AIR®

AMERICAN INSTITUTES FOR RESEARCH®

Education, Human Development, and the Workforce

Combining Multiple Measures in Performance-Based Compensation Systems

Lauren Bivona

SEPTEMBER 2012

Combining Multiple Measures in Performance-Based Compensation Systems

September 2012

Lauren Bivona

AMERICAN INSTITUTES FOR RESEARCH®
1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
202.403.5000 | TTY 877.334.3499
www.air.org

2815_09/12

Contents

Combining Multiple Measures in Performance-Based Compensation Systems

No single measure of educator practice or student outcomes can adequately capture the wide range of knowledge and skills an educator employs to demonstrate effectiveness. Thus, states and districts should consider incorporating multiple measures[1] of performance and outcomes into their performance-based compensation systems. Strategic use of multiple measures can increase the legitimacy of the system, promote stakeholder buy-in, and provide a more complete picture of teacher practice (Burnett, Cushing, & Bivona, 2012). In addition, use of multiple measures emphasizes the range of objectives the school or district values (Hamilton & Li, 2009).

Many educator evaluation systems already include multiple measures. States and districts may decide to incorporate the measures used in their evaluation systems into their performance-based compensation system, or they may want to consider incorporating additional measures that will align with local contexts and needs. In this paper, we provide a broad overview of the measures available to be used in a compensation system and then discuss different potential approaches to combining measures in a performance-based compensation system.

What are some measures that might be incorporated into a compensation system?

Education performance appraisal systems generally include one or both of two types of measures: measures of professional practice and measures of student learning (Leo & Lachlan-Haché, 2012). Classroom observations have been and continue to be a primary tool for measuring professional practice, but survey data, leadership projects, peer reviews, and artifact reviews are additional measures that may be used. Approaches to measuring student learning often include value-added modeling and student growth percentiles, both of which are calculations based upon student achievement data (Miller & Scott, 2012). Other measures of student learning include student portfolios, end-of-course assessments, school- or teacher-developed assessments, performance assessments, and Student Learning Objectives (SLOs). Each of these measures has its own strengths and weaknesses. For complete descriptions of measures, as well as their strengths and limitations, see Appendix A. Although the list does not contain all possibilities, it will provide a great starting point for states and districts.

It is important to ensure that metrics are predictive of or correlated to educator effectiveness and that the measurement tools selected accurately capture performance (Leo & Lachlan-Haché, 2012). Other practical considerations when selecting measures include data availability, the resources available to dedicate to implementation, and ease of communication with stakeholders. When used in human capital decisions, it is critical that data collected by measurement tools be reasonably accurate and consistent over time. In other words, "multiple measures should not mean 'anything goes'" (The New Teacher Project, 2010, p. 6). For a list of questions to consider when selecting measures, see Appendix B.

---

[1]Multiple measures refers to measures of student performance (student achievement scores, value-added scores, student learning objectives, performance portfolios), as well as measures of teacher practice (observations of instruction).

We have selected measures—now what?

Careful consideration of the combination of measures can signal to stakeholders the district's priorities for educator effectiveness and recognize the relative strengths of each measure selected. In this section, we profile multiple approaches to combining measures into a final performance rating.

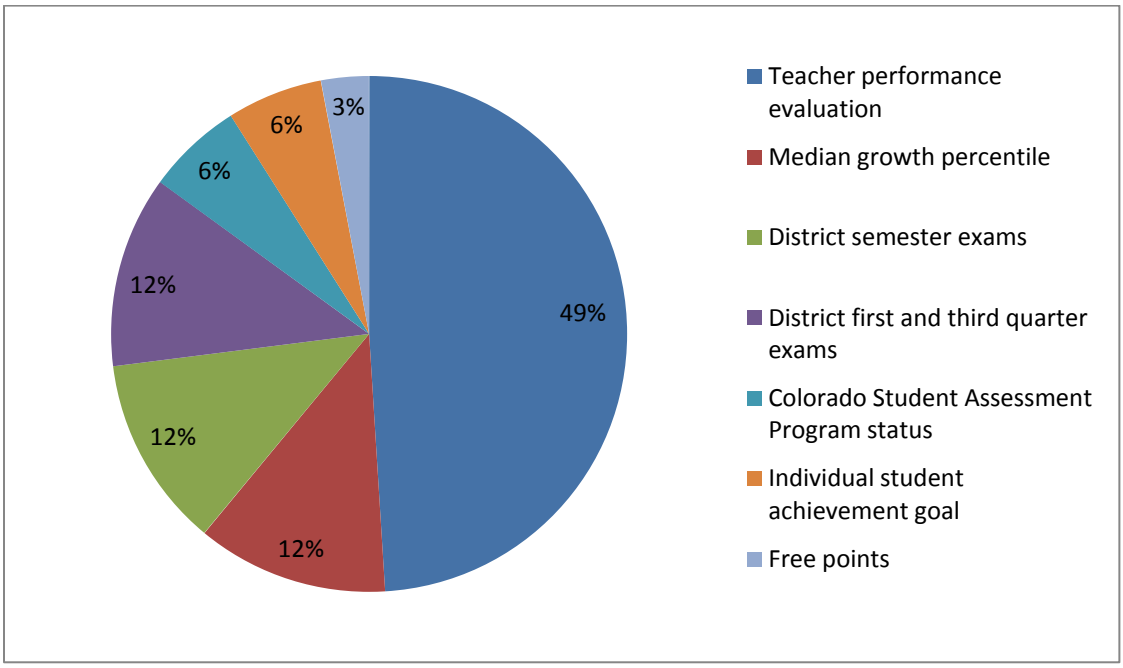
The Numerical Approach. The numerical approach requires districts to quantify measures of educator performance and then either add or average them to generate a final score. The calculation may be a straight average or a weighted average where some measures are given greater emphasis than others. The resultant summative score is then associated with a performance level. These performance levels may be connected to compensation, such as a career ladder.

Strengths of the Approach. The numerical approach is intuitive to most people and can easily be clearly communicated in pie charts and tables. In this approach, educators can compensate for poor performance in one area with stronger performance elsewhere; similarly, the numerical approach minimizes the effect of any biased metric or data outlier related to variations over time. This approach is also very adaptable because weights can change over time to permit the phase in of components over time or to differentiate the weights to better fit the unique contexts and skills educators need.

Limitations of the Approach. Although the numerical approach permits easy calculations, it may overlook unacceptably low performance in one area. Thus, educators may still receive rewards or additional compensation even though they may have negatively impacted student growth. This approach may also fail to provide educators meaningful feedback on their practice. To compensate for these limitations, districts may want to consider setting minimum performance standards for each measure and then report both aggregated and disaggregated scores to encourage conversation about educator performance and how educators might improve in the future.

Example. Harrison School District in Colorado combines multiple measures using a numerical approach to determine a teacher's evaluation score and subsequent compensation. The evaluation comprises a performance evaluation using a rubric (50 percent) and measures of student achievement (50 percent). The performance evaluation is worth 50 points, which are distributed across seven standards with the remaining point given to teachers as a "freebie." The student performance portion is also worth 50 points and is distributed across eight parts, with each part worth 6 points (Miles & Belcher, 2012). The remaining two points are given to teachers as "freebies." Six of the eight student performance parts vary based upon the grade, discipline, or specialty of the teacher, but the remaining two parts are the same for all teachers. One part must be tied to the school's standardized achievement scores and the other must be based on accomplishment of an individual goal set in collaboration with the principal. Figure 2 shows what the combination of measures might look like for a fourth-grade teacher.

Figure 2: Sample Combination of Measures for a Fourth-Grade Teacher in Harrison School District, CO



To standardize expectations, the district offers student achievement templates, which provides a list of acceptable measures for each part of a teacher's evaluation and the scoring rubrics to be used for each measure. Because the evaluation is worth 100 points, weighting by percentages is not needed, making scoring easier, as shown in Table 1.

