions that states and districts make in the redesign of teacher evaluation systems, how to combine multiple measures of teacher performance into an accurate, consistent, and defensible summative teacher rating is one of the more challenging. This white paper is intended to assist states and districts in strategically combining measures into a summative score in a way that reflects their goals and priorities while accurately and consistently representing teacher effectiveness. While most evaluation systems typically draw upon multiple approaches to combining measures, this paper delineates three unique methods for combining measures: numerical, profile, and holistic and then introduces the more common hybrid approach. These distinctions aim to support readers in building a common language and in conceptualizing the variety of combinations that are possible. Beyond these basic approaches, it is further noted that states and districts will need to develop specific decision rules that accompany each approach while also taking into account the goals and priorities of the evaluation system.

### Types of Measures

In any performance appraisal system, evaluators have the option of examining two kinds of information on an individual: *behavior* and *results*. In education, this translates into two general categories of teacher evaluation metrics: "professional practices" measures and "student learning" measures. States and districts choose measurement tools for each metric they select as part of a teacher evaluation system. Classroom observations have been a primary tool for measuring professional practices, while recent efforts to include results-oriented or outcome metrics in teacher evaluation focus on student growth as a measure, with standardized tests an important measurement tool. As states and districts are beginning to think more innovatively about their standards and expectations for teachers, they are considering other tools to measure behavior and results. For example, student and family satisfaction is increasingly used as a metric in teacher evaluations, as evidence is emerging that student survey results are correlated to teacher effectiveness (Measures of Effective Teaching Policy Brief, 2011).

### Selecting Appropriate Measures

It is important that the metrics selected are predictive of or correlated to teacher effectiveness, and that the measurement tools accurately and consistently capture performance for a metric. In addition, evaluation design teams face practical considerations when selecting teacher effectiveness metrics, such as data availability, resources dedicated to implementation, and ease of communication with stakeholders.

Selecting appropriate measures requires states and districts to be explicit about the purposes and uses of teacher evaluation data as well as an overall aggregation approach. If a primary goal of the evaluation system is to help teachers improve practice, then evaluators require instruments that support them in providing teachers detailed feedback as well as a summative score. Further, if a primary goal of the system is to use data for human capital decisions, data gathered through these instruments must be reasonably accurate and consistent over time.

In selecting appropriate measures of teacher effectiveness, decision makers need to explore the following questions:

### **General Design Issues**

- Are the evaluation measures chosen aligned to state or district teaching standards? That is, does the evaluation aim to measure behaviors that are consistent with the expectations for practice laid out for teachers in state or local teaching standards?
- Are the measures chosen observable and/or measurable?
- Is there any evidence that a given evaluation measure is a correlate of teacher effectiveness? For example, is there any relationship between the chosen measure and student achievement?

### **Technical Properties**

- How accurately does each of the measurement tools (student assessments, growth models, observation tools, and so forth) assess an aspect of teacher effectiveness? What research or other information supports inclusion of the metrics and tools to be used in teacher evaluations?
- How much data are required, for consistency, to ensure the metric captures an accurate snapshot of performance? For example, how many years of student growth data, and what number of observations, observers, or survey responses are needed?
- How will the accuracy and consistency of each measure factor into decisions about its inclusion in summative ratings?

### **Local Policy and Stakeholder Context**

- What measures or evaluation tools and methods, if any, are required or proscribed by state statute, grant awards, and/or collective bargaining agreements?
- Where do important stakeholders (e.g., union leaders, legislators, principals, and teachers) stand on including measures in teacher evaluations?
- What actions will be informed by summative evaluation results? Are there any actions that will be informed by individual measures?

### **Communication Considerations**

- How easy is it to present the evaluation system to stakeholders, including teachers, evaluators, students, parents, and the general public?

- How clear is the statistical underpinning of the evaluation system? How will the evaluation system—expectations, benchmarks, and calculations—evolve over time and how will this message be communicated up front?

