HANOVER RESEARCH

1750 H Street NW, 2nd Floor, Washington, DC 20006
P 202.756.2971 F 866.808.6585 hanoverresearch.com

Best Practices for Including Multiple Measures in Teacher Evaluations

In this report, Hanover Research addresses the use of multiple-measure teacher evaluation models. The report examines the advantages of a multiple-measure approach, the types of data sources commonly used in multiple-measure evaluations, and best practices in the implementation of multiple-measure teacher evaluations. The report draws on a review of the literature, as well as real-world examples from four public school districts.

Introduction and Key Findings

Effective teacher evaluation is a growing concern among educators and policymakers. While single-measure approaches, namely value-added models, have grown in prominence since the launch of No Child Left Behind (NCLB), there is a growing impetus to take a more holistic approach to teacher evaluation. In this report, Hanover Research addresses the use of multiple-measure teacher evaluation models. The report examines the advantages of a multiple-measure approach, the types of data sources commonly used in multiple-measure evaluations, and best practices in the implementation of multiple-measure teacher evaluations.

The report is structured in three primary sections. For comparative purposes, the first section looks briefly at the use of value-added models and the drawbacks of single-measure evaluations, then turns the focus to multiple-measure evaluations, with a discussion of the advantages of this model, the types of data commonly collected, and best practices in implementation. The second section provides a closer look at five common categories of data considered in multiple-measure evaluations: classroom observations; administrator reports; student perception data; parent perception data; and documentation of professional activity. This section also discusses the ways in which districts have handled evaluations of teachers who teach non-tested subjects. Finally, the third section provides a brief introduction to the 360 degree feedback model, then presents profiles of four school districts that utilize multiple-measure teacher evaluations.

Key Findings

The following are the key findings of this report:

- Along with placing a spotlight on student achievement, NCLB has changed the landscape of teacher evaluation. While evaluations have become more closely linked to student achievement, some scholars and educators alike have called for a more holistic approach to teacher evaluations, and many districts have opted to use multiple-measure teacher evaluations. Oftentimes, these evaluation models continue to look at student achievement data, but also take into account numerous other sources, such as classroom observations and student and parent perceptions.

- Multiple-measure evaluations carry significant benefits in the way of accuracy, with increased validity of observations and reliability of feedback being cited as advantages. However, multiple-measure evaluations can also require a significant commitment of time and resources.

- Designing and implementing teacher evaluation systems is often a complex process that involves many stakeholders. School districts will likely want to take an approach that engages different groups (e.g., teachers, parents, students, administrators, and the community) and that communicates clearly how evaluations will be used.

- There are no universally agreed upon best practices for multiple-measure evaluation systems, and opinions vary widely over what types of measures ought to be included in evaluations. It is common, however, to include student achievement (value-added) data, classroom observations, administrator reports, parent surveys, student reports, portfolios, examples of work, and, sometimes, peer reviews.

- Educators and researchers alike have viewed administrator reports as controversial, but recent research has indicated that administrators' perceptions of teacher performance are often an accurate indicator of actual quality.

- Parent and student surveys have often been overlooked as tools to evaluate teachers. However, students' perceptions have been found to be surprisingly accurate, leading to their inclusion in some evaluation systems. Parent surveys may be a way of obtaining perception data on overall teacher quality for formative purposes and to build community inclusion.

- Because many evaluation systems place high importance on value-added data, many school districts have begun to try to formulate ways to obtain similar measures from less commonly tested subjects. A number of approaches exist that attempt to replicate the types of data produced on standardized tests. Still, there are few approaches that may be regarded as universal best practice.

Section One: Introduction to the Value-Added Model and Multiple-Measure Approaches to Teacher Evaluations

In order to provide context for our discussion of multiple-measure evaluations, this section first discusses single-measure, value-added models that focus solely on student achievement gains and the potential drawbacks of this approach. The discussion then turns to the alternative of multiple-measure evaluations. We describe the advantages of this model, the types of data commonly used in such evaluations, and general best practices and considerations for implementation.

Standardized Tests and Value-Added Models

The enhanced emphasis on teacher quality that has grown out of educational reforms such as No Child Left Behind (NCLB) and the Race to the Top Initiative (RTT) has led to rapid evolution and innovation in the way teacher evaluations are conducted. In general, research on teacher quality and performance—the focal points of teacher evaluations—suggests three fundamental aspects to be considered: 1) teacher qualifications, 2) teacher behaviors, and 3) teacher outcomes. A number of mechanisms exist for the determination of the first two; teacher qualifications are evidenced by certification assessments and degree levels, for example, while teacher behaviors are determined through routes such as classroom observations, portfolios, surveys, and administrator reports.[1] It is the third aspect—teacher outcomes (i.e., the impacts a teacher has in the classroom)—that has traditionally been more difficult to measure and evaluate.

Coinciding with the launch of NCLB, a recent push toward measuring teacher outcomes has largely been addressed through a focus on achievement test data. The approach has been lauded as a legitimate procedure for teacher evaluations as students' test performance—a quantitative, outside measure—leaves out all subjectivity from the process. It has been suggested that evaluations based purely on observations and portfolios leave too much room for subjectivity and often lack consistency from one evaluation to the next.[2] Evaluations based on achievement tests, or "value-added models" (VAM), eliminate such subjectivity and have come to occupy a prominent position in debates on teacher evaluation models. **Relying on growth in student test scores to determine the "value" teachers contribute to student outcomes, value-added models aim to provide a single, objective means to measure and quantify teachers' contributions to student learning.**

[1] Blackmon, L., E. Harden, E. Reynolds, M. Shepherd, K. Skinner, V. Wilburn. "The Use of Student Achievement Data in Teacher Evaluation: A Field Based Research Project." The College of William and Mary. P. 5. http://ginnywilburn.wmwikis.net/file/view/Final+Paper-Student+Achievement+%26+Teacher+Evaluation.doc

[2] Joshua, M., A. Joshua, and W. Kritsonis. "Use of Student Achievement Scores as Basis for Assessing Teacher's Instructional Effectiveness: Issues and Research Results." *National Forum of Teacher Educational Journal*, 17:3. P. 3. http://www.nationalforum.com/Electronic%20Journal%20Volumes/Joshua,%20Monday%20Use%20of%20Student%20Achievement.pdf

Despite their objectivity, however, value-added models have themselves been the source of criticism by some educators and scholars. The heavy reliance on test score data, for example, has been questioned, as researchers have found that teacher effectiveness measured by test data is highly unstable, with evident fluctuations from class to class and from year to year, as well as from one test or statistical model to another.[3] These inconsistencies may arise from factors outside of the teacher's realm of influence. Several external factors are difficult to reliably include in value-added models, including student IQ, family background, peer group, and interest in school.[4] Furthermore, student performance is unlikely to be the result of one teacher alone. One study, for example, conducted in 2005, found that the performance of 5th grade students continued to be affected by the quality of their teachers two years earlier, at the 3rd grade level. An analysis of teacher performance based solely on quantitative student achievement data, then, may not be able to accurately determine the influence of a single teacher on student performance.[5]

Several other concerns also cast doubt on the efficacy of using student achievement data to evaluate teachers, including factors such as test quality, teacher acceptance of the system, and a number of factors that may affect student performance but that teachers have no control over. Furthermore, **value-added models may create incentives for educators to "teach to the test," as well as disincentives for teachers to seek positions in classrooms with high-need students. Furthermore, teachers who specialize in subjects such as music, art, and foreign languages are not typically accounted for by standardized test results.**[6] Many teachers oppose VAM and have been shown to reject the notion that their performance will be accurately reflected by student test data.[7]

For all of these reasons, **there is an emerging consensus that VAM is not appropriate as the sole measure for teacher evaluations.** Studies published by the Rand Corporation, the National Research Council, and the Educational Testing Service (ETS) have all argued unequivocally that standardized tests should never *completely* determine high-stakes personnel decisions that may impact teachers' future paths in the profession. A 2005 RAND Corporation report, for example, notes, "The research base is currently insufficient to support the use of VAM for high-stakes decisions about individual teachers or schools."[8] Similarly, a 2005 ETS report notes that VAM results should not be used as the "sole or principal basis" for

[3] Darling-Hammond, L. and A. Beardsley. 2011. "Evaluating Value-Added Models and Other Methods." *Phi Delta Kappan*. http://opendev.stanford.edu/sites/default/files/Evaluating%20Teacher%20Evaluation,%20Linda%20Darl.H-%20PDK.doc

[4] Joshua, *et al. Op. cit.*, p. 4.

[5] Blackmon, *et al. Op. cit.*, p. 9.

[6] Blackmon, *et al. Op. cit.*, pp. 7-9. And: Darling-Hammond and Beardsley. *Op. cit.*, pp. 3-4.