Table 1: Sample Combination of Measures for a Fourth-Grade Teacher in Harrison School District, CO

Type of Measure	Data Source	Score
Performance (50 points)	Teacher performance evaluation (out of 50)	36
Achievement (50 points)	Median growth percentile (out of 12)	8
	District semester exams (out of 12)	8
	District first- and third-quarter exams (out of 12)	9
	Colorado State Assessment Program status (out of 6)	5
	Individual student achievement goal (out of 6)	4
	Free points	2
	Total Score (out of 100)	72

In the Harrison School district, the final evaluation score ties to a teacher's effectiveness rating and to a teacher effectiveness scale, a sequence of seven levels of effectiveness that ties the teacher evaluation score to a salary, as shown in Table 2. When the teacher's evaluation score falls into a higher score range than it did the previous year, the teacher moves to the next category on the teacher effectiveness scale and receives additional pay. However, no categories may be skipped. For example, if a Progressing I teacher scores 58 on the evaluation, the teacher

would only move to a Progressing II level. The teacher effectiveness scale is not a one-way scale. If the teacher scores at a lower level for three consecutive years, then the teacher moves down the teacher effectiveness scale to a lower level.[2] The final levels of the effectiveness scale exist for teachers who meet additional requirements beyond scoring high on the evaluation.

Table 2: Teacher Salary Scale in Harrison School District, CO

	Unsatisfactory	Progressing I	Progressing II	Proficient I	Proficient II	Proficient III	Exemplary I	Exemplary II	Master
Evaluation Score Range	10–18	19–29	30–42	43–57	58–71	72–85	86–100	*	**
Salary (in thousands of dollars)	35	38	40/44	48	54	60	70	80	90

* The teacher must have received an Exemplary II score from the Distinguished Teacher Evaluation Review team and score at least 42 points on the measures of student growth.

** The teacher must be rated Exemplary II for two consecutive years and either hold a National Board Certification or have successfully taught for five months in a school outside the district that is challenged by poverty.

For additional examples of using a numerical approach in performance-based compensation systems, see Appendix C.

The Profile Approach. In this approach, multiple measures are collected and the data are maintained in separate categories rather than added together. The measures are combined in a matrix, so evaluators can look up the "answer" to each possible combination of measures. The final product is a profile that defines strengths and areas in need of refinement for each educator.

Strengths of the Approach. States or districts set minimum performance expectations for each of the components in the model, so educators and stakeholders have a common understanding of acceptable and unacceptable performance for each measure. The approach also provides a visual representation of educator performance that highlights strengths and weaknesses across multiple dimensions. Unlike the numerical approach, the profile approach permits the combination of multiple types of metrics (scaled, qualitative, or binary) as well as different types of data (qualitative, narrative, and quantitative) without reducing all measures to a single number or "score."

Limitations of the Approach. This approach can be difficult to explain because calculations of a final score often require multiple steps, combining two measures at a time using a series of matrixes. The complexity of the profile approach also can make it difficult to sum up and weight results for school-level reports, compliance with state laws, or data reporting purposes. Even though the profile approach often provides more disaggregated data than a numerical approach, it places educators into performance categories but cannot rank order them within any rating

[2] In 2014, the span will be reduced to two years.

category. This lack of information can present challenges when using data to inform compensation.

Example. Although we currently do not have examples of a district or state using a profile approach to compensation, Rhode Island uses a profile approach in its teacher evaluation system and will soon offer districts grants to develop performance-based compensation systems. Using this approach, Rhode Island combines multiple measures, some of which are scored holistically and some of which are scored numerically but on separate scales. Teachers are evaluated on three criteria: professional practice (measured by teacher observations), professional foundations (measured by a Teacher Professional Foundations rubric), and student learning (measured by Student Learning Objectives (SLOs) and, beginning in 2013-14 in applicable grades and subjects, Rhode Island Growth Model ratings scores). Scores and ratings from each of the three criteria are combined to produce one of four final effectiveness ratings. This process consists of multiple steps.

Step 1. First, the evaluator scores each observation of the teacher separately. After all observation scores have been gathered, the evaluator computes an average score for each component to the nearest tenth. The average component scores are added together for a total Teacher Professional Practice Rubric score. The evaluator then determines the Professional Practice Rating based upon where the score falls in scoring ranges. The four possible ratings are Exemplary, Proficient, Emerging, and Unsatisfactory.

Step 2. The evaluator calculates a professional foundations rating using a rubric, then determines the Teacher Professional Foundations rating based upon where the professional foundations scores falls within scoring ranges. The three possible ratings are Exceeds Expectations, Meets Expectations, and Does Not Meet Expectations.

Step 3. Using a matrix, the evaluator determines the Professional Practice and Professional Foundations score on a scale of 1 to 4. The matrix is shown in Table 3.

Table 3: Matrix Used to Determine the Professional Practice and Professional Foundations Score in Rhode Island

		Professional Practice			
		Exemplary	Proficient	Emerging	Unsatisfactory
Professional Foundations	Exceeds Expectations	4	4	2	2
	Meets Expectations	4	3	2	1
	Does Not Meet Expectations	2	2	1	1

Source: Rhode Island Department of Education, 2012, p. 55

Step 4. The evaluator rates each individual SLO holistically. Possible ratings for individual SLOs are Exceeded, Met, Nearly Met, and Not Met. Using a series of matrices, the evaluator then determines an overall student learning objective rating. Possible overall ratings are Exceptional Attainment, Full Attainment, Partial Attainment, and Minimal Attainment. The matrix for the educator with three learning objectives is included in Table 4.

Table 4: Matrix Used to Determine the Final Student Learning Objective Rating in Rhode Island for Teachers with Three SLOs

Student Learning Objective 1	Student Learning Objective 2	Student Learning Objective 3	Final
Exceeded	Exceeded	Exceeded	Exceptional Attainment
Exceeded	Exceeded	Met	Exceptional Attainment
Exceeded	Exceeded	Not Met	Full Attainment
Exceeded	Exceeded	Not Met	Partial Attainment
Exceeded	Met	Met	Full Attainment
Exceeded	Met	Nearly Met	Full Attainment
Exceeded	Met	Not Met	Partial Attainment
Exceeded	Nearly Met	Nearly Met	Partial Attainment
Exceeded	Nearly Met	Not Met	Partial Attainment
Exceeded	Not Met	Not Met	Minimal Attainment
Met	Met	Met	Full Attainment
Met	Met	Nearly Met	Partial Attainment
Met	Met	Not Met	Partial Attainment
Met	Nearly Met	Nearly Met	Partial Attainment
Met	Nearly Met	Not Met	Partial Attainment
Met	Not Met	Not Met	Minimal Attainment
Nearly Met	Nearly Met	Nearly Met	Partial Attainment
Nearly Met	Nearly Met	Not Met	Partial Attainment
Nearly Met	Not Met	Not Met	Minimal Attainment
Not Met	Not Met	Not Met	Minimal Attainment

Source: Rhode Island Department of Education, 2012, p. 72.

Step 5. Beginning in 2013-2014, ratings based on value-added scores will be supplied to evaluators by the Rhode Island Department of Education.

Step 6. The evaluator then determines the overall Student Learning Score. In 2012-2013, SLOs are the only measures of student growth. In 2013-2014 and beyond, evaluators will use a matrix like the one shown below in Table 5 to determine the overall student learning score on a scale from 1 to 4.

Table 5: Matrix Used to Determine the Overall Student Learning Score in Rhode Island

		Student Learning Objectives			
		Exceptional Attainment	Full Attainment	Partial Attainment	Minimal Attainment
Growth Model	High Growth	4	4	3	2
	Typical Growth	4	3	2	1
	Low Growth	2	2	1	1

Source: Rhode Island Department of Education, 2012, p. 57

Step 7. Finally, the evaluator combines the Professional Practice and Professional Foundations Score from Step 3 and the overall Student Learning Score from Step 6 to determine the final effectiveness rating using a final effectiveness rating matrix, shown in Table 6. Possible ratings are Highly Effective, Effective, Developing, and Ineffective.