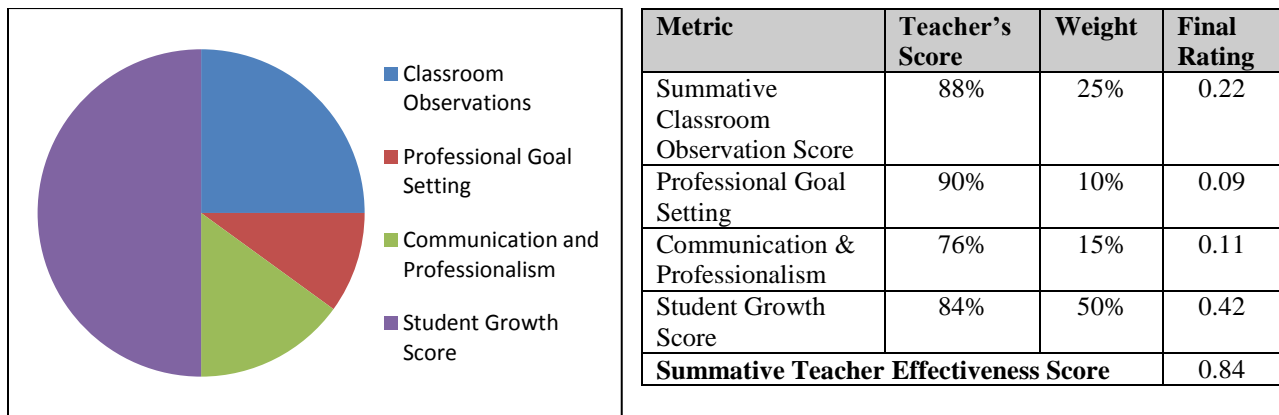
## Approaches to Combining Multiple Measures

As states build teacher evaluation systems that draw upon several different measures of teacher effectiveness, they need to consider how to combine many distinct data points into a single summative performance rating for each teacher. Multiple approaches for combining measures of teacher effectiveness have emerged: numerical, profile, and holistic. In practice, most teacher evaluation systems utilize a hybrid model that draws upon all three of these approaches for combining measures within and across categories such as professional practice, student learning, and stakeholder feedback.

### Combining Measures Using the Numerical Approach

In the **numerical** approach, the various measures of teacher performance are quantified and either added or averaged in order to generate a single teacher effectiveness “score.” The calculation may involve a straight average or a weighted average in which some measures are adjusted to contribute more than others before they are combined into a single score. Ranges of scores are then established to arrive at a summative teacher effectiveness rating. The weighting approach can be illustrated as a pie chart (see Figure 1 below) that depicts the weight assigned to each type of teacher performance data.

**Figure 1. Illustration of the Numerical Approach**



Does Not Meet Standards	Partially Meets Standards	Meets Standards	Exceeds Standards
0.0 – 0.19	0.20 – 0.54	0.55 – 0.89	0.90 – 1.0

In this example, a teacher receives a numerical score for each of four classroom observations. These scores are combined into a summative observation score, then measures of professional goal setting and communication are added and incorporated with multiple combined student learning data points for a student growth score. Each metric is then weighted according to the

state or district teacher evaluation guidelines. Next, these weighted metrics are added (or averaged) to create a single summative score for each teacher. Each of the summative teacher effectiveness ratings is associated with a range of scores, so that the summative score places a teacher into a rating category. The example above illustrates a teacher who has earned a 0.84 summative rating on a 1-point scale, or 84 percent out of 100. The score places the teacher into the category of “meets standards.”

### **Advantages of the Numerical Approach**

In psychometric literature, the numerical approach is known as the *compensatory* approach to combining measures for evaluation purposes. Advantages include the following:

- Teachers can compensate for weak performance in one area with stronger performance elsewhere.
- This approach is intuitive to many people, so it is relatively easy to describe to stakeholders.
- The weighting approach minimizes the effect of any biased metric or data outlier related to expected variations in performance over time. Thus, this approach is preferred by measurement experts.
- The specific measures in a system and/or their weight may change over time. This approach allows the system more flexibility to 1) change weights as well as specific components, 2) phase in specific components over time, or 3) differentiate the weights and components to better fit the unique contexts and skills teachers experience and develop (e.g., evaluations of new teachers may have more emphasis on observation and less on student growth).
- The numerical approach provides a single summative “score” for each teacher that translates well into human capital decisions about tenure, promotion, layoffs, and rewards. Districts can set score cut-points that vary by decision, making this approach more flexible to changing district needs.