[7] Blackmon, *et al. Op. cit.*, p. 9.

[8] McCaffrey, D. F., Koretz, D., Lockwood, J.R., Hamilton, L.S. 2005. *Evaluating Value-Added Models for Teacher Accountability*. RAND Corporation. P. xx. http://www.rand.org/pubs/monographs/MG158.html

"consequential decisions" about teacher effectiveness.[9] The report contends that there are numerous pitfalls to teacher effectiveness being casually linked to the data commonly collected by school districts, and that the education community still lacks sufficient knowledge of how technical problems might "threaten the validity of such interpretations."[10] Finally, in a 2009 document, the National Research Council points out that VAM results based on data for just one class of students are insufficient to support operational decisions due to the previously noted instability of results.[11]

Multiple-Measure Approaches

Benefits of the Multiple-Measure Approach

Despite the flaws of value-added models, they are likely to endure for the foreseeable future even if absorbed into multiple-measure models. **In a review of teacher outcomes, the data produced by value-added models may simply be too valuable to completely ignore. Furthermore, VAM measures may be particularly useful if the information they provide can be corroborated by other sources.** Just as scientific inferences may be strengthened by multiple observations, teacher evaluations benefit from a wealth of data. In the *Handbook on Teacher Evaluation*, authors James H. Stronge and Pamela D. Tucker build a case for multiple measures to be adopted in evaluations of teacher effectiveness and performance. They hold that good teacher evaluations are driven by similar principles as those used in the social sciences or statistics. They summarize the strengths of multiple-measure evaluations as follows:[12]

- **Increased validity:** Validity is achieved through an increase in the number of performance components being evaluated, which ultimately leads to more accurate information than could be achieved through just one example of a teacher's performance.

- **Increased reliability:** Reliability (or consistency) typically comes with an increase in sample size. A diverse collection of data, including input from multiple perspectives, will increase the possibility for corroboration of other observations.

- **Decreased subjectivity:** While value-added models are often purported to be more objective than traditional means of evaluation, models that utilize multiple sources may also help build objectivity in the evaluation process as

[9] Braun, H. 2005. *Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models.* (Princeton, NJ: Educational Testing Service). P. 8. http://www.isbe.state.il.us/peac/pdf/primer.pdf
[10] *Ibid.*
[11] National Research Council, Board on Testing and Assessment. 2009. "Letter Report to the U.S. Department of Education." P. 10. http://www.nap.edu/openbook.php?record_id=12780
[12] Stronge, J.H. and Tucker, P.D. *Handbook on Teacher Evaluation*. Larchmont: Eye on Education, 2003. P. 64.

data can be collected from numerous sources, with one source of input checked against another.

Stronge and Tucker also point out that the use of multiple measures often leads to an **increased comfort level for teachers and evaluators alike**; teachers will find comfort in the fact that their evaluations will consider input from numerous and varied sources, and evaluators will feel that a burden is lifted when they are no longer considered the sole reviewer. The use of multiple measures further allows the evaluator to glean a **more realistic "portrait" of the teacher's performance** and to **integrate both primary and secondary data sources** into the evaluation process.[13]

Recommended Data Sources for Multiple-Measure Evaluations

In *Teacher Evaluation: A Comprehensive Guide to New Directions and Practices*, author Kenneth Peterson lays out a number of recommendations for changing the nature of teacher evaluation, which he sees as being vital, but too often characterized by mismanagement and poor practice. While written over a decade ago, Peterson's recommendations still offer a useful guide on the topic and address many of the concerns that educators and administrators continue to tackle today as they approach evaluations. According to Peterson, **school districts should use multiple data sources to inform judgments about teacher quality.** More specifically, Peterson recommends that evaluators consider the following types of data:[14]

- Student reports
- Peer review of materials
- Student achievement
- Teacher tests
- Parent reports
- Documentation of professional activity
- Systematic observation
- Administrator reports

In Peterson's view, the use of different data sources offers a way to organize different types of information relevant to teacher quality. The process allows evaluators to see multiple dimensions of a teacher's performance that might not be captured through a single source. Figure 1.1 highlights how Peterson views the data sources that correspond with various areas of teacher quality.

[13] *Ibid.*

[14] Peterson, K. *Teacher Evaluation: A Comprehensive Guide to New Directions and Practices*. Thousand Oaks: Corwin Press, 2000. Pp. 91-103.

Figure 1.1: Teacher Performance Areas and Corresponding Data Sources[15]

	Student reports	Systematic observation	Pupil achievement	Peer review of materials	Teacher tests	Professional Activity	Administrator Reports	Parent Reports
Creates opportunities to learn	X	X		X				
Student gains			X	X				
Academic quality				X	X	X		
Follows district and state guidelines				X			X	
Member of school community						X	X	
Maintains health and safety conditions							X	
Ethical Practice							X	
Parent Relations								X

Adapted from: Peterson, K. 2000. *Teacher Evaluation: A Comprehensive Guide to New Directions and Practices.*

While it is important to gather input from numerous sources, in the multiple-measure approach, *more input* should not be prized over *valuable input*. Indeed, Peterson cautions against the use of several sources that may be unreliable, invalid, or misleading, including:[16]

- Testimonials
- Peer visits to classrooms
- Peer consensus
- Graduate follow-up
- Microteaching performances
- Self-evaluation
- Observations of classroom environment

While Peterson recommends against the inclusion of these sources in teacher evaluations, it should be noted that some advocates of multiple measure evaluations argue for their inclusion (or inclusion of subtle variations). Figure 1.2 lists data sources that have been recommended by various authors and organizations for inclusion in multiple-measure approaches. While there appears to be substantial consistency regarding the merits of certain data sources, particularly observations and student achievement, other sources are less universally regarded as important measures.

[15] *Ibid.*, p. 98.
[16] *Ibid.*

Figure 1.2: What to Measure? Potential Data Sources for Multiple Measure Teacher Evaluations

Peterson: *Teacher Evaluation: A Comprehensive Guide to New Directions and Practices*[17]	Organisation for Economic Co-operation and Development[18]	National Comprehensive Center for Teacher Quality[19]	Stronge and Tucker: *Handbook on Teacher Evaluation: Assessing and Improving Performance*[20]
✓ Student reports ✓ Peer review of materials ✓ Student achievement ✓ Teacher tests ✓ Parent reports ✓ Documentation of professional activity ✓ Systematic observation ✓ Administrator reports[21]	✓ Classroom observation ✓ Objective setting and individual interviews ✓ Teacher self-evaluation ✓ Teacher portfolio ✓ Teacher tests ✓ Student results ✓ Surveys of students ✓ Surveys of parents	✓ Classroom observation ✓ Principal evaluation ✓ Instructional artifacts ✓ Portfolios ✓ Teacher self-report measures ✓ Student surveys ✓ Value-added model	✓ Observations ✓ Portfolios ✓ Student learning measures ✓ Annual goals ✓ Client surveys ✓ Self-evaluation ✓ Documentation (lesson plans, grades, tests)

The lists in Figure 1.2 provide examples of the types of multiple measures put forth in the literature on teacher evaluations. Each source has strengths and drawbacks that districts must consider as they implement evaluation systems, which will be addressed in greater depth in the second section of this report. First, however, this section will briefly review general factors that school districts should consider as they implement multiple-measure evaluation systems.

Implementation of Multiple-Measure Evaluation Systems

The use of a multiple-measure approach to teacher evaluations is not without challenges. The implementation of any teacher evaluation system involves challenges and numerous considerations, such as the "accuracy of the measurement, inclusion of all the dimensions of what is meant to be measured, consistency with the goals of the feedback exercise, adaptation to the needs of those who will use the results (teachers, school leaders, educational authorities), cost-effectiveness, and practical feasibility."[22] **Multiple-measure evaluations clearly require time and initiative from teachers and administrators.** Buy-in from teachers is critical, as the use of multiple data sources not only opens the door for a broader scope of excellent or effective performance, but also the prospect of additional criticism for weak performance in

[17] *Ibid.*
[18]" Teacher Evaluation: A Conceptual Framework and Examples of Country Practices." 2009. OECD. Pp. 14-16. http://www.oecd.org/dataoecd/16/24/44568106.pdf
[19] Goe, L., C. Bell, and O. Little. 2008. "Approaches to Evaluating Teacher Effectiveness: A Research Synthesis." *National Comprehensive Center for Teacher Quality.* Pp. 16-18. http://www.tqsource.org/publications/EvaluatingTeachEffectiveness.pdf
[20] Stronge and Tucker. *Op. Cit.*, pp. 91-103.
[21] *Ibid.*
[22] Teacher Evaluation: A Conceptual Framework…" *Op. cit.*, p. 11.

other areas.[23] Ultimately, best practice will result from thorough assessment of a school system's unique needs and the use of evaluations in a purposeful manner.