Table 6: Matrix Used to Determine the Final Effectiveness Rating in Rhode Island

		Student Learning Score			
		4	3	2	1
Professional Practice and Professional Foundations Score	4	Highly Effective	Effective	Developing	Developing
	3	Highly Effective	Effective	Developing	Developing
	2	Effective	Effective	Developing	Ineffective
	1	Developing	Developing	Ineffective	Ineffective

Source: Rhode Island Department of Education, 2012, p. 58

The Component Approach. Rather than combine measures into a summative score, some performance-based compensation systems will keep the measurements separate or combine similar measures into distinct components, and then reward educators for achievement of certain standards within each component. Often these components exist separate from the educator's evaluation.

Strengths of the Approach. The final results usually appear in a chart that explains an educator's performance by measure. This format does not require the measures to be converted to a common scale. Districts can emphasize their priorities by assigning greater payouts to some measures than others. Unlike in the numerical approach, minimum expectations are usually set for each measure. This approach is often used in performance-based incentives, especially when eligibility for certain components of payouts varies by educator.

Limitations of the Approach. Depending on how results are reported, the component approach may provide insufficient feedback to inform professional development plans. In addition, understanding the various measures that make up an award can be confusing, especially if not all educators are eligible for the same payouts. For example, in Prince George's County, Maryland, educators did not always clearly understand the components of the program for which they were eligible or the measurements that were used to assess their performance (Malen et al., 2011). When using this approach, clear communication of educator expectations and the compensation attached to each component is critical. Because this approach does not aggregate the measures to a final performance measure, it may be ill-suited for performance-based compensation that affects base pay and other benefits.

Example. Financial Incentive Rewards for Supervisors and Teachers (FIRST) is a TIF program that operated in Prince George's County, Maryland, from fall 2008 through spring 2012. The program provided monetary awards to administrators and teachers for their involvement or achievement in various categories: student performance, performance growth and contribution, evaluation, and hard-to-staff subject certifications (see Table 3). Each component had its own eligibility requirements and monetary value attached to it. For example, all teachers could earn

up to $2,500 if their school demonstrated growth over time by meeting adequate yearly progress (AYP) and district-set growth targets, but only teachers who taught mathematics or reading in self-contained classes were eligible for the classroom value-added award. When scoring items, each component had separate scoring systems and payout scales associated with the component.

Table 7: Award Components of Financial Incentive Rewards for Supervisors and Teachers (FIRST)

Component	Data Sources	Payout
School growth over time	AYP calculations Attainment of district-set growth targets	Up to $2,500
Classroom value added	Value-added model	Up to $2,500
Professional development	Attendance at required professional development or quality of individualized professional development as captured by a district-determined rubric	Up to $1,000
Leadership project	Quality of leadership project as measured by a district-created rubric	Up to $1,000
Observations using Danielson's *Framework for Teaching*	Growth between observations Attainment of Exemplary performance	Up to $1,000
Hard-to-staff certification	Holding a certification in a state- and district-identified hard-to-staff area	$1,500

For additional examples of using a component approach in performance-based compensation systems, see Appendix D.

Conclusion

The selection and weighting of measures in performance-based compensation systems are critical decisions in the design process. Selecting credible and accurate measures that align with district priorities, research on best practices, and input from stakeholders is a critical step in creating a fair and defensible compensation system. District systems may look very different in terms of the measures they include and the combination approach they employ. No one best approach exists, but, regardless of the decisions made, decision makers should understand the strengths and challenges associated with their design choices.

References

Achievement First. (n.d.). *AF Teacher Career Pathway executive summary*. Retrieved from http://tntp.org/assets/tools/AF%20Teacher%20Career%20Pathway%20Executive%20Summary%20Both%20Regions%20TSLT_0311.pdf

Bandeira de Mello, V. (2011). *Mapping state proficiency standards onto the NAEP scales: Variation and change in state standards for reading and mathematics, 2005–2009* (NCES 2011-458). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC: Government Printing Office.

Beaupré, D. (1995-2011). Testing our schools: A guide for parents. *PBS Frontline.* [Website]. Retrieved from http://www.pbs.org/wgbh/pages/ frontline/shows/schools/etc/guide.html

Braun, H. (2010). *Issues in measuring student growth and conducting productivity analyses*. Austin, TX: Educational Testing Service, Center for K–12 Assessment and Performance Management. Retrieved from http://www.k12center.org/rsc/pdf/BraunPresenterSession2.pdf

Buckley, K., & Marion, S. (2011). *A survey of approaches used to evaluate educators in non-tested grades and subjects.* Dover, NH: National Center for the Improvement of Educational Assessment. Retrieved from http://colegacy.org/news/wp-content/uploads/2011/10/Summary-of-Approaches-for-non-tested-grades_7-26-11.pdf

Burnett, A., Cushing, E., & Bivona, L. (2012). *Uses of multiple measures for performance-based compensation systems.* Washington, DC: Center for Educator Compensation Reform. Retrieved from http://cecr.ed.gov/pdfs/CECR_MultipleMeasures.pdf.

Community Training and Assistance Center (CTAC). (2008). *Tying earning to learning: The link between teacher compensation and student learning objectives.* Boston, MA: Author. Retrieved from http://www.ctacusa.com/PDFs/Rpt-TyingtoLearning-2008.pdf

Excellence for All. (2012). Rewards and recognition. http://www.pps.k12.pa.us/1483204413273860/blank/browse.asp?a=383&BMDRN=2000&BCOB=0&c=65084

Goe, L. (2008). *Using value-added models to identify and support highly effective teachers* (Key Issue). Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved February from http://www2.tqsource.org/strategies/het/UsingValueAddedModels.pdf

Goe, L. (2010). *Evaluating teaching with multiple measures.* Washington, DC: American Federation of Teachers. Retrieved from http://www.lauragoe.com/LauraGoe/EvalTchgWithMultipleMeasures.pdf

Goe, L., Bell, C., & Little, O. (June 2008). *Approaches to evaluating teacher effectiveness: A research synthesis.* Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from http://www.tqsource.org/publications/EvaluatingTeachEffectiveness.pdf

Goe, L., Holdheide, L., & Miller, T. (2011). *A practical guide to designing comprehensive teacher evaluation systems.* Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from http://www.tqsource.org/publications/practicalGuideEvalSystems.pdf

Goldhaber, D., & Hansen, M. (2010). *Is it just a bad class? Assessing the stability of measured teacher performance* (CEDR working Paper 2010–3). , Seattle: University of Washington, Center for Education Data and Research. Retrieved from http://www.cedr.us/papers/working/ CEDR%20WP%202010-3_Bad%20Class%20Stability%20(8-23-10).pdf

Goldstein, J. (2004). Making sense of distributed leadership: The case of peer assistance and review. *Educational Evaluation and Policy Analysis, 26*(2), 173–197.

Graham, M., Milanowski, A., & Miller, J. (2012*). Measuring and promoting inter-rater agreement of teacher and principal performance ratings*. Washington, DC: Center for Educator Compensation Reform. Retrieved from http://cecr.ed.gov/pdfs/Understanding_Basics.pdf.

Hamilton, L., & Li, J. (2009). *Designing effective pay-for-performance in K–12 education*. Santa Monica, CA: RAND Corporation.

Henrico County Public Schools. (2012). Learning Leaders Teacher Incentive Fund grant. Retrieved from http://www.henrico.k12.va.us/LearningLeaders/LearningLeadersIncentives.html.

Herndon, E. B. (1980). *Your child and testing*. Washington, DC: National Institute of Education. (ERIC Document Reproduction Service No. ED195579). Retrieved from http://www.eric.ed.gov/PDFS/ED195579.pdf

Hinchey, P. H. (2010). *Getting teacher assessment right: What policymakers can learn from research*. Boulder: University of Colorado, National Education Policy Center. Retrieved from http://nepc.colorado.edu/files/PB-TEval-Hinchey_0.pdf

Holdheide, L., Browder, D., Warren, S., Buzick, H., & Jones, N. (January 2012). *Summary of using student growth to evaluate educators of students with disabilities: Issues, challenges, and next steps.* Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from http://www.isbe.net/peac/pdf/using_student_growth_summary 0112.pdf

Houston Independent School District. (2012a). 2011-2012 ASPIRE Awards: School leader performance pay model. Retrieved from http://static.battelleforkids.org/images/HISD/LeaderAwardDiagram2012.pdf.