### **Drawbacks of the Numerical Approach**

The weighting approach has several drawbacks as follows:

- Adding or averaging ratings may level out specific areas of strength and weakness to an average performance level, obscuring an unacceptably low level of performance in one area. If minimum performance standards are not set, a teacher with negative student growth could theoretically still be rated as effective if he or she receives sufficiently high ratings for teacher practice and professional responsibility.
- Through the numerical approach, some nuance can be lost as strong and weak ratings are averaged into a single score. This process may deny teachers meaningful feedback for improvement. Ideally, scores would be disaggregated to illustrate specific areas where teachers need refinement or reinforcement.



- Through the adding/averaging process, all measures are fit to the same scale. When more than one type of measure is involved, this mathematical process can require technical expertise beyond the reach of many districts. Further, a single scale for student growth scores, professional practice measures, and stakeholder feedback may be impractical or politically unpalatable. Finally, if multiple metrics capture the same aspects of teacher performance, there is a great risk that the actual weighting for each metric will be significantly different from the nominal weights. Measurement error can be amplified when metrics overlap, so metrics should be selected with care.

## The Profile Approach

In the **profile** approach, more than one type of evaluation data is collected, but each measure is considered and scored separately before the data are combined in a matrix with different categories of metrics on the horizontal and vertical axes. Rating categories are used for each of the measures, resulting in a multifaceted *profile* that defines areas of refinement or reinforcement for each teacher.

### Illustration of the Profile Approach

In the profile approach, all of the evidence for each group of metrics is considered (e.g., observations, goal setting, communication and professionalism, or student growth measures) in order to make a separate determination about each dimension of performance. Evaluators rate teachers on each dimension, providing multiple perspectives on teacher performance. Decision rules, often in the form of a matrix or an if-then flowchart, guide the evaluator to map performance in each dimension to a summative rating. First, as shown in Table 1 below, multiple observation ratings are combined for a summative observation rating.

**Table 1. Determining Summative Observation Rating**

		Performance Level				
		Unsatisfactory	Emerging	Proficient	Accomplished	Distinguished
	Observation 1		X			
	Observation 2			X		
	Observation 3			X		
	Observation 4				X	
<b>Summative Observation Rating</b>				<b>X</b>		

Next, as shown in Table 2, summative observation ratings are combined with other measures of teacher performance for a summative score of professional practice and responsibility.

**Table 2. Calculating Summative Score of Professional Practice and Responsibility**

	Unsatisfactory	Emerging	Proficient	Accomplished	Distinguished
Summative Observation Rating			X		
Goal Setting					X
Communication and Professionalism				X	
<b>Professional Practice and Responsibility Rating</b>			<b>X</b>		

Then, as displayed in Table 3, student growth ratings from district, state, and local measures are combined for a summative growth rating.

**Table 3. Determining Summative Growth Rating**

		1 (Low)	2	3	4 (High)
<b>Growth Score Measures</b>	District Growth Score				X
	State Growth Score			X	
	Student Learning Objectives			X	
<b>Summative Student Growth Rating</b>				<b>X</b>	

Finally, the two ratings, summative student growth and summative professional practice and responsibility, are combined in a matrix to determine the overall rating. In Table 4, the summative student growth rating is 3, and the professional practice and responsibility rating is “proficient,” leading to the teacher’s overall summative rating of “effective.”

**Table 4. Matrix of Overall Summative Rating**

		Summative Professional Practice and Responsibility Rating				
		Distinguished	Accomplished	Proficient	Emerging	Unsatisfactory
Summative Student Growth Rating	4	Highly effective	Highly effective	Effective	Effective	Minimally effective
	3	Highly effective	Effective	Effective	Minimally effective	Ineffective
	2	Effective	Effective	Minimally effective	Minimally effective	Ineffective
	1	Minimally effective	Minimally effective	Minimally effective	Ineffective	Ineffective

### Advantages of the Profile Approach

In psychometric literature, the profile approach is called the *disjunctive* approach; teachers must meet all of the minimum competencies on the evaluation metric in order to meet overall performance expectations. Advantages include the following:

- Through this approach, minimum performance expectations are set by a state or district for each of the components in the model, creating a common understanding of the performance thresholds in each area and ensuring that these thresholds are met by all teachers.
- Expectations can be differentiated to unique sets of teachers (based on teachers' experience, grade/subject, or school/district context).
- This approach is well suited to providing performance feedback to teachers during a summative rating conversation because specific strengths and weaknesses are not obscured by an averaging process.
- Unlike the numerical approach, the measurement scale does not need to be the same for every metric. Multiple types of metrics (e.g., scaled, qualitative, or binary) can be combined through the profile approach. For example, a student growth rating may be rated on a numeric scale of 1 to 4 and combined with a qualitative observation rating (e.g., unsatisfactory, developing, effective, accomplished) as well as a binary professionalism rating (e.g., does not meet expectations/meets expectations).
- The profile allows for the inclusion of both qualitative and narrative data in the summative rating process, which do not always translate well into numbers for the numerical approach. The resulting summative rating is associated with a multidimensional teacher profile that supports goal-oriented conversations about specific aspects of effectiveness during the summative evaluation conference.

## **Drawbacks of the Profile Approach**

Before implementing the profile approach, states should help districts consider the intended purposes of the system along with the communications, human capital, and measurement complexities of this approach. A teacher evaluation system based on the profile approach may be difficult to explain to stakeholders because there are multiple tiers of ratings, first within metrics and then across performance dimensions. This complexity also may make it difficult to sum up the formative results for school-level reports or principal evaluation purposes. The profile approach categorizes teachers by grouping them into similar categories, obscuring individual differences in performance within a category. Therefore, this approach can make the data more difficult to interpret for the purpose of human capital decisions around compensation, tenure, and promotion.

It is also critical to acknowledge that when errors or misrepresentations of a teacher's true performance are part of the body of evidence, the profile approach requires the evaluator to consider these data points with equal weight to the rest of the information. In such a case, the weighting approach has merit in that it generally evens out the errors in the data through averaging.

Consequently, a teacher must meet an overall set of standards that is more stringent under the profile approach than in the numerical approach. A final measurement-related drawback of the profile approach is the difficulty of explicitly weighting the various performance dimensions for a final summative score, as is required by several state laws.

## **The Holistic Approach**

The third approach to combining multiple measures, the **holistic** approach, emphasizes evaluators' professional judgment over a formulaic approach to combining measures into a single rating. An evaluator using the holistic approach reviews the body of collected evidence, looks for patterns in performance and trends over time, and compares the evidence to a performance rubric or similar set of multidimensional performance criteria. The evaluator interprets the evidence within the context of these performance benchmarks to draw a conclusion about overall performance and determine a teacher effectiveness rating.

### **Advantages of the Holistic Approach**

The holistic approach is the most flexible option for evaluators, accounting for factors such as a teacher's experience with the curriculum, additional leadership or teaching responsibilities, and other contextual variables, while emphasizing the big picture over any individual data point. In addition, this approach highlights the role of evaluator judgment in rating performance; the other two rating approaches may provide a false sense of numbers-based objectivity when, in reality, evaluators should and do play a pivotal role in any teacher evaluation system. Finally, the holistic approach lends itself well to focused implementation of the evaluation system and targeted feedback aligned to district priorities, a school-wide area of focus, or a teacher's individual goals. In contrast, the numerical and holistic approaches require more complete information on all of the teacher effectiveness measures outlined in the evaluation system.

## Drawbacks of the Holistic Approach

Drawbacks related to the holistic approach center on the strength of training and implementation for evaluators in applying this more nuanced strategy to teacher evaluation. It is difficult to ensure the consistent and accurate application of ratings across teachers, schools, and districts without intensive, ongoing training, calibration, and monitoring of evaluators across schools and districts. At the same time, it is more difficult to compare performance over time or across teachers when using the holistic approach since the evaluation parameters or areas of focus vary by teacher and year. In addition, if evaluators do not make a concerted effort to provide regular updates and feedback to teachers, a lack of transparency about teachers’ performance may result. A final shortcoming of the holistic approach is that it is not compatible with a weighting strategy for various performance measures, as many state or district policies require.

## Summary of the Numerical, Profile, and Holistic Approaches

None of the three approaches described above are mutually exclusive: Each approach can support feedback to teachers and be linked to administrative decisions, but each method has its strengths and limitations (as illustrated in Table 5 below). It is up to the state and district to choose the right application of each approach to support the stated aims of their evaluation system.