Articulation of Evaluation Goals: Formative and Summative Assessments

Critical in the initial stage of implementation is the ability of district administrators to clearly articulate the purpose of the formative or summative components of the evaluation process. Summative evaluations generally reveal information about participants' overall competencies, whereas formative assessments rate competencies at multiple points in time in order to help guide future strategies.[24] In other words, **formative assessments allow teachers and administrators to utilize constructive feedback in order to improve instructional or leadership techniques**.

Schools and school districts must define whether the evaluation activities are to play a formative/improvement function or an accountability/summative function, or a combination of both. According to a review by the OECD, there is an **inherent tension** between the two functions:[25]

> When the evaluation is oriented towards the improvement of practice within schools, **teachers are typically open to reveal their weaknesses**, in the expectation that conveying that information will lead to more effective decisions on developmental needs and training. However, when teachers are confronted with potential consequences of evaluation on their career and salary, the inclination to reveal weak aspects of performance is reduced, i.e. the improvement function is jeopardized. Also, using the same evaluation process for both purposes undermines the usefulness of some instruments (such as self-evaluation), and creates an additional burden on evaluators as their decisions have somewhat conflicting consequences (e.g. tension between improving performance by identifying weaknesses and limiting career progression, if the evaluation prevents teachers from advancing in their career).

Administrators must think carefully about the goal of evaluation procedures and the conditions that will best promote the summative or formative nature of the process. Both contexts may present challenges. Summative evaluations with an emphasis on accountability, for example, may lead to insecurity or fear among teachers, while formative evaluations may lead to teacher or union expectations for social recognition of quality work or enhanced professional development opportunities.

---

[23] Stronge and Tucker. *Op. cit.*, p. 99.
[24] Condon, C. and M. Clifford. 2010. "Measuring Principal Performance: How Rigorous Are Commonly Used Principal Performance Assessment Instruments?" Learning Point Associates. P. 1. http://www.learningpt.org/pdfs/QSLBrief2.pdf
[25] "Teacher Evaluation: A Conceptual Framework…" *Op. cit.*, pp. 8-9.

Conditions that will support **formative teacher evaluation processes aimed at improvement** include the following:[26]

- A non-threatening evaluation context;
- A culture of mutually providing and receiving feedback;
- Clear individual and collective objectives with regard to improving teaching within the school as well as sharing of school objectives;
- Simple evaluation instruments such as self-evaluation forms, classroom observation, and structured interviews;
- A supportive school leadership;
- Opportunities to enhance competencies as well as resources and means to improve practice;
- Teacher evaluation integrated in a system of school self-evaluation and quality assurance.

Meanwhile, conditions that will support **summative teacher evaluation processes used for accountability purposes** include the following:[27]

- An independent and objective assessment of the teacher's performance;
- National-level standards and criteria across schools;
- An evaluation component external to the school and more formal processes;
- Well-established rules regarding the consequences of the evaluation;
- Clear individual objectives with regard to all aspects of a teacher's performance;
- Well-trained, competent evaluators of teaching performance;
- Impact on professional development plan;
- Possibilities for appeal for teachers who feel they have not been treated fairly.

Best Practices in Implementation of Multiple Source Evaluations

As districts plan and implement multiple-measure teacher evaluation systems, whether within the context of formative (growth) or summative (accountability) goals, the following best practices should be considered: [28]

[26] *Ibid.*
[27] *Ibid.*
[28] *Ibid.* And: Goe *et al. Op. cit.*

Communication should be a high priority	• Procedures should be developed through a collaborative, public forum that involves multiple stakeholders.
Individual performance should be linked to overall school goals	• Evaluation systems that rely on multiple sources of data are more effective when they use data or combinations of data that reflect the school as a whole.
The context of the evaluation system should be taken into account	• Administrators must understand and account for context and make realistic plans for what evaluation will involve at an institutional level. Multiple measure evaluation is likely to involve an ongoing process that includes substantial commitments of time and energy.
Evaluation should facilitate professional growth:	• Finally, evaluation should be directed towards improvement, with recognition for exemplary performance and actionable feedback given at intermediate levels

The National Comprehensive Center for Teacher Quality also stresses the importance of communication in the design and implementation of new teacher evaluation systems. **In particular, it stresses that districts should use systematic communication to involve teachers in evaluation design and to ensure that they understand the system prior to evaluation time.** The NCCTQ also recommends that district administrators think carefully about who evaluators should be, as research shows that "teachers highly regard evaluators with deep knowledge of curriculum, content, and instruction who can provide suggestions for improvement."[29] In multiple-measure models, particularly 360 degree feedback models, districts may consider the use of other evaluators in addition to traditional administrator (principal) evaluators. **Also of importance is the provision of training opportunities to ensure that evaluators thoroughly understand the evaluation rubric and the characteristics and behaviors expected in high-quality classroom instruction**.[30]

[29] Mathers, C., M. Oliva, and S.W.M. Laine. 2008. "Effective Teacher Evaluation: Options for States and Districts." National Comprehensive Center for Teacher Quality. Pp. 9-10. http://www.tqsource.org/publications/February2008Brief.pdf

[30] *Ibid.*

Section Two: Best Practices and Considerations for the Integration of Multiple Measures into Teacher Evaluations

The previous section introduced several sources of data that might be included in multiple-measure teacher evaluation systems. In this section, we examine how those measures might be integrated into a multiple-measure evaluation system. As some of these approaches have been the subject of extensive research and debate in their own right, the goal of this section will be to provide a brief overview and snapshot of recent research on the use of each measure in teacher evaluations, rather than an extensive literature review. For reference, the techniques profiled in this section include:

- Classroom Observation
- Administrator Reports
- Student Perception Data
- Parent Perception Data
- Documentation of Professional Activity

In addition to these measures, this section also presents a brief overview of approaches to teacher evaluations in subjects not traditionally evaluated through student achievement exams.

Classroom Observation

Teacher observation has historically been one of the most popular methods of teacher evaluation in the United States. **This approach provides rich information about classroom behaviors and activities, is generally considered a fair and direct measure by stakeholders, and can be used for formative or summative purposes.** At the same time, however, the use of classroom observations suffers from a lack of a strong research base, and in practice, a lack of rigor. Research and validity findings are highly dependent on the instrument used, sampling procedures, and the training received by observers. Finally, classroom observations can be expensive and time consuming.[31]

Classroom observation can take both formative and summative forms. For example, in a summative evaluation (e.g., a performance review) a supervisor might conduct an announced or unannounced classroom visit when a teacher presents a new lesson. The feedback from this evaluation would then be provided at a fixed point in time during an annual or periodic performance review. By contrast, formative evaluations are more focused on continuous, incremental improvement, and might include

[31] Goe, *et al. Op. cit.*

incidental observations intended to provide more frequent information on a wider variety of contributions made by teachers in the classroom or community.[32]

Recently, the nation's eye has been turned towards the use of observations to evaluate teachers, in part due to the Bill and Melinda Gates Foundation's Measures of Effective Teaching (MET) project. The MET project, an innovative approach to evaluating teacher effectiveness that centers on observation, has issued a number of guidelines for implementing and utilizing observations in multiple-measure evaluation systems:[33]

1. Choose an observation instrument that sets clear expectations
2. Require observers to demonstrate accuracy before they rate teacher practice.
3. When high-stakes decisions are being made, multiple observations are necessary.
4. Track system-level reliability by double scoring some teachers with impartial observers.
5. Combine observations with student achievement gains and student feedback.
6. Regularly verify that teachers with stronger observation scores also have stronger student achievement gains on average.

One technique that may be particularly valuable in formative evaluation is instructional rounds, an approach based on medical rounds in hospitals and other health care settings. As described by Richard Elmore, **while walkthroughs are typically used to evaluate teachers, instructional rounds take a more focused, slower, and deliberate approach as participants visit classrooms and describe what they see using structured protocols.** Instructional rounds are generally conducted with the aim to improve instruction, and discussions and an exchange of information focused on improving practice occur after the debriefing. In Elmore's model of instructional rounds, groups of superintendents from other schools serve as the observers.[34]

For more formal, summative observations, school districts should develop or adopt a protocol to guide evaluation. Fortunately for districts and schools, there are numerous examples of observation protocols that have been produced and refined through research. Examples include:[35]

- University of Virginia's Classroom Assessment Scoring System

[32] *Ibid.*
[33] Quoted verbatim from: "Gathering Feedback for Teaching." 2012. The Bill and Melinda Gates Foundation. http://metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf
[34] Elmore, R.F. 2007. "Professional Networks and School Improvements." *The School Administrator.* http://pdisl.magicaer.com/suptnetwork/elmore_on_networks.pdf
[35] "Gathering Feedback for Teaching." 2012. The Bill and Melinda Gates Foundation. http://metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf

- Reformed Teaching Observation Protocol (RTOP) (for math and science)
- Quality of Mathematics in Instruction (QMI)
- TEX-IN3 (literacy)
- Framework for Teaching and Research
- Mathematical Quality of Instruction (MQI)
- UTeach Teacher Observation Protocol (UTOP)
- Protocol for Language Arts Teaching Observations (PLATO)

In sum, teacher observation may be considered one of the foundations of multiple-measure evaluation systems, and while observations entail a relatively substantial time commitment, models and rubrics generally have fairly straightforward implementation procedures.