Houston Independent School District. (2012b). 2011-2012 ASPRIE Awards for teachers: Award model diagram. http://static.battelleforkids.org/images/HISD/hisd_ASP12Awards_Teach_4_30.pdf

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains.* Seattle, WA: Bill & Melinda Gates Foundation. Retrieved from http://www.metproject.org/downloads/MET_Gathering_Feedback_Practioner_Brief.pdf

Lavy, V. (2007). Using performance-based pay to improve the quality of teachers. *The Future of Children, 17*(1), 87–109.

Leo, S. F., & Lachlan-Haché, L. (2012). *Creating summative educator effectiveness scores: Approaches to combining measures.* Washington, DC: American Institutes for Research.

Little, O., Goe, L., & Bell, C. (2009). *A practical guide to evaluating teacher effectiveness.* Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from http://www.tqsource.org/ publications/practicalGuide.pdf

Malen, B., Rice, J. K., Jackson, C., Hoyer, J. M., Hyde, L., Bivona, L., et al. (2011). *Implementation, payouts, and perceived effects: A formative analysis of Financial Incentive Rewards for Supervisors and Teachers (FIRST).* Upper Marlboro, MD: Prince George's County Public Schools.

Marion, S., & Buckley, K. (2011). *Approaches and considerations for incorporating student performance results from 'non-tested' grades and subjects into educator effectiveness determinations.* Dover, NH: National Center for the Improvement of Educational Assessment. Retrieved from http://colegacy.org/news/wp-content/uploads/2011/10/Marion-Buckley_Considerations-for-non-tested-grades_2011.pdf

May, H., Perez-Johnson, I., Haimson, J., Sattar, S., & Gleason, P. (2009). *Using state tests in education experiments: A discussion of the issues* (NCEE 2009–013). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved from http://ies.ed.gov/ncee/pdf/2009013.pdf

Meyer, R. H., & Dokumaci, E. (2010). *Value-added models and the next generation of assessments*. Austin, TX: Educational Testing Service, Center for K–12 Assessment and Performance Management. Retrieved from http://www.k12center.org/rsc/pdf/MeyerDokumaciPresenterSession4.pdf

Milanowski, A. (2011). *Resolving some issues in using value-added measures of productivity for school and teacher incentives: Ideas from technical assistance and TIF grantee experience*. Washington, DC: U.S. Department of Education, Office of Elementary and Secondary Education, Center for Educator Compensation Reform. Retrieved from http://cecr.ed.gov/ pdfs/CECR_HP_ValueAdded.pdf

Miles, F. M., & Belcher, E. (2012). *Teacher compensation based on effectiveness: The Harrison (CO) School District's pay-for-performance plan*. Washington, DC: Thomas B. Fordham Institute.

Miller, J. & Scott, J. (2012). Understanding the Basics of Measuring Student Achievement. Washington, DC: Center for Educator Compensation Reform.

National Center on Response to Intervention (2011). *Progress monitoring tools* [Website]. Retrieved from http://www.rti4success.org/ progressMonitoringTools

Newton, X. A., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). Value-added modeling of teacher effectiveness: An exploration of stability across models and contexts. *Education Policy Analysis Archives, 18*(23). Retrieved from http://epaa.asu.edu/ojs/article/view/810/858

Peterson, K. D., Wahlquist, C., & Bone, K. (2000). Student surveys for school teacher evaluation. *Journal of Personnel Evaluation in Education, 14*(2), 135–153.

Peterson, P. E., & Hess, F. M. (2008). Few states set world-class standards. *Education Next.* Retrieved from http://educationnext.org/files/ ednext_20083_70.pdf

Prince, C. D., Schuermann, P. J., Guthrie, J. W., Witham, P. J., Milanowski, A. T., & Thorn, C. A. (2009). *The other 69 percent: Fairly rewarding the performance of teachers of nontested subjects and grades*. Washington, DC: U.S. Department of Education, Office of Elementary and Secondary Education, Center for Educator Compensation Reform. Retrieved from http://cecr.ed.gov/guides/other69Percent.pdf

Rhode Island Department of Education.(n.d.). Rhode Island model: teacher evaluation and support system. Edition II. Retrieved from http://www.ride.ri.gov/EducatorQuality/EducatorEvaluation/Docs/Teacher_Model_GB-Edition_II_FINAL.pdf.

Rivkin, S. G. (2007). *Value-added analysis and education policy* (Brief 1). Washington, DC: Urban Institute, National Center for Analysis of Longitudinal Data in Education Research (CALDER). Retrieved from http://urbaninstitute.org/UploadedPDF/411577_value-added_analysis.pdf

Sartain, L., Stoelinga, S. R., & Brown, E. R. (2011). *Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation.* Chicago: Consortium on Chicago School Research: University of Chicago Urban Education Institute. Retrieved from http://ccsr.uchicago.edu/publications/ Teacher%20Eval%20Report%20FINAL.pdf

Steele, J. L., Hamilton, L. S., & Stecher, B. M. (2010). *Incorporating student performance measures into teacher evaluation systems.* Santa Monica, CA: RAND. Retrieved from http://www.rand.org/pubs/ technical_reports/TR917.html

The New Teacher Project. (2010). *Teacher evaluation 2.0.* Retrieved from http://tntp.org/assets/documents/Teacher-Evaluation-Oct10F.pdf?files/Teacher-Evaluation-Oct10F.pdf

Toch, T. (2008). Fixing teacher evaluation: Evaluations pay large dividends when they improve teaching practices. *Education Leadership, 66*(2), 32–37.

Tucker, P. D., Stronger, J. H., Gareis, C. R., & Beers, C. S. (2003). The efficacy of portfolios for teacher evaluation and professional development: Do they make a difference? *Educational Administration Quarterly, 39*(5), 572–602.

Tyler, J. H., Taylor, E. S., Kane, T. J., & Wooten, A. L. (2010). Using student performance data to identify effective classroom practices. *American Economic Review: Papers and Proceedings, 100,* 256–260.

U.S. Department of Education. (2004). *Testing: Frequently asked questions* [Website]. Retrieved from http://www2.ed.gov/nclb/accountability/ ayp/testing-faq.html

Valli, L., Croninger, R. G., & Walters, K. (2007). Who (else) is the teacher? Cautionary notes on teacher accountability systems. *American Journal of Education, 113,* 635–662.

Weems, D. M. & Rodgers, C. B. H. (2010). Are U.S. teachers making the grade? A proposed framework for teacher evaluation and professional growth. *Management in Education, 24*(1), 19–24.

Wilkerson, D. J., Manatt, R. P., Rogers, M. A., & Maughan, R. (2000). Validation of student, principal, and self-ratings in 360° feedback™ for teacher evaluation. *Journal of Personnel and Evaluation, 14*(2), 179–192.