**Table 5. Strengths and Challenges of Both Approaches**

Approach	Examples	Advantages	Implementation Considerations/Challenges
Numerical	<ul style="list-style-type: none"> <li>• Achievement First</li> <li>• DC Public Schools</li> <li>• TAP™: <a href="#">The System for Teacher and Student Advancement</a></li> </ul>	<ul style="list-style-type: none"> <li>- Ease of communication</li> <li>- Use of results for human capital decisions</li> <li>- Application of weights easier, variations by teacher group are straightforward and possibly more easily understood by teachers and evaluators</li> <li>- Minimization of “noisy” data or rater problems</li> </ul>	<ul style="list-style-type: none"> <li>- Necessity of fitting all metrics to the same linear scale</li> <li>- Obscuring performance differences across dimensions</li> <li>- Lack of concrete, specific developmental feedback</li> </ul>
Profile	<ul style="list-style-type: none"> <li>• New Haven Public Schools</li> <li>• Rhode Island Department of Elementary and Secondary Education</li> </ul>	<ul style="list-style-type: none"> <li>- Creation of minimum performance expectations by metric</li> <li>- Map for developmental feedback</li> <li>- Incorporation of narrative evidence, metrics with different scales</li> </ul>	<ul style="list-style-type: none"> <li>- Application of results to human capital decisions more difficult</li> <li>- Magnification of measurement errors</li> <li>- Challenges in aggregating to school-level ratings</li> <li>- Complexity of weighting metrics</li> </ul>
Holistic	<ul style="list-style-type: none"> <li>• Massachusetts Department of</li> </ul>	<ul style="list-style-type: none"> <li>- Responsiveness to local context or individual needs</li> </ul>	<ul style="list-style-type: none"> <li>- Difficult to calibrate evaluators and compare</li> </ul>

	Elementary & Secondary Education • Ohio Department of Education	- Emphasis on trends and patterns over individual data points - Clarity around the role of professional judgment - Focused implementation	ratings across teachers or over time - Lack of transparency for teachers during the process - Not compatible with weighting multiple measures differently
--	--	--	--

## Real-World Application: A Hybrid Model

States and districts have the option to blend the strengths of each approach described above while mitigating some of the drawbacks of each by using a **hybrid** approach. In the real-world application of these approaches, nearly all states and districts use a hybrid approach to some degree. Even the examples highlighted in Table 5 above demonstrate aspects of a hybrid approach; they were categorized into the numerical, profile, or holistic approach based upon their most defining feature.

A thoughtful review of each metric will inform decisions about how and when to employ the numerical, profile, or holistic approach. For example, averaging multiple data points across evaluators or time periods allows a district to moderate the impact of bias from a single rater. This same district can then set minimum standards for the most critical aspects of effectiveness, thereby reducing the likelihood that teachers with serious weaknesses will be misclassified as effective.

Although building a hybrid model allows a district to tailor its strategy to the local context and needs, there are a few communication and implementation concerns related to this approach. The multiple levels of analysis and understanding required—when to analyze a body of evidence holistically, quantify and average data, or apply decision rules about minimum standards—make the hybrid approach more difficult to explain to stakeholders. Additionally, a hybrid approach mitigates but does not eliminate the drawbacks of the numerical, profile, and holistic approaches to combining teacher effectiveness measures. States and districts that choose this route should be mindful of the pitfalls associated with each of the approaches and work to proactively address such concerns (Massachusetts Teachers Association, 2011).

The Houston Independent School District (HISD) teacher evaluation model provides a more detailed example of a hybrid model that combines the numerical and profile approaches into a single evaluation system. HISD is in the process of redesigning its teacher evaluation system; the model was approved in May 2011 and implementation began in the 2011–12 school year.

### Multiple Measures

HISD measures teacher effectiveness from three perspectives: instructional practice (IP), professional expectations (PE), and student performance as follows:

- **Instructional Practice.** HISD has developed 13 IP criteria focused on instruction and planning as well as a detailed four-point rubric that describes performance at each level.

Teachers receive criterion-level ratings of 1–4 at the end of the year based on evidence gathered from classroom observations, conversations, and instructional artifacts such as lesson plans, portfolios, and student work samples.

- **Professional Expectations.** There are nine PE criteria described by a four-point rubric. Teachers will be assessed on each criterion using evidence gathered through observations, instructional artifacts, and conversations. They will receive a rating of 1–4 on each criterion at the end of the year.
- **Student Performance.** The student performance component may include a number of different measures of learning and growth, depending on the grade levels and subjects taught by a given teacher. Some of the potential measures include value-added growth on state standardized assessments, districtwide end-of-course or end-of-year assessments, and student performance on culminating performance tasks. The number and type of measures will vary by group of teachers, and this component is still under development.