Administrator Reports

Administrator reports **may utilize an assortment of data, including firsthand observations, as well as reviews of a teacher's role in the school community, professional development pursuits, and relationship with parents.** The principal is generally the evaluator. Administrator reports are generally used for summative purposes, most commonly for tenure decisions for beginning teachers. While administrator reports are one of the most common means of teacher evaluation, they vary significantly in format—some are conducted as formal, scheduled observations with the use of validated instruments and pre- and post-observation interviews with teachers, while others are conducted merely as informal, drop-in visits.[36] Given the increased focus on accountability for student achievement, administrator evaluations have received renewed attention. In particular, there has been a shift from the use of summative evaluations to the use of formative evaluations (i.e., toward greater emphasis on performance improvement).[37]

While popular within the context of teacher evaluations, administrator reports often rely on a weak base of knowledge about instructional practice and classroom techniques, which can weaken their value. Due to supervisors' lack of familiarity with excellent teaching practices and/or lack of training in the identification of high-quality and effective practices, in some districts, a majority of teachers receive a "satisfactory" rating on their evaluations. This is sometimes the case even in schools where many students fail to achieve satisfactory results on the state's standardized tests. Other common pitfalls that districts should be careful to avoid in the use of administrator reports include: the use of outdated evaluation criteria in a "checklist" form; simplistic evaluation comments (e.g., "satisfactory" or "needs improvement"); the use of the same evaluation for novice and experienced educators; and one-way,

[36] Goe, *et al. Op. cit.*
[37] Wilkerson, D., *et al.* 2000. "Validation of Student, Principal, and Self-Ratings in 360[degrees] Feedback for Teacher Evaluation." *Journal of Personnel Evaluation in Education*, 14:2. P. 180.

top-down evaluation procedures that inhibit communication and the development of professional goals.[38]

Historically, research into whether subjective supervisor performance ratings match actual, objective performance has suggested a weak relationship. Research from a number of professions suggests that supervisor ratings are often heavily influenced by a number of demographic and interpersonal factors.[39] However, in an analysis that compared principals' ratings with test scores, scholars Brian Jacob and Lars Lefgren found that principals' perceptions of teacher performance were strongly correlated with test scores.[40] Similarly, in their research, Douglas Harris and Tim Sass have found that **principals' evaluations are generally, though not always, better predictors of a teacher's contribution to student outcomes than traditional measures of experience or education**. Principals often assign a high level of importance to a teacher's ability to raise student test scores, though **they are also interested in a combination of subject knowledge, instructional skill, intelligence, and to a lesser extent, with ability to work with others.**[41] Still, principals' ratings may have limitations. For example, Jacob and Lefgren note that, while principals are able to accurately distinguish between very good and very poor teachers, they often encounter trouble in identification of those in between.[42]

In order to combat some of the problems cited above, the evaluation processes used by principals and other administrators to assess teacher effectiveness should follow several best practice guidelines, including:[43]

- The use of a **consistent definition** for high-quality educational quality
- Application of a **common language** to describe good practice
- **Training and preparation** for classroom observations so that administrators can identify good practice, interpret evidence, and engage teachers in productive post-observation conversations about performance

A revised classroom observation process recommended by scholar Charlotte Danielson incorporates the following steps:[44]

---

[38] Danielson, C. "Evaluations that Help Teachers Learn." *The Effective Educator*. 68:4. Pp. 35-39. http://www.danielsongroup.org/UserFiles/files/Evaluations%20That%20Help%20Teachers%20Learn%20-%20Danielson.pdf

[39] Jacob, B.A. and L. Lefgren. 2005. "Principals as Agents: Subjective Performance Measure in Education. NBER Working Paper Survey. Pp. 5-6. http://www-personal.umich.edu/~bajacob/files/Teacher%20Labor%20Markets/principals%20as%20agents.PDF

[40] *Ibid.*

[41] Harris, D. N. and T.R. Sass. 2007. "What Makes for a Good Teacher and Who Can Tell?" NBER Working Paper. http://www.nber.org/public_html/confer/2007/si2007/ED/sass.pdf

[42] Jacob and Lefgren. *Op. cit.*

[43] Danielson. *Op. cit.*

[44] *Ibid.*

- Thorough notes taken by the administrator during observation
- Review of these notes by the teacher post-observation
- Analysis of the notes against evaluative criteria and performance levels
- Self-reflection by the teacher
- Post-observation conference in which the teacher can offer the administrator context for the lesson and describe challenges
- Discussion of strengths and areas for improvement, as well as recommendations for strategies

Student Perception Data

Several research studies have demonstrated that students' ratings of their teachers can provide useful information about teacher performance. It has even been suggested that student surveys may provide information that is as valid as judgments made by college students, a population among which the practice of rating teachers is much more widespread. Despite indications that student reports are a potentially promising means of teacher evaluation, **their validity is dependent on the instrument used and they are generally recommended for formative use only.** Student reports have not been validated for summative assessment, nor should they be used as a primary measure of teacher evaluation. Most significantly, students cannot provide information on aspects of a teacher's subject knowledge, adherence to the curriculum, or other aspects of their professional activities.[45]

Despite these drawbacks, student surveys provide valuable information for teachers as they seek to improve their classroom practice and have garnered attention from education scholars as an important area of research. Several researchers have begun to develop rubrics for future teacher evaluation efforts in this area. For example, the Tripod Project survey, developed through a partnership between Cambridge Education and Harvard University's Dr. Ronald F. Ferguson, seeks to assess "whether students' perceptions of the teaching they experience help in predicting how much those students learn."[46] The Tripod Project survey has been administered in large, urban school districts and has been applied through the Bill and Melinda Gates Foundation's two-year Measures of Effective Teaching (MET) project.[47] The Tripod Project surveys are designed to measure students' perceptions of teacher performance in the "Seven Cs:"

- **Caring** about students (Encouragement and Support)
- **Captivating** students (Learning Seems Interesting and Relevant)

[45] Goe, *et. al. Op. Cit.*
[46] "Student Perceptions and the MET Project. 2010. The Bill and Melinda Gates Foundation. http://metproject.org/downloads/Student_Perceptions_092110.pdf
[47] *Ibid.*, p. 2.

- **Conferring** with students (Students Sense their Ideas are Respected)
- **Controlling** behavior (Culture of Cooperation and Peer Support)
- **Clarifying** lessons (Success Seems Feasible)
- **Challenging** students (Press for Effort, Perseverance and Rigor)
- **Consolidating** knowledge (Ideas get Connected and Integrated)[48]

Previous administrations of the survey have demonstrated a direct relationship between student-assigned ratings in these categories and greater student achievement. According to the Tripod Project's past research, "classrooms in which students rated their teachers higher on the Seven Cs tended also to produce greater average achievement gains."[49] In a MET Project study of teacher evaluations and ultimate effectiveness, the report found that "combining classroom observations with student feedback and student achievement gains on state tests" was more effective than a master's degree and/or years of experience in "predicting which teachers would have large gains [in student achievement]" when placed with a new group of students.[50]

The study also found that **student feedback provides a more reliable indication of teacher effectiveness than do classroom observations** because such data "includes many more perspectives based on many more hours in the classroom."[51] However, such feedback was found to be **less indicative of a teacher's "achievement gains with other students"** than were value-added student achievement gains. Combining student feedback with classroom observations and value-added student achievement gains was determined to be the most effective manner in which to evaluate teacher effectiveness, since integrating the three approaches capitalizes on the strengths of all three and helps to offset the weaknesses of any one approach.[52]

School districts may choose to use simple questionnaires to collect feedback from students at different levels of the K-12 grade span. As an example, Figure 2.1 presents the survey instruments used in the Davis School District's Educator Assessment system. The surveys are administered at the early education (K-2), elementary (3-6), and secondary grade levels and are used to supplement teacher evaluations. Respondents rate their teachers on a simple scale of: "no," "sometimes," or "yes" (coded as ☹, 😐, or ☺ for students in grades K-2).[53]

[48] Bulleted points taken verbatim from: *Ibid.*
[49] *Ibid.*
[50] "Gathering Feedback." *Op. cit.*, p. 29.
[51] *Ibid.*
[52] *Ibid.*
[53] "Educator Assessment System (EAS): Acknowledging and Honoring Quality Performance." Davis School District. P. 39. http://www.nctq.org/docs/72-08.pdf

Figure 2.1: Davis School District, Educator Assessment System, Student Questionnaires

K-2 (Non-reader) Student Survey	Elementary Student Survey (3-6)
• My teacher shows me how to do new things. • My class is a good place for learning. • I like to come to this class. • My teacher is a good teacher. • My teacher is nice to me. • My teacher's rules are fair. • I know what I am supposed to do in this class.	• I learn new things in this class. • My class is a good place for learning. • I like to come to this class. • My teacher is a good teacher. • I know what I am supposed to do in this class. • I understand the rules in my class. • My teacher treats me fairly. • I know how well I am learning in this class. • My teacher is nice to me.
Secondary Student Survey	
• I learn new things in this class. • My class is a good place for learning. • This teacher treats me with care and respect. • This is a good teacher. • I know what I am supposed to do in this class.	• I understand the class rules. • This teacher treats me fairly. • I know how well I am doing in this class. • This teacher maintains class discipline. • I usually understand how to do my assignments.