Appendix A. Measures for Teacher Evaluation and Performance-Based Compensation[3]

Measure	Description	Strengths	Limitations
Student Assessment Measures			
State Achievement Tests	No Child Left Behind (NCLB) requires states to assess students in Grades 3 through 8 and high school (U.S. Department of Education, 2004). One measure of teacher performance is student performance on state assessments (Miller & Scott, 2012).	State achievement tests provide a common measure of student performance across a state and allow for comparisons across teachers and schools within and across districts. These assessments often tie to state and district standards, and the test content reflects state priorities. Further, many state achievement tests have high levels of validity (they measure what they are designed to measure) and reliability (they produce consistent and accurate results). Finally, because states already utilize these tests for NCLB reporting, using state achievement tests to assess teacher performance incurs few additional costs (Goe et al., 2008).	Although student achievement tests reflect state standards, they do not capture teachers' efforts to expand student knowledge beyond what is required for grade-level proficiency. Subsequently, test scores may not capture true performance of low- or high-performing students (Little, Goe, & Bell, 2009; May et al., 2009). State achievement tests usually consist of 40 to 50 multiple choice questions; consequently, they often contain only one or two test items per assessed objective and do not assess all objectives (May et al., 2009). State achievement tests also may not adequately measure higher level thinking skills, such as analysis and evaluation (Toch, 2008). Because states do not typically test all subjects, the use of student achievement tests is limited to select subjects and grades. In addition, studies comparing the rigor of state proficiency standards against nationally normed[4] standards, such as the National Assessment of Educational Progress (NAEP), suggest that standards and assessments vary widely across states (Bandeira de Mello, 2011; Peterson & Hess, 2008). Thus, student achievement tests are not appropriate for interstate comparisons. Student performance on state assessment tests can help teachers identify the strengths and weaknesses of their performance, but do not directly evaluate instructional quality (Goe, 2010).
Standardized Achievement Tests	Standardized achievement tests are nationally normed and measure student knowledge and skills in	Test makers typically design these tests to have high validity and reliability, and many have been evaluated in research studies (National Center on Response to	Standardized achievement tests are expensive to purchase, and administering these exams for every grade and subject can be an expense too large for most states and

[3]This chart originally appeared in Burnett, Cushing, and Bivona, 2012.
[4]A nationally normed test uses a common standard for the entire nation.

Measure	Description	Strengths	Limitations
	relation to other students (Beaupré, 1995–2011; Herndon, 1980). Examples include the Terra Nova tests, Stanford Achievement Tests, Iowa Tests of Basic Skills, NAEP exams, and Advanced Placement (AP) exams (Beaupré, 1995–2011).	Intervention, 2011). Because these tests are nationally normed, evaluators can compare test results across classrooms, districts, and states (Herndon, 1980). Unlike state achievement tests, which tend to be high-stakes assessments, many standardized achievement tests are considered low-stakes exams.	districts (Buckley & Marion, 2011). Further, because tests are not typically created for particular states or districts, they need to be aligned to the specific curriculum standards of the district, state, or school (Buckley & Marion, 2011).
End-of-Course Assessments	End-of-course assessments are summative assessments of student learning. States or districts often mandate them, and they correspond to the content of the course (Buckley & Marion, 2011).	End-of-course assessments serve as a way to measure student achievement in otherwise untested subjects and grades (Prince et al., 2009). Because states and districts frequently develop end-of-course assessments, they can often use the results to compare educators across schools within a state or district (Buckley & Marion, 2011). Locally created end-of-course assessments can be customized to fit local standards and can align with local curricula more easily than standardized or state tests (Prince et al., 2009).	Few states and districts have the financial resources to create tests in every subject and validate them sufficiently for high-stakes accountability (Buckley & Marion, 2011). Further, when used alone, end-of-course assessments cannot measure growth (Prince et al., 2009).
School- or Teacher-Developed Assessments	A school, teacher, or group of teachers develop these tests to assess student knowledge or mastery. Examples could include an end-of-chapter test, a midterm, student portfolios, or performance tasks (Buckley & Marion, 2011).	These tests are relatively inexpensive to create and administer. They serve as a way to measure student achievement of typically untested subjects and grades. Because teachers typically develop tests for their own students, they have the flexibility to tailor the tests to specific goals and may be more likely to support the use of student assessment for evaluation purposes (Buckley & Marion, 2011). If assessments are developed by a group of teachers teaching the same subject and content, teachers can compare results across classes. Group discussions of student performance may encourage teachers to share best practices or develop collective plans for improving student achievement. Finally, the test development process itself can improve teachers' practices through the exercise of designing and improving their assessments (Community Training and Assistance Center, 2008).	Because individual schools or teachers develop the tests, they do not permit comparison beyond school walls. The rigor and quality of the tests are also questionable because the testing environment is unlikely to be as controlled as with standardized tests (CTAC, 2008). A teacher may use the same questions for all classes year after year, which could enable students to decrease the test's validity by circulating questions (CTAC, 2008). Also, individual schools and teachers are unlikely to have sufficient resources to develop assessments with the validity needed for high-stakes accountability (Buckley & Marion, 2011). Finally, unless teachers administer well-aligned pretests, teacher-developed tests cannot measure growth.

Measure	Description	Strengths	Limitations
Teacher Observation Measures			
Classroom Observation by Principal or Outside Evaluator	A principal or outside evaluator evaluates a lesson, often using a protocol or rubric, either during an informal walk-through or a formal session (Hinchey, 2010).	A variety of stakeholders, including teachers, principals, and community members, generally consider classroom observations to be an effective measure of teacher quality (Little et al., 2009). When the observation process includes use of valid observation forms and rigorous training, observers can provide detailed information about teacher practices that can be useful for both formative and summative purposes (Goe et al., 2008). Observations can also capture information that is not included in student achievement scores, including student engagement, classroom environment, and teacher questioning techniques—all of which are important for student learning.	Valid observation forms, rigorous training and recalibration, and sufficient observation time are necessary in order for observations to be valid and reliable for high-stakes accountability (Graham, Milanowski, & Miller, 2012). Properly training evaluators and ensuring agreement between observers can be costly in time and financial resources (Graham et al., 2012). Furthermore, the time required to observe teachers, document evidence, and conduct pre- or post-observation conferences may strain the capacity of already busy school administrators (Malen et al., 2011; Sartain et al., 2011). Observations may not be able to provide teachers with formative feedback if observers are not experts in the subjects of the teachers they observe.
Peer Review	Peer and/or master teachers conduct classroom observations. These peer teachers can be from the observed teacher's school or another school. Peer observers usually specialize in the same content area as the teachers they observe (Goldstein, 2004).	Administrators, teachers, and peer observers can benefit from the peer review process. By having expert teachers observe their peers, administrators are able to reduce their administrative burden (Goldstein, 2004). Teacher observers with subject area expertise may be able to provide peers with more detailed feedback on the lesson content than administrators (Goldstein, 2004). In return, peer observers may benefit through exposure to a variety of instructional strategies, some of which may be new to the reviewer (Weems & Rodgers, 2010). Compared to the cost of hiring external evaluators, it is generally less expensive to use existing staff. Additionally, research suggests that peer review models can be effective at pushing ineffective teachers out of the classroom (Goldstein, 2004).	Peer review, like other classroom observation methods, presents similar challenges. Successful peer review processes require valid observation forms, rigorous training, and sufficient time to observe and meet with teachers (Goe et al., 2008). Providing time to master teachers or expert teachers to observe and meet with teachers can be a costly logistical challenge. Although the peer review process can be a useful exercise for everyone involved, it can also create tensions between the peer observer and the observed teacher, especially if the observer's feedback is negative.
Other Approaches to Teacher Evaluation			
Instructional Artifacts	Using standardized protocols, raters select and evaluate specific	Evaluators can use artifacts to assess teachers of all subjects and grade levels. Artifacts supplement other	Although artifact collection requires teachers to collect only evidence of work they already do, evaluators