### Adding Metrics for Component-Level Ratings

Final overall IP ratings are calculated by adding all 13 of the IP criterion-level ratings and translating this total to an overall rating as shown below in Table 6.

**Table 6. Proposed Instructional Practice Score Ranges**

Overall IP Rating	Total Criterion Scores
1	13–24
2	25–34
3	35–43
4	44–52

The same procedures are used to calculate the overall PE rating. Methods for calculating an overall student performance rating that utilizes multiple types of student growth data are still under review.

### Profile Aspects of the Model

The overall component ratings are combined with matrix look-up tables that group teachers into categories based on their performance on each component. First, the IP and PE ratings are combined to create a joint rating called IP x PE as displayed in Table 7.

**Table 7. Calculating Joint Rating of Instructional Practice/Professional Expectations**

		Instructional Practice			
		1	2	3	4
Professional Expectations	1	1	1	2	2
	2	1	2	3	3
	3	2	2	3	4
	4	2	2	4	4

Similarly, a student performance matrix, which is still under development, will help evaluators combine teacher value-added scores, where available, with other student learning measures.

The resulting IP x PE rating is combined with the overall student performance rating in a summative look-up table to determine a summative evaluation rating for each teacher (see Table 8 below).

**Table 8. Determining Summative Teacher Rating**

		Student Performance			
		1	2	3	4
IP x PE	1	Ineffective	Ineffective	Needs improvement	Needs improvement
	2	Ineffective	Needs improvement	Effective	Effective
	3	Needs improvement	Needs improvement	Effective	Highly effective
	4	Needs improvement	Effective	Effective	Highly effective



## **Conclusion**

Credible and accurate measures, local context, and stakeholder input are critical factors when crafting a summative teacher evaluation system that is fair, defensible, and transparent. Therefore, whichever approach a state or district adopts, it is important that decision makers understand and communicate their design choices while making thoughtful plans for system refinement.

It is important to note that even well-selected, technically sound measures do not eliminate the role of managerial judgment in evaluating performance. Any teacher evaluation system is only as good as the level of expertise with which it is administered. Well-trained, thoughtful evaluators will issue consistent evaluation information that is useful for teachers and district human capital managers. State and district leaders should consider the strengths and capacity of potential evaluators when identifying the purposes and uses of a new teacher evaluation system, and build the system to promote and develop these assets.

## References and Additional Resources

- Baker, J., & Best, A. (2011). *Effective teachers initiative: Designing a new teacher appraisal & development system*. Houston, TX: Houston Independent School District. Retrieved May 17, 2011, from [http://hisdeffectiveteachers.org/assets/HISD Appraisal and Dev System April 7 Board Workshop Presentation FINAL.pdf](http://hisdeffectiveteachers.org/assets/HISD_Appraisal_and_Dev_System_April_7_Board_Workshop_Presentation_FINAL.pdf)
- Center for Educator Compensation Reform. (2010). *TAP: The System for Teacher and Student Advancement*. Washington, DC: U.S. Department of Education, Office of Elementary and Secondary Education. Retrieved May 17, 2011, from <http://cecr.ed.gov/pdfs/summaries/TAPCaseSummary.pdf>
- Curtis, R. (2011). *Achievement First: Developing a teacher performance management system that recognizes excellence*. Washington, DC: The Aspen Institute. Retrieved May 13, 2011, from [http://www.aspeninstitute.org/sites/default/files/content/docs/education%20and%20society%20program/AI\\_Achievement%20First\\_performance%20mangmt.pdf](http://www.aspeninstitute.org/sites/default/files/content/docs/education%20and%20society%20program/AI_Achievement%20First_performance%20mangmt.pdf)
- Curtis, R. (2011). *District of Columbia Public Schools: Defining instructional expectations and aligning accountability and supports*. Washington, DC: The Aspen Institute. Retrieved May 13, 2011, from [http://www.aspeninstitute.org/sites/default/files/content/docs/education%20and%20society%20program/AI\\_DCPS\\_teacher%20evaluation.pdf](http://www.aspeninstitute.org/sites/default/files/content/docs/education%20and%20society%20program/AI_DCPS_teacher%20evaluation.pdf)
- District of Columbia Public Schools. (2010). *IMPACT: The District of Columbia Public Schools effectiveness assessment system for school-based personnel, 2010–2011. GROUP 1 general education teachers with individual value-added student achievement data*. . Washington, DC. Retrieved May 13, 2011, from <http://www.dc.gov/DCPS/Files/downloads/TEACHING%20&%20LEARNING/IMPACT/IMPACT%20Guidebooks%202010-2011/DCPS-IMPACT-Group1-Guidebook-August-2010.pdf>
- Goe, L. Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality.
- Houston Independent School District. (2011). Appraisal and development system proposal materials. *Strategic direction: Effective teachers initiative*. Retrieved May 25, 2011, from <http://hisdeffectiveteachers.org/final-appraisal-and-development-system-proposal>
- Houston Independent School District. (2011) *Building a better teacher appraisal and development system: Instructional practice in detail* (Draft Document). Houston, TX: Author. Retrieved May 17, 2011, from [http://hisdeffectiveteachers.org/assets/Appendix\\_2\\_Instructional\\_Practice\\_in\\_Detail\\_-\\_Criteria\\_and\\_Rating\\_Methodology.pdf](http://hisdeffectiveteachers.org/assets/Appendix_2_Instructional_Practice_in_Detail_-_Criteria_and_Rating_Methodology.pdf)