Source: Davis School District Educator Assessment System (EAS)

Parent Perception Data

Parents are an important audience within the school system, and past research has shown that parents can make important contributions in the evaluation of teachers.[54] Parental input, however, generally does not play a central role in teacher evaluations, as parent surveys can only offer an indirect view of teacher performance.[55] Many evaluation systems offer teachers a certain degree of control by featuring *optional* parent surveys and by allowing teachers to choose whether to share the results after they have had the opportunity to review the responses themselves.[56] Parent reports are often used on a limited basis and for largely formative purposes. **They should be used with a degree of caution, as high parent ratings do not necessarily translate to quality instruction.**[57] However, high parent ratings in conjunction with other positive data sources may be an indicator of quality teaching.

One state department of education offers in its evaluation handbook the following set of **six best practices for conducting parent surveys**: [58]

- ❖ Use global items as the central datum for evaluation decisions

[54] Peterson. *Op. cit.*, p. 172.
[55] Goe, *et al. Op. cit.*
[56] Peterson, K., *et al.* 2003. "Parent Surveys for Teacher Evaluation." *Journal of Personnel Evaluation in Education*, 17:4. Pp. 317-330.
[57] *Ibid.*
[58] Bullet list taken verbatim from: "Evaluation Handbook for Professional Alaska (HB 465) Educators." Alaska Department of Education. http://www.eed.state.ak.us/evaluationhandbook.pdf

- Include questions which elicit information about how involved parents have been with the school
- Establish and publicize minimum return rate expectations
- Take into account a number of factors in analyzing the results, such as age of pupil and differences in parent populations.
- Help teachers interpret the information
- Vary the frequency of parent surveys

As an example of a parent questionnaire, Figure 2.2 presents the 13 questions administered to parents of Davis School District students. Parents respond to each survey item with: "no opinion," "no," "sometimes," or "yes."[59]

Figure 2.2: Davis School District, Educator Assessment System, Parent Survey Questionnaire

Question
My son/daughter is learning in this class.
This classroom is a good place for learning.
This teacher treats my son/daughter with care and respect.
I am satisfied with my son's/daughter's experience in this class.
The learning activities in this class are appropriate for my son/daughter.
My son/daughter knows what is expected in this class.
This teacher treats my son/daughter fairly.
This teacher is accessible.
Homework in this class helps my son/daughter learn.
I have reviewed the class content and goals for this class.
When I contact this teacher I get a timely response.
The Student Information System (online) is updated in a timely fashion.
This teacher communicates with me as a parent.

Source: Davis School District Educator Assessment System (EAS)

A study focused specifically on the parent survey facet of the Davis Public Schools EAS made several recommendations for the improvement of the system that should be noted in conjunction with the preceding sample questionnaire. First, statements such as "My child is learning" may not be suitable for parents who are biased observers or ill-equipped to answer the question accurately.[60] Second, the study recommended the inclusion of some "popularity" items, such as whether the student was "treated as an individual," despite these questions previously being considered too superficial. The "factor analysis of themes that underlay the literal item reports revealed an unexpectedly strong sentiment on the part of parents for humane personal treatment of pupils."[61] Third, the authors suggested a move to a **five-point scale in order to offer more room for differentiation** among teacher's abilities.[62]

[59] "Educator Assessment System (EAS): Acknowledging and Honoring Quality Performance." Davis School District. p. 40. http://www.nctq.org/docs/72-08.pdf
[60] Peterson, K., *et al.* 2003. "Parent Surveys for Teacher Evaluation." *Op. cit.*, p. 327.
[61] *Ibid.*, pp. 327-328.
[62] *Ibid.*, p. 328.

Fourth, the report recommended the use of global items, such as "overall satisfaction with this teacher," a suggestion corroborated by various other researchers, as well.[63] Finally, the researchers suggested the addition of a question that would pertain to "teacher attention to culture, language, and physical condition among teacher, classroom, students, and parents."[64]

Documentation of Professional Activity

Documentation of professional activity has the potential to take many forms. **In practice, one of the most common means entails the creation and review of a teacher portfolio.** Teacher portfolios provide key evidence about a teacher's work and may contain a number of elements, such as lesson plans, instructional materials, samples of students' work, and comments on students' work. While documentation of professional activity and the use of portfolios has widespread potential for application in teacher evaluations, the practice has a relatively small research base. The practice may take a role in both formative and summative performance assessments.[65]

Portfolios offer evaluators an opportunity to **examine aspects of instruction that are not readily observable in the classroom** and may be used with teachers in all fields. Evaluators and teachers are both likely to view them as a fair and largely authoritative reflection of teachers' skills, as well as a useful tool for self-reflection on performance. Despite these advantages, portfolios are time-consuming for teachers and scorers alike. Portfolio work is difficult to assess and standardize and, because portfolio samples reflect a teacher's best work, may not be representative of daily practice. Some research on the validity and reliability of portfolios has raised concerns about consistency and stability in scoring, as well as a lack of research linking portfolios to student achievement.[66] Portfolios and similar means of evaluation have been implemented and included in teacher evaluation procedures through various approaches in individual school settings, though some standard rubrics have also been produced.

Teacher Evaluation in Subjects Not Traditionally Evaluated

Due to No Child Left Behind and Race to the Top, teacher evaluation has been closely tied to student achievement, but with most standardized exams focused on the two core areas of mathematics and language arts, other subjects have been neglected. Many educational leaders note that teacher evaluation instruments and reward structures address only the "tested" subjects and thereby exclude the majority of teachers. Indeed, some studies suggest that **the proportion of teachers not**

[63] *Ibid.*, p. 327. And: "Evaluation Handbook for Professional Alaska (HB 465) Educators." *Op. cit.*, p. 106.
[64] "Evaluation Handbook for Professional Alaska (HB 465) Educators." *Op. cit.*
[65] "Teacher Evaluation: A Conceptual Framework and Examples of Country Practices. 2009." OECD. *Op. cit.*
[66] Goe, *et al. Op. Cit.*

eligible for participation in such evaluation systems could be as high as 69 percent.[67]

To date, there has been relatively little research conducted in the area of achievement-based teacher evaluations for non-tested subjects and grades. Indeed, recent research briefs commissioned by the National Center for the Improvement of Education Assessment and the National Comprehensive Center for Teacher Quality indicate that **there are currently "no research-based models" for incorporating student learning growth into teacher evaluations in non-tested areas**, and that states and districts have been left to experiment with a variety of strategies—many of which are imperfect or incomplete at the time of implementation.[68]

There are **three primary categories of data** that can be collected by schools for use in evaluations for teachers in non-tested areas: school/district level data, subject-level standardized exam data, and data from other standardized subject-level measures.[69] Evaluation models based on these data vary widely in cost and complexity. No one model is appropriate or feasible for all districts or states. Many states and districts have adopted classroom-level assessments such as portfolios, projects, or Student Learning Objectives (SLOs) to measure student performance growth in non-tested areas. While this approach is popular due, in large part, to the fact that teachers are already using many of these measures in their classrooms, many educators and researchers have raised concerns regarding the validity and comparability of these measures. To address these concerns, states and districts have created extensive rubrics for standardizing these assessments—a process which can be both costly and time consuming.[70] Below, we review different types of measures that can be used to evaluate teachers teaching non-tested subject areas:

- **Pro-Rated School/District-Wide Value-Added Model:** A related but slightly different approach to the school/district-wide value added model is to "pro-rate" student achievement growth **based on a teacher's actual contribution to student learning in a "tested" subject area**. One such model is Battelle for Kids, which offers a web-based solution called BFK-Link which assists teachers and administrators with this kind of pro-rated scoring.