Measure	Description	Strengths	Limitations
	artifacts of teachers' work, such as letters to parents, open-house handouts, student assessments, grading guidelines, lesson plans, or student work (Goe, Holdheide, & Miller, 2011).	measures well because they assess areas of teachers' practice that may not be evident in achievement data or classroom observations. For example, artifacts can assess teachers' professionalism, instructional preparation, communication with parents, and engagement with the community. Another advantage of artifacts is that they require little additional effort on behalf of teachers because artifacts are preexisting student and teacher work products. When rubrics and protocols are valid and raters receive sufficient training, artifacts can serve as a useful indicator of instructional quality (Goe, 2011).	need sufficient training to ensure that they score artifacts consistently and provide teachers with appropriate guidance (Goe et al., 2008). The review process is time intensive, requiring review of multiple materials and provision of feedback so that teachers can improve their practice. Studies suggest that the subjectivity of such assessments renders instructional artifact analysis insufficiently reliable for high-stakes accountability (Steele, Hamilton, & Strecher, 2010). Additionally, raters need sufficient expertise in all subject areas (Little et al., 2009). With the exception of student work, most artifacts do not demonstrate student performance or growth.
Portfolios	A portfolio is a collection of artifacts that can include teacher videos, lesson plans, rationales for teaching, and teaching philosophies. Often, the portfolio process requires teachers to reflect on their practice (Tucker, Stronger, Gareis, & Beers, 2003).	Portfolio creation is a reflective process that requires teachers to think about their effectiveness and provide evidence of their practice (Danielson, 2007; Weems & Rodgers, 2010). A well-compiled portfolio provides a comprehensive overview of a teacher's practice, including aspects that achievement data or classroom observations may not demonstrate. Portfolios enable teachers to demonstrate excellence in a variety of ways, rather than using measures that focus on one aspect of teaching (Weems & Rodgers, 2010), and give teachers control over the basis of their evaluation (Goe et al., 2008; Tucker et al., 2003). Additionally, districts can use portfolios to evaluate teachers of all subjects and grade levels.	Portfolios tend to focus on teachers rather than students. Unless student work or test scores are included, portfolios cannot demonstrate student achievement or growth. The preparation and review of portfolios can be very time-consuming. Portfolios require teachers to devote a significant amount of nonteaching time to compile materials, organize them appropriately, and write rationales or reflections on their practice. Portfolios can also be unwieldy documents that are difficult and time-consuming to review; often, rater training is needed to increase reliability (Tucker et al., 2003). Teachers tend to select their most exemplary work for portfolios, thereby giving raters an unrepresentative impression of their practice (Goe et al., 2011). The process may also unfairly favor teachers who are more articulate or are skilled writers. As with observations and artifacts, judging portfolio quality and teacher practice is challenging (Weems &

Measure	Description	Strengths	Limitations
			Rodgers, 2010). Developing specific rubrics and training raters is an expensive but necessary step to ensure consistent scoring. It may also be necessary to find raters with expertise in each teacher's subject area (Little et al., 2009).
Student and Parent Surveys	Students and/or parents complete surveys about teachers' behaviors. Topics often include teachers' efforts to engage and challenge students (Goe et al., 2011).	Surveys provide the perspective of those most affected by teachers: their students (Goe et al., 2011). Surveys can gauge intangible aspects of teacher performance, such as perceived teacher expectations and rapport with students. These measures can provide teachers with specific feedback on how they can improve their interactions with students and parents. Because they can be administered across an entire school or district at once, surveys are often cost- and time-efficient (Goe et al., 2011). Systemwide distribution enables evaluators to compare survey results across classrooms within a school or district. A recent study from the Measures of Effective Teaching (MET) project found that the relationship between observation scores and student growth grew stronger when researchers combined observation results with student surveys (Kane & Staiger, 2012). Other studies have found that high-quality survey instruments can be valid measures of teacher performance and predictors of student achievement (Peterson et al., 2000; Wilkerson, Manatt, Rogers, & Maughan, 2000).	As with most surveys, achieving high response rates can be challenging. Schools may struggle to get parents to return the surveys. Additionally, the survey process may prevent illiterate or non-English-speaking parents from completing the survey. Although surveys can provide valuable insight into aspects of teacher practice that other measures do not capture, some argue that students and parents may not be knowledgeable about the complexity of teaching (Goe et al., 2008). Peterson, Wahlquist, and Bone (2000) caution that having high survey ratings does not necessarily equate with being a good teacher. Furthermore, collecting feedback from young students may be difficult. Districts should also consider that students will likely find out when surveys affect teacher compensation or employment status. Some students may see such a survey as an opportunity to punish a teacher they dislike. Teachers may also attempt to manipulate the survey results, even if they are not present during survey administration.
Self-Report Measures	These measures can take a variety of forms. They require teachers to reflect upon and document their practice, using surveys, instructional logs, or interviews (Goe et al., 2011).	Self-report measures provide information about teacher beliefs, intentions, and expectations that other measures often do not capture. The self-report process promotes teachers' reflection on their practice because it requires them to document and describe their areas of strength and professional growth needs. Self-report surveys are relatively inexpensive and easy to implement because large amounts of data can be collected at once (Goe et al., 2011).	Reliability assessments of these measures produce mixed results (Goe et al., 2008). Teachers may overreport or underreport practices either intentionally or unintentionally, may intentionally misreport their practices in order to receive higher ratings, or may unintentionally misreport their practices because they misperceive the correctness of their implementation (Goe et al., 2008).

Measure	Description	Strengths	Limitations
Approaches to Measuring Student Growth			
Simple Growth or Gain	Measures of simple growth or gain compute the difference between student performance on a pretest and posttest (Miller & Scott, 2012).	Simple growth or gain scores use longitudinal measures so they are able to capture student performance over time (Miller & Scott, 2012).	These measures are only effective when at least some of the same content appears on the pretest and posttest (Miller & Scott, 2012). In addition, simple growth or gain approaches do not account for contextual factors that may mitigate student achievement.
Value-Added Models (VAMs)	VAMs use previous student test data to predict students' level of achievement in the next school year. Models use these data to determine a particular teacher's effect on student growth. A variety of VAMs exist. Most are regression or analysis of variance (ANOVA) based, and many consider student and school characteristics (Miller & Scott, 2012).	Because these measures often control for some school and nonschool factors that may affect student achievement, VAMs provide more valid comparisons of student outcomes than student achievement measures (Miller & Scott, 2012). VAMs focus on growth rather than proficiency and thus do not penalize teachers for working with students below proficiency levels (Holdheide et al., 2012). For these reasons, many supporters of VAMs perceive them to be more objective than other measures (Little et al., 2009). Rivken (2007) argues that principals who are knowledgeable about their schools can contextualize VAM results and use them to make informed decisions about the teachers and instruction of their schools. VAMs can also be useful when looking at larger patterns, such as the distribution of "effective" teachers across schools (Goe, 2008).	Given the complex formulas used in VAMs, teachers and other stakeholders may have difficulty understanding how VAMs assess teacher performance. VAMs are unreliable when teachers have small class sizes or a high percentage of students with missing records (Milanowski, 2011). They also tend to be highly unstable from year to year due to measurement error (Goldhaber & Hansen, 2010; Meyer & Dokumaci, 2010; Newton, Darling-Hammong, Haertel, & Thomas, 2010; Steele et al., 2010). VAMs also present design issues. First, the quality of VAMs is dependent upon the quality of the measures incorporated into the model because VAMs use tests as proxies of student achievement (Rivkin, 2007); measurement error, test coverage of topics, and a lack of construct validity can all affect VAMs (Braun, 2010; Meyer & Dokumaci, 2010; Rivken, 2007). Though VAMs control for some variables, they may not account for all factors that affect student achievement (i.e., the effects of a recent parent divorce). In addition, the collaborative nature of teaching may make it difficult to distinguish one teacher's contributions to student learning from another's (Valli, Croninger, & Walters, 2007). In addition, tutors, support staff, mentors, and other teachers can all contribute to students' learning (Steele et al., 2010). When using VAMs, multiple decision rules decide how student growth will be attributed to teachers. Additionally, VAMs can only assess teachers of tested grades and subjects. As with