- Jerald, C., & Van Hook, K. (2011). *More than measurement: The TAP system's lessons learned for designing better teacher evaluation systems*. Santa Monica, CA: National Institute for Excellence in Teaching. Retrieved May 25, 2011, from [http://www.tapsystem.org/publications/eval\\_lessons.pdf](http://www.tapsystem.org/publications/eval_lessons.pdf)
- Massachusetts Teachers Association. (2011). *Reinventing educator evaluation: Connecting professional practice with student learning*. Retrieved November 29, 2011, from <http://www.massteacher.org/advocating/~media/Files/PDFs/CEPP/evalreport.pdf>
- Measures of Effective Teaching Project. (2010). *Learning about teaching: Initial findings from the Measures of Effective Teaching project* (Policy Brief). Retrieved November 29, 2011, from [http://www.metproject.org/downloads/Preliminary\\_Finding-Policy\\_Brief.pdf](http://www.metproject.org/downloads/Preliminary_Finding-Policy_Brief.pdf)
- New Haven Public Schools. (2010). *NHPS evaluation and development system recommendations: Executive summary for board*. Retrieved May 25, 2011, from [http://www.nhps.net/sites/default/files/NHPS\\_TEVALDEV\\_ExecutiveSummary\\_041510\\_Final.pdf](http://www.nhps.net/sites/default/files/NHPS_TEVALDEV_ExecutiveSummary_041510_Final.pdf)
- Rhode Island Department of Education. (2011). *The Rhode Island model educator evaluation system* (Working Draft). Retrieved May 17, 2011, from [http://www.ride.ri.gov/EducatorQuality/EducatorEvaluation/Docs/RhodeIslandModelEducatorEvaluationSystemWorkingDraft\\_01\\_7\\_11.pdf](http://www.ride.ri.gov/EducatorQuality/EducatorEvaluation/Docs/RhodeIslandModelEducatorEvaluationSystemWorkingDraft_01_7_11.pdf)
- Schmitt, L., & Ibanez, N. (n.d.) *AISD REACH program update: 2009–2010. Texas Assessment of Knowledge and Skills (TAKS) results and Student Learning Objectives (SLOs)*. Austin, TX: Austin Independent School District. Retrieved May 17, 2011, [http://archive.austinisd.org/inside/docs/ope\\_09-83\\_RB\\_Reach\\_TAKS\\_and\\_SLOs.pdf](http://archive.austinisd.org/inside/docs/ope_09-83_RB_Reach_TAKS_and_SLOs.pdf)
- Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. Brooklyn, NY: The New Teacher Project.



AMERICAN INSTITUTES FOR RESEARCH®

1000 Thomas Jefferson Street NW  
Washington, DC 20007-3835  
202.403.5000 | TTY: 877.334.3499

[www.air.org](http://www.air.org)

*Making Research Relevant*

1963\_3/12