---

[67] Prince, C. *et al.* "The Other 69 Percent: Fairly Rewarding the Performance of Teachers of Nontested Subjects and Grades." The Center for Educator Compensation Reform. P. 4. http://cecr.ed.gov/guides/other69Percent.pdf

[68] "Measuring Teachers' Contributions to Student Learning Growth for Nontested Grades and Subjects." 2011. The National Comprehensive Center for Teacher Quality. Research and Policy Brief. P. 1. http://www.tqsource.org/publications/MeasuringTeachersContributions.pdf

[69] Marion, S. and K. Buckley. 2011. "Approaches and Considerations for Incorporating Student Performance Results from "Non-Tested" Grades and Subjects into Educator Effectiveness Determinations." The National Center for the Improvement of Education Assessment and Harvard University. http://www.nciea.org/publications/Considerations%20for%20non-tested%20grades_SMKB2011.pdf
"Measuring Teachers' Contributions to Student Learning Growth for Nontested Grades and Subjects." *Op. cit.*
"Measuring Student Growth for Teachers in Non-Tested Grades and Subjects: A Primer." The Southwest Comprehensive Center at WestEd. http://www.swcompcenter.org/educator_effectiveness2/NTS__PRIMER_FINAL.pdf

[70] *Ibid.*

The system allows teachers to "review and correct data used for teacher-level measures of effectiveness, including value-added analysis, by ensuring that all students taught are 'claimed' by teachers for all subjects, accounting for student mobility and shared instruction/co-teaching."[71]

- **New Pre- and Post-Tests for Non-Tested Subjects:** Due, in part, to the problems associated with basing all teacher evaluations on student achievement data for only tested subject areas, some states and districts have instead opted to create or select new standardized exams for all grades and subject areas. For example, Delaware is in the process of developing such exams for use in its schools as part of its Race to the Top initiative. Similarly, the Hillsborough County School District in Florida has created over 500 pre- and post-tests for 429 different courses not tested by the Florida Comprehensive Assessment Test (FCAT).[72]

- **Curriculum-Based End-of-Course Exams:** Many curriculum designers include standardized end-of-course (EOC) exams designed to measure student learning as part of their curriculum packages. Some districts and schools have considered using these existing EOC exams as measures to assess student learning growth by administering these exams as pre-tests at the beginning of the year and post-tests at the end of the learning period.[73]

- **Interim Assessments:** Similar to curriculum-based EOC exams, interim assessments can be included in curriculum or intervention packages, and are generally administered at specified intervals throughout the school year "to evaluate student knowledge and skills relative to a specific set of academic standards."[74] Like pre- and post-tests, these assessments can also be developed at the school or district level, provided that these educational systems possess the necessary resources for development and maintenance.

- **Locally-Created Assessments:** Many states and districts are considering the use of "locally-created assessments" such as portfolios, performances, products, or projects (the "Four Ps") as reliable measures of student learning growth. In order to utilize locally-created assessments effectively, teachers and administrators must first agree on a "pre-test," administered at the beginning of the course, which is used to assess students' existing knowledge base. Students are then asked to perform the same or similar task at the end of the

[71] "Measuring Teachers' Contributions to Student Learning Growth for Nontested Grades and Subjects." *Op. cit.,* P. 6.
[72] Buckley and Marion. *Op. cit.*, p. 14.
[73] "Measuring Teachers' Contributions to Student Learning Growth for Nontested Grades and Subjects." *Op. cit.*, p. 8.
[74] "Measuring Student Growth for Teachers in Non-Tested Grades and Subjects: A Primer." *Op. cit.*, p. 8.

course in order to demonstrate the knowledge they have gained.[75]

- **Student Learning Objectives:** The Race to the Top Technical Assistance Network defines Student Learning Objectives (SLOs) as "a participatory method of setting measurable goals, or objectives, based on the specific assignment or class, such as the students taught, the subject matter taught, the baseline performance of the students, and the measurable gain in student performance during the course of instruction."[76] SLOs are similar in nature to locally-created assessments; indeed, any of the Four Ps can be used as assessment tools within an SLO framework. The key difference between the two assessment methods is that SLOs are typically developed at the classroom level, and are designed with the needs and capabilities of students within a specific learning environment in mind.

[75] "Measuring Teachers' Contributions to Student Learning Growth for Nontested Grades and Subjects." *Op. cit.*, p. 10.

[76] "Measuring Student Growth for Teachers in Non-Tested Grades and Subjects: A Primer." *Op. Cit.*, p. 1.

Section Three: Models and Case Studies

The previous section presented an overview of the use of multiple measures in teacher evaluations, though the measures were discussed independently of one another. This section builds on that discussion to consider how multiple measures of teacher performance can be combined to allow for more comprehensive evaluations. Rather than draw on research studies and scholarly literature, as the previous section did, this section turns to real-world examples of how districts have integrated multiple measures into their teacher evaluations. We first discuss one model that has enjoyed popularity among school districts—the 360 degree feedback model for evaluation. We then turn our attention to four teacher evaluation case studies:

- Cincinnati Public Schools
- Memphis City Schools
- Austin Independent School District
- Charlotte-Mecklenburg Schools

360-Degree Evaluations

One major, but perhaps understated focus of this report has been on how to integrate perceptual data from various stakeholders into teacher evaluations. One strategy to achieve this end is the 360-degree evaluation model, an approach that has been favored by the private sector and that has begun to see greater adoption by school districts. **In the 360-degree model, different groups that have an interest in teacher quality—"stakeholders"—report their perceptions of an educator's performance, whether through a survey, questionnaire, interview, or focus group.** The collection of stakeholder perception data is also known as a "multiple-source feedback system;" under this definition, data that summarize multiple viewpoints are collected in order to provide a "detailed and accurate picture of individual performance."[77] This approach allows for the inclusion of "many perspectives and viewpoints on the actions" of an individual educator, but also protects the anonymity of contributors.[78]

360-degree assessments, or "evaluations based on input from a full circle of appraisers," rely on evaluations from an educator's students, peers, parents, staff, community, and supervisors to provide the most complete picture of performance possible.[79] **The possible functions of this type of evaluation are: (1) "for developmental purposes (for the employee's eyes only), (2) for appraisal, and (3) for compensation."**[80]

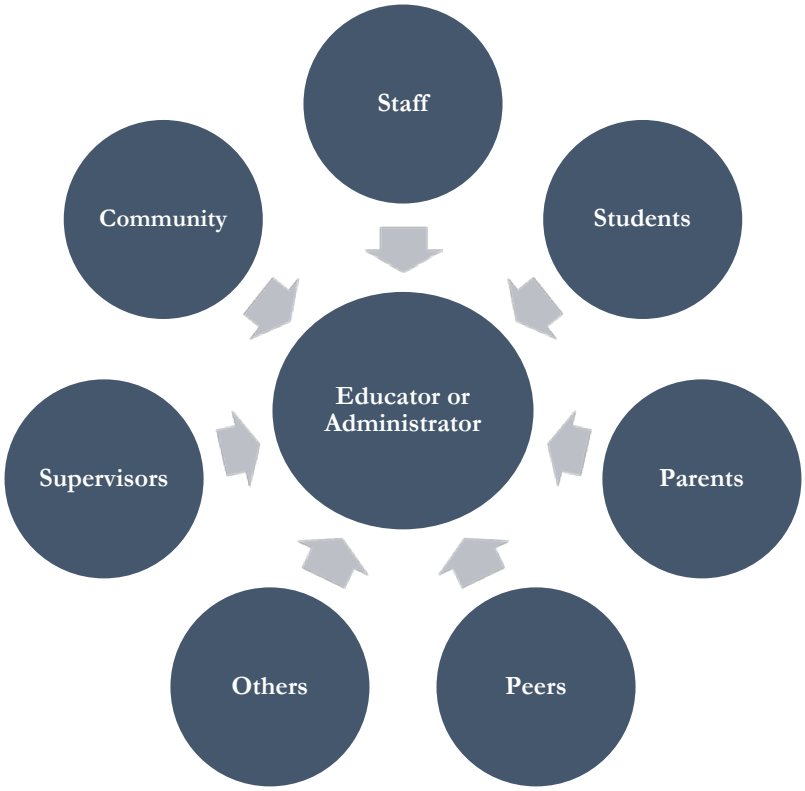
---

[77] "School Leaders." Ecra Group: Research and Analytic Solutions. http://www.ecragroup.com/360-school-leaders
[78] *Ibid.*
[79] Wilkerson, *et al.* 2000. *Op. Cit.*, p. 181.
[80] Bulleted list taken verbatim from: Manatt, R. 2000. "Feedback at 360 Degrees." *The School Administrator.* http://www.aasa.org/SchoolAdministratorArticle.aspx?id=14530&terms=360

According to one study of teachers' perceptions of both traditional and 360 degree feedback processes in a suburban New York school district, **only 29.6 percent of participants identified traditional evaluations as contributing to student achievement outcomes, compared to 66.7 percent of participants who "found the 360-degree feedback process to be more focused on student achievement.**"[81] Furthermore, teachers participating in the study found the multisource evaluation to be better at assisting them in identifying professional development needs.[82]

The following figure demonstrates how multiple sources can contribute to form a composite evaluation for one administrator or educator.