Measure	Description	Strengths	Limitations
			student achievement measures, VAMs do not document teachers' use of effective or ineffective practices.
Schoolwide VAM	Schoolwide VAMs use student achievement data to estimate the contributions of teachers to student academic growth. Unlike individual value-added measures, schoolwide VAMs measure student growth at the school level (Holdheide et al., 2012; Miller & Scott, 2012).	This model allows teachers of nontested subjects and grades to be accountable for student performance, and can encourage elective teachers to incorporate tested subjects (reading and mathematics) into instruction. Schoolwide growth models also have the potential to increase teacher collaboration for the good of all students (Holdheide et al., 2012).	Schoolwide models do not directly measure the effectiveness of individual teachers (Buckley & Marion, 2011). Teachers of nontested subjects and grades may believe that they have less ability to contribute to outcomes, which may decrease their buy-in and satisfaction with the evaluation system. Conversely, schoolwide VAM models may present a "free rider" problem, wherein teachers of nontested subjects receive evaluations based on the efforts of their peers rather than their own efforts (Lavy, 2007). In addition, schoolwide VAM can devalue nontested subjects and grades that are not included in the model (Holdheide et al., 2012).
Student Growth Percentiles (SGPs)	Like VAMs, SGPs use previous student test data to determine teacher contributions to student learning. With SGPs, each student's percentile rank is calculated from one year to the next to determine student growth associated with a particular teacher (Marion & Buckley, 2011).	As with VAMs, SGPs evaluate an educator's contribution to student learning. They allow districts and states to compare test score growth across groups of students with similar test histories in the same grade and subject (Miller & Scott, 2012).	VAMs and SGPs share many limitations. SGPs do not describe teacher use of effective or ineffective practices and are only useful in assessing teachers of tested grades and subjects. These measures cannot attribute the cause of student gain to teachers because schools do not randomly assign students to teachers (Steele et al., 2010). Like VAMs, SGPs attribute growth to one teacher when multiple teachers could be involved in students' learning—such as support staff, tutors, mentors, and student teachers (Steele et al., 2010).
Student Learning Objectives (SLOs)	SLOs are goals set by a teacher or group of teachers that specify specific learning targets for students. These targets include what students will know or be able to perform after completing a quarter, semester, or school year (Miller & Scott, 2012).	SLOs are highly flexible and can be set for students of any grade level or subject. Unlike some other measures, all teachers can create SLOs because they do not depend on the availability of standardized achievement tests. Though teachers are not required to standardize test scores when creating SLOs, the SLO process requires teachers to analyze trend and baseline data to set rigorous yet achievable targets. The SLO process encourages teachers to reflect on their practice and their students' outcomes (CTAC, 2008).	SLOs may not be comparable across classrooms if the process is not standardized or objective (Holdheide et al., 2012). In addition, attaching SLOs to high-stakes decisions could incentivize teachers to set easily obtainable goals (Holdheide et al., 2012). Implementing SLOs requires significant time and resources at both the district and school levels. Teachers need training on how to set appropriate growth targets, interpret student data, identify

Measure	Description	Strengths	Limitations
		Additionally, the process provides teachers the opportunity to give input on how schools measure student learning, which may increase teacher support for a performance-based evaluation system (Buckley & Marion, 2011).	trends, and adjust instruction. They also need time to complete the process. Principals need professional development so that they know how to ensure that SLOs are comparable, rigorous, and realistic. They also need sufficient training to ensure that they evaluate teacher performance in a systematic and fair way. In addition, principals must make time to review each teacher's SLOs, and district or state personnel must monitor SLO quality across schools (Holdheide et al., 2012).

Appendix B. Questions to Consider When Selecting Measures[5]

General Design Issues

- Are the measures chosen aligned to state or district teaching standards? 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 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 inclusion in summative ratings or final calculations?
- Will the measures consistently produce reliable results that would permit inclusion in a performance-based compensation system?

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 stand on including measures in teacher evaluations?
- What actions or rewards will be informed by summative scores versus results of individual measures?

Communication Considerations

- How easy is it to present the evaluation system to stakeholders?
- How clear is the statistical underpinning of the compensation system? How will the system—expectations, benchmarks, award amounts, and calculations—evolve over time and how will this message be communicated early in the process?

[5]This list as it appears is a slightly modified list of questions from Leo and Lachlan-Haché, 2012, pp. 2–3.

Appendix C. Examples of Using the Numerical Approach in Performance-Based Compensation Systems

Achievement First

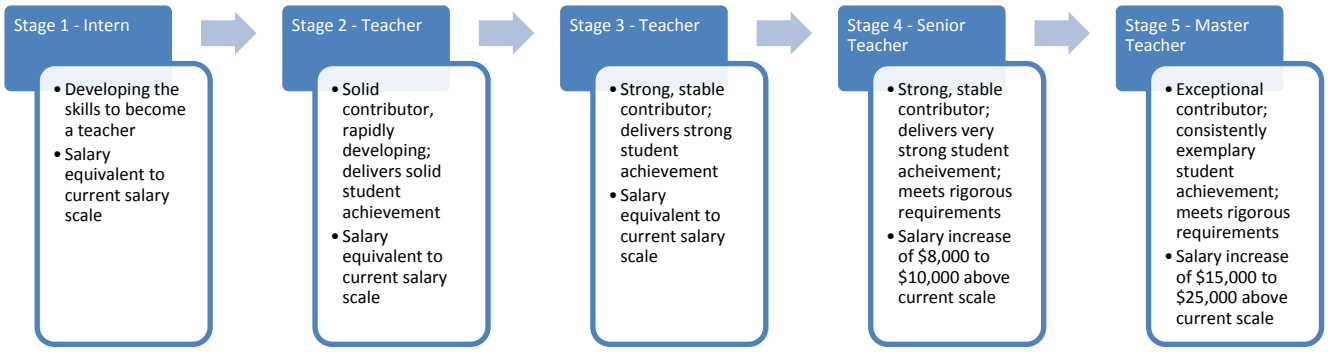
Achievement First, a charter school network in New York City, created a Teacher Excellence Framework comprised of outcomes and inputs. Table C-1 describes the measures, measurement instruments, and weighting of the evaluation system. Different weighting schemes exist for teachers who teach classes that have standardized assessments and those who do not.

Table C-1.Components, Data Sources, and Weights for Achievement First's Performance-Based Compensation System

Components	Data Sources	Weight in Tested Subjects	Weight in Untested Subjects
Student Achievement	Value-added model Principal assessment of data accuracy	40%	20%
Student Character and Development	Student survey Parent survey	15%	15%
Quality Instruction and Planning	Lesson observations Planning assessment	30%	50%
Core Values and Contributions to Team Achievement	Peer survey Principal survey	15%	15%

The Teacher Excellence Framework is tied to a Teacher Career Pathway, as shown in Figure C-1. Teachers who meet certain cut-scores have the opportunity to receive increases in salaries ($8,000 to $25,000) beyond their current scale. Teachers can also earn a schoolwide bonus of up to 10 percent of their salary at every stage of the career pathway.

Figure C-1.Achievement First's Teacher Career Pathway



Source: Achievement First via The New Teacher Project.

LEAP Academy University Charter School

The Leadership Education and Partnership (LEAP) Academy University Charter School is a public charter school serving students in grades kindergarten through 12th grade in Camden, NJ. The school's teacher compensation system is performance based. Each teacher's salary upgrade is determined each year based upon the teacher's overall evaluation score on three core elements: Teacher Effectiveness, Leadership and Professional Contributions, and Student Academic Growth and Achievement. Teachers may earn up to 100 points on their evaluation, as shown in Table C-2. The number of points determines the teacher's salary upgrade, as shown in Table C-3.

Table C-2. Distribution of Available Points in LEAP Academy University Charter School's Evaluation and Performance Based Compensation System

	In Action Points	In Reflection Points	Total Possible Points
Core Element 1: Teacher Effectiveness			
1. Planning and Preparation			
a. Setting Instructional Outcomes	3	3	6
b. Designing Coherent Instruction	3	3	6
c. Designing Student Assessments	3	3	6
2. The Classroom Environment			
a. Establishing a Culture for Learning	3	3	6
b. Managing Student Behavior	3	3	6
3. Instruction			
a. Using Questioning and Discussion Techniques	3	3	6
b. Engaging Students in Learning	3	3	6
c. Using Assessment in Instruction	3	3	6
Subtotal for Core Element 1:			**48**
Core Element 2: Leadership and Professional Contributions			
1. Professional Responsibilities			
a. Reflecting on Teaching and Student Academic Growth	3	3	6
b. Communicating with Families	3	3	6
2. Contributions to Professional Development and Growth of Colleagues			
a. Contributing to Professional Development	3	3	6
b. Engaging on and Initiating Innovative Learning Projects	3	3	6
Subtotal for Core Element 2:			**24**
Core Element 3: Student Academic Growth and Achievement			
Met NJASK, HSPA Performance or Improvement Criteria in Each Subject *OR* Met Local Assessment Performance or Improvement Criteria in Each Subject			28
Subtotal for Core Element 3:			**28**
Total Points Used to Determine Salary Increments:			**100**

Source: LEAP Academy University Charter School, 2011, p. 30.

Table C-3. LEAP Academy's Salary Upgrade Calculation Chart

Average Cumulative Scores	% Increment
100 – 86	2.25-2.34%
85 - 80	2.00-2.24%
79 - 66	1.75-1.99%
65 - 60	1.25-1.74%
59 -48	.75-1.24%
47 – 39	.5-0.76%
38 – 29	.4%
28 – 0	0.0%

Source: LEAP Academy University Charter School, 2011, p. 31.

Teachers at LEAP Academy University Charter School may also earn a one-time leadership bonus for exemplary leadership. To earn these bonuses, teachers must contribute to the school's mission or engage in and initiate innovative projects. These efforts may impact the grade level in which the teacher is working, the entire building, or the entire school or district. Teachers are rated at the end of the year on these contributions and can earn up to 15 points towards a salary bonus, which is calculated based upon the chart below.

Table C-4. LEAP Academy's Bonus Calculation Chart

Average Score	Percent Bonus
15 –10	1.0-1.5%
9-5	.75-1.0%
4-1	.5-.75%
0	0.0%

Source: LEAP Academy University Charter School, 2011, p. 33.

Appendix D. Examples of Using the Component Approach in Performance-Based Compensation Systems

Houston Independent School District

Houston Independent School District offers bonuses in three components, or strands as part o fits Accelerating Student Progress, Increasing Results and Expectations (ASPIRE) program. Eligibility for awards and maximum possible award amounts vary based upon the educator's position and the subjects taught or supported, as shown in Tables D-1 and D-2 below. For example, operational support providers are only eligible for up to $500 of the Campus Progress Award, but principals are eligible for all awards and can earn up to $13,500 in ASPIRE awards.

Table D-1. Components, Data Sources, and Awards for Teachers and School Leaders in the Houston Independent School District's ASPIRE Program

Component	Data Sources	Payout
Campus progress	Value-added model	Up to $1,850
Classroom progress	Value-added model Campus second-grade Stanford/Aprenda test	Up to $7,000
Campus achievement	Results of Stanford/Aprenda tests AP/IB participation and performance Four-year longitudinal dropout rate	Up to $1,000

Source: Houston Independent School District, 2012a, 2012b

Table D-2. Maximum Payout Amounts by Component and Educator

Educators	Maximum for campus progress	Maximum for classroom progress	Maximum for campus achievement	Maximum ASPIRE award
Principals	$1,850	$10,000	$825	$13,500
Assistant principals/deans of instruction	$925	$5,000	$425	$6,750
Teachers with value-added scores	$1,000	$7,000	$1,000	$9,000
PreK- grade 2 teachers and special educators	$1,000	$3,500	$1,000	$5,500
Ancilary and elective teachers	$1,000		$1,000	$2000
Instructional support providers	$750		$600	$1,350
Teaching assistants	$750		$400	$1,150
Operational support providers	$500			$500

Source: Houston Independent School District, 2012a, 2012b

Henrico County Public Schools

The Learning Leaders program is a Teacher Incentive Fund program in Henrico County, VA. The program offers differentiated incentives for teachers and principals. Teachers may earn up to $9,000 and principals up to $10,000 in one-time bonuses.

Table D-3. Components and Awards for Teachers and Principals in Henrico County Public Schools

Teachers		Principals	
Component	**Payout**	**Component**	**Payout**
Teachers will receive an incentive for meeting *100% of their implementation of the *PQRs. This will be determined by walk-through and classroom observation data.	Up to $3,000	Principals will receive an incentive for meeting *100% of their targets for supporting teacher growth in implementation of the PQRs. This will be determined by examination of walk-through and observation data collected and their documented post-observation conferencing skills.	Up to $4,000
Teachers will receive an incentive for meeting *100% of their student learning goals for each student on their roster. Student learning	Up to $5,000	Principals will receive an incentive for meeting identified school-wide student achievement targets. These will be set by the	Up to $3,000

targets will be based on state test scores and value-added growth measure assessments.		principals' supervisor and include state test pass rate, attendance, discipline referrals, parent involvement, and stakeholder perception survey results.	
Secondary math, science, and special education teachers will receive a one-time signing bonus	$1,000	Principals will receive an incentive if the teachers they supervise make *100% of their student learning targets for each individual student on their roster. Since the principal's role is to ensure teachers and students are successful, this target will be tied to teacher success	$3,000
Total possible teacher incentive compensation	**$9,000**	**Total possible principal incentive compensation**	**$10,000**

Source: Henrico County Public Schools, 2012

Pittsburgh Public Schools

Since the 2010-11 school year, Pittsburgh Public Schools has offered its teachers a variety of rewards and recognition opportunities. Each opportunity has different eligibility and performance requirements. For example, the Promise-Readiness Corps Cohort Award is offered only to teams of high school teachers who move a cohort of 9th grade students to 11th grade on track to graduate Promise-Ready. Another award, the Students and Teachers Achieving Results (STAR) award, is available to all union-represented staff at schools that have achieved significant growth compared to the statewide average. In addition to having different requirements, the form of compensation for each opportunity varies as well. For example, the career ladders and new teacher salary schedule impact the teacher's base salary whereas the Promise-Readiness Corps Cohort, STAR, AYP, and Voluntary Incentive Earnings at Work (VIEW) awards compensate teachers in the form of one-time bonuses.

Table D-4. Performance-Based Compensation Opportunities and Payouts in Pittsburgh Public Schools

Level	Opportunity	Description	Launched
District	AYP Award	In each year that the District attains AYP, teachers at the top step of the salary schedule receive a $1,000 reward.	2010-11
School	Students and Teachers Achieving Results (STAR)	All PFT-represented staff at schools achieving significant growth compared to schools statewide earn an award of up to $6,000.	2011-12
Team	Promise-Readiness Corps Cohort Award	Teams of teachers that move a cohort of 9th graders to 11th grade on track to graduate Promise-Ready earn an award of up to $20,000 upon completion of a two year "loop" with their students.	2010-11
Individual	Career Ladders	Roles for effective teachers to take on leadership responsibilities, discover unique career pathways and earn between a $10,000 - $14,000 differential.	2010-11
	New Teacher Salary Schedule	A new salary plan where earnings potential for new teachers exceeds $100,000 for effective performance.	2010-11
	Voluntary Incentive Earnings at Work (VIEW)	Teachers who choose to participate earn awards up to $8,000 per year by demonstrating their ability to grow student achievement. Up to 225 teachers will participate in VIEW through the duration of the pilot.	Anticipated launch in 2013-14

Source: Excellence for All, 2012

ABOUT AMERICAN INSTITUTES FOR RESEARCH

Established in 1946, with headquarters in Washington, D.C., American Institutes for Research (AIR) is an independent, nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance both domestically and internationally. As one of the largest behavioral and social science research organizations in the world, AIR is committed to empowering communities and institutions with innovative solutions to the most critical challenges in education, health, workforce, and international development.

LOCATIONS

Domestic

Washington, D.C.
Atlanta, GA
Baltimore, MD
Chapel Hill, NC
Chicago, IL
Columbus, OH
Frederick, MD
Honolulu, HI
Naperville, IL
New York, NY
Portland, OR
Sacramento, CA
San Diego, CA
San Mateo, CA
Silver Spring, MD
Waltham, MA

INTERNATIONAL

Egypt
Ethiopia
Georgia
Haiti
Honduras
Kenya
Liberia
Malawi
Nicaragua
Pakistan
South Africa
Zambia

2815_09/12

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

www.air.org

Making Research Relevant

2815_09/12