Figure 3.1: Perceptual Feedback in K-12 Public Schools



In an effort to determine how such feedback compares with other methods, one study examined a school district that partially implemented the 360 degree evaluation model by soliciting feedback from students, principals, and teachers (self-ratings). The researchers found "**high positive correlations between student feedback of**

[81] Mahar, J. and B. Strober. 2010. "The Use of 360-Degree Feedback Compared to Traditional Evaluative Feedback for the Professional Growth of Teachers in K-12 Education." *Planning and Changing*, 41:3/4. P. 152, 156.
[82] *Ibid.*

teacher performance and student achievement in all three core subject areas (math, reading, and language arts)."[83] Teachers' self-ratings were also highly to slightly positively correlated with student achievement scores, though principal ratings were only slightly correlated with test scores. Student feedback was the strongest predictor of achievement as evidenced by test results.[84]

Case Study #1: Cincinnati Public Schools

Multiple-Measure System with an Observational Focus

Cincinnati Public Schools (CPS) uses a system for teacher evaluation that has been widely regarded in the literature. The CPS evaluation system involves multiple classroom observations and detailed written feedback for teachers. Studies have found that these evaluations have led to marked improvements in the performance of mid-career teachers, as reflected in enduring improvements in student achievement. According to one researcher who studied the CPS framework, **substantial correlations were found between differences in student achievement and teacher performance when combined across grades and subjects.** The study author concluded that "a rigorous teacher evaluation system" such as that used by CPS "can be substantially related to student achievement and provide criterion-related validity evidence for the use of the performance evaluation scores as the basis for a performance-based pay system or other decisions with consequences for teachers."[85]

CPS also appears to have managed the implementation of the evaluation system in an exemplary manner, taking multiple stakeholder views into consideration and testing and evaluating the system rigorously over time. CPS' Teacher Evaluation System (TES) was developed from a framework outlined in Charlotte Danielson's publication, *Enhancing Professional Practice: A Framework for Teaching*. Danielson's framework divides skills and responsibilities into four domains:

- Planning and Preparing for Student Learning
- Creating an Environment for Student Learning
- Teaching for Student Learning
- Professionalism[86]

CPS's current system began in 1997 when the teachers' union contract called for the implementation of a new teacher evaluation system and the exploration of a new compensation system. To achieve these goals, the Teacher Evaluation

[83] *Ibid.*

[84] Wilkerson, D.J., R. Manatt, M.A. Rogers, R. Maughn. 2000. "Validation of Student, Principal, and Self-Ratings in 360 Degree Feedback for Teacher Evaluation." *Journal of Personnel Evaluation in Education* 14:2. Pp. 179-192.

[85] Milanowski, A. 2004. "The Relationship Between Teacher Performance Evaluation Scores and Student Achievement: Evidence From Cincinnati." *Peabody Journal of Education* 79:4. Pp. 33-53. http://www.tandfonline.com/doi/abs/10.1207/s15327930pje7904_3

[86] "Teacher Evaluation." Cincinnati Public Schools. http://www.cps-k12.org/employment/tchreval/tchreval.htm

Committee, the Skills and Knowledge Compensation Committee, and the Local Professional Development Committee worked together with an umbrella "Committee of the Whole" to define goals for quality education within the CPS system. **The umbrella committee studied several potential frameworks, ultimately deciding on the one prescribed in Charlotte Danielson's book.** The committee then went about establishing standards for good teaching and defining the skills that high-quality teachers should possess. Finally, the Teacher Evaluation Committee designed an evaluation system using the defined standards, as well as practical determinations regarding the appropriate frequency of evaluations, standards, rubrics, and field testing.

After approval by the umbrella committee, **the Teacher Evaluation System was field tested in ten schools during the 1999-2000 school year.** After being refined by the district, the Consortium for Policy Research in Education (CPRE) evaluated the field test and baseline data were established for rubric scores. Additional revisions were made on a yearly basis during the first few years, taking into account information in the CPRE evaluation and recommendations from teachers and principals throughout the district.[87] In 2005, the district modified specific elements to make it more efficient and responsive to teachers' needs. These modifications included:

- Providing immediate feedback
- Streamlining assessments
- Reducing the number of classroom observations
- Providing conferences immediately after observations,
- Developing tools such as checklists and evidence packets for assessments
- Providing a wide variety of professional opportunities.[88]

Within the CPS system, **Comprehensive Evaluations** occur during a teacher's first year as a new hire, his or her fourth year, and every five years thereafter. A Comprehensive Evaluation consists of the following components:

- An orientation meeting for the teacher to learn about the evaluation process
- A readiness conference with the evaluator to share details about the teaching assignment
- Completion of at least four classroom observation sessions

Each year that they are *not* scheduled for a Comprehensive Evaluation, teachers undergo an **Annual Assessment**—one classroom observation conducted by their administrator.

[87] "History of the Teacher Evaluation System." Cincinnati Public Schools. http://www.cps-k12.org/employment/tchreval/TESHistory.htm

[88] "CPS' Evaluation System Strengthened." 2005. CPS News Release. May 19, 2005. http://www.cps-k12.org/employment/tchreval/TESstrengthened.pdf

In CPS's evaluation system, good teaching is further defined within each of the four domains outlined in Danielson's framework. A total of 16 standards have been developed and form the basis for the Comprehensive Evaluation. The district has also designed rubrics (scoring guides) that establish standards for good teaching and delineate clear expectations for performance and professional development in the district.

With the support of the teachers' union, the district created the **Peer Assistance and Evaluation Program (PAEP)** to help teachers with instructional deficiencies and to support and retain more experienced and expert teachers. The program helps teachers refine their instructional skills and provides an orientation to the district, including its goals, curriculum, and organization. Through this component, each teacher is evaluated and assisted by a consulting teacher.

An **Intervention Component** works with experienced teachers who exhibit serious instructional deficiencies. Teachers may be referred to this program component when a principal has concerns about a teacher's performance or failure to meet expected performance standards. Consulting teachers work with those teachers referred to the program to improve their instructional skills. When improvement does not occur, a panel of teachers and administrators may recommend a second year of intervention or the non-renewal of the teacher's contract.

Finally, a **Career-In-Teaching Program** was developed to provide incentives to attract and retain quality teachers in the profession, to improve and encourage professional growth opportunities, and to give teachers broader roles and responsibilities for the improvement of student achievement. Through five levels, teachers work to achieve lead teacher status:

- **Level One, Apprentice**: An apprentice is a teacher without previous teaching experience. This level prepares teachers to pursue a career in teaching.
- **Level Two, Novice:** A novice is a teacher who has met licensure requirements and is developing the skills required for a teaching career.
- **Level Three, Career**: A career teacher has demonstrated the skills needed to have a career in teaching.
- **Level Four, Advanced**: An advanced teacher is continuing to master the art of teaching, demonstrating a distinguished level of teaching.
- **Level Five, Accomplished**: An accomplished teacher is a teacher who has demonstrated outstanding teaching.
- **Lead Teacher**: Lead teachers support quality instruction, demonstrate leadership, effective communication skills, a consistent pattern of professional growth, cooperation and collaboration, and a commitment to teaching. They may serve at both the school and district level as consulting teachers, teacher

evaluators, curriculum specialists, subject or team leaders, or program facilitators.[89]

Case Study #2: Memphis City Schools

Integration of Achievement Data, Observations, Stakeholder Perceptions, and Teacher Knowledge

Funded by the Bill and Melinda Gates Foundation, Memphis City Schools' Teacher Effectiveness Initiative is a $90 million project to "improve teacher effectiveness in order to empower teachers for student success."[90] The Initiative has established a Teacher Effectiveness Measure (TEM) that provides insight into how stakeholder perception data is factored into larger evaluations of teacher effectiveness. The TEM is comprised of four main components, which are illustrated in Figure 3.2 below.

Figure 3.2: Components of Memphis City Schools' Teacher Effectiveness Measure (TEM)

Component		Percentage	Description
Student Growth & Achievement	**Student Growth**	**35%**	TEI uses a value-added assessment system that measures "whether a teacher's class has achieved less than one year, one year, or more than one year of academic growth" per academic year of instruction.
	Student Achievement	**15%**	TEI calculates student achievement according to "one of several possible measures" chosen by a teacher and or his principal from a selection provided by the state of Tennessee.
Observation of Practice		**40%**	This component will be scored based on frameworks from the Memphis City Schools Teaching & Learning Framework.
Stakeholder Perceptions		**5%**	Memphis City Schools uses the TRIPOD survey to gather insights about students' classroom experiences. According to TEI, "questions focus on specific observable teaching practices to limit the impact of any subjective student opinions."

[89] "Teacher Evaluation." Cincinnati Public Schools. http://www.cps-k12.org/employment/tchreval/tchreval.htm

[90] "$90 Million Grant to Fund Teacher Effectiveness Initiative: Empowering Teachers for Student Success." Bill & Melinda Gates Foundation. http://www.gatesfoundation.org/press-releases/Pages/memphis-city-schools-intensive-partnership-grants-091119.aspx

Component	Percentage	Description
Teacher Knowledge	**5%**	Teacher knowledge is measured by TEI using one of several measures, which include professional development experience, portfolios, National Board Certification, and observation by a "content-area specialist."[91]

Source: "The Teacher Effectiveness Measure Manual, 2011-2012." Memphis City Schools.

Case Study #3: Austin Independent School District

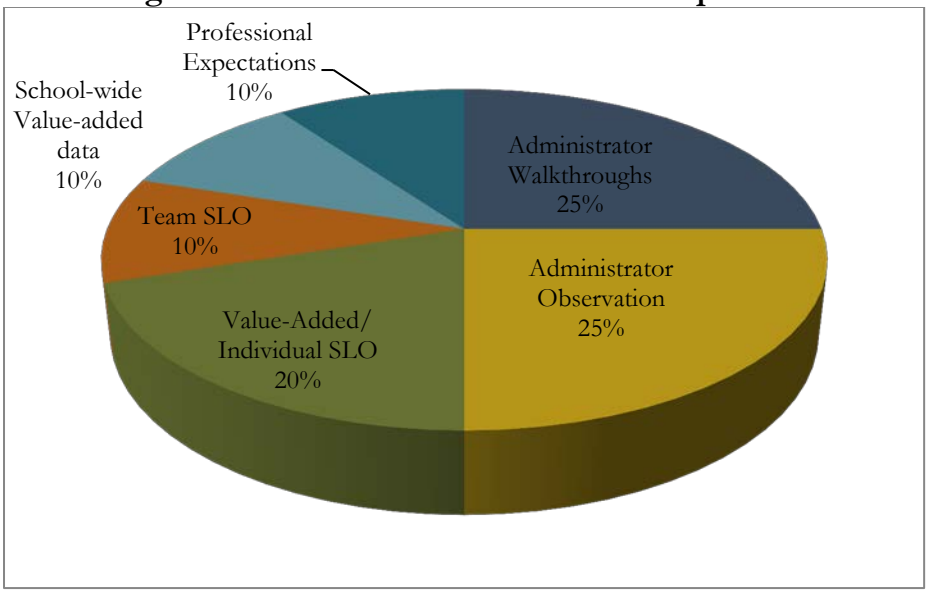
Measures for Teachers of Non-Tested Subjects

Austin Independent School District (AISD) has been working to implement a multiple measure evaluation system that includes value-added components—including for subjects not traditionally tested—as well as a substantial emphasis on observations of teacher performance. During the 2011-2012 school year, the Austin Independent School District is piloting its REACH Teacher and Principal Evaluation System, which **"provides a significant weight to student growth" but focuses heavily on the three domains "of instructional practice, classroom climate, and professional expectations."**[92] Under REACH, teacher evaluations are comprised of six different components, including administrator walkthroughs, administrator evaluations, professional expectations, individual and team Student Learning Objectives (SLOs), and value-added data. The specific weights of each of these components are detailed in Figure 3.3.

[91] "The Teacher Effectiveness Measure Manual, 2011-2012." Teacher Effectiveness Initiative. 2011-2012. P. 10. http://www.nctq.org/docs/Memphis_Evaluation_handbook_2011-12.pdf

[92] "Teacher Evaluation System: AISD REACH 2011-2012." The Austin Independent School District. P. 4. http://archive.austinisd.org/inside/initiatives/compensation/docs/SCI_Teacher_Evaluation_System_Handbook.pdf

Figure 3.3: Breakdown of REACH Components



Source: Austin Independent School District, 2011[93]

For teachers of subjects assessed by the Texas STAAR exams (grades 4-9 math and reading; grades 5, 8, and 9 science; and grades 6, 8, and 9 social studies), **20 percent of the evaluation is based on student value-added growth data**. For teachers of all other subjects, 20 percent of the evaluation **is based on an individual SLO developed in collaboration and consultation with colleagues and district administrators**.[94] Additionally, 10 percent of any teacher evaluation is based on **team SLOs,** which are developed in collaboration and consultation with colleagues and administrators. Finally, *all* district teachers are evaluated on school-wide value added data on the Texas STAAR exam.

The Austin Independent School District provides an SLO development guide to all district teachers and administrators. **This guide assists users in creating appropriate SLOs by prompting them to answer a series of questions regarding the goals, target population, and outcomes associated with each potential objective.** Additionally, the district provides teachers with an SLO Rigor Rubric designed to help teachers self-evaluate the validity of their chosen objectives. To develop an SLO, teachers set a target of student growth at the start of the school year. Targets are based on a review of data on students' skills and are set and approved after collaboration and consultation with colleagues and administrators.[95]

Administrator observations and administrator walkthroughs each compose 25 percent of a teacher's evaluation and are intended to measure performance in the

[93] "Teacher Evaluation Overview." The Austin Independent School District. http://archive.austinisd.org/inside/initiatives/compensation/docs/SCI_Teacher_Evaluation_Overview.pdf
[94] "Teacher Evaluation System: AISD REACH 2011-2012." *Op. Cit.*, P. 6.
[95] *Ibid.*

AISD's Instructional Practice and Classroom Climate rubrics. One observation is performed per year; these are announced and last 45 minutes. Three 15-minute walkthroughs are also performed. Both seek to measure instructional practice and classroom climate.

Instructional practice (25 percent) includes the following elements:

- Actively engages students during instructional activities
- Checks for student understanding and modifies instruction to address student misconceptions
- Differentiates instruction for student needs utilizing a variety of instructional strategies
- Develops problem-solving and critical thinking skills for all students
- Sets rigorous academic expectations for students
- Collects, tracks, and uses student data to develop lesson plans and assessments
- Provides relevant and useful feedback to students
- Designs effective objective driven lessons and assessments that reflect the standards

Classroom climate (25 percent) consists of the following elements:

- Sets and implements classroom routines and procedures that support student learning
- Establishes and maintains standards for student behavior
- Creates a safe and secure classroom environment that is organized and engages students
- Establishes a climate that promotes fairness, respect, and diversity
- Provides responsive communication to parents throughout the year

Finally, evaluations include professional expectations (10 percent). Observation in this area is ongoing throughout the year and considers whether the teacher:

- Establishes professional goals, participates in professional development, and applies learning to practice
- Engages in meaningful collaboration to attain school goals and a positive campus climate
- Complies with district and school policies and procedures
- Fulfills professional responsibilities while modeling professional integrity[96]

[96] "Teacher Evaluation System: AISD REACH 2011-2012." *Op. cit.*, pp. 6-8

Case Study #4: Charlotte-Mecklenburg Schools

Use of a Value-Added Model and SLOs in Performance-Based Compensation

Using funds awarded by the U.S. Department of Education's TIF grant, Charlotte-Mecklenburg Schools (CMS) implemented a new, performance-based compensation system—Leadership for Educators' Advanced Performance (LEAP)—during the 2007-2008 school year. During the 2011-2012 school year, CMS is revising the LEAP initiative by piloting two separate compensation packages based on Student Learning Objectives (SLOs) and value-added growth measures. The specific goals of these new initiatives are to:[97]

- Build teacher and principal capacity to increase student achievement by aligning and improving district systems in support of the schools;
- Create a compensation system for teachers and principals that provides differentiated levels of compensation based on student achievement gains and teacher/principal evaluations that include multiple classroom observations;
- Support the recruitment and retention of qualified teachers and principals in hard-to-staff schools and subjects; and
- Develop district capacity to implement, scale-up, evaluate, and sustain a performance-based compensation system, with measurable impact on student achievement.

Only instructors who teach End-of-Course/End-of-Grade tested courses (grades 4-8 reading and math, grades 5 and 8 science, Algebra I, Algebra II, Biology, English I, Physical Science, Civics and Economics, and U.S. History) are eligible to participate in the value-added compensation program.[98] However, **all certified teachers who meet the following requirements are eligible to participate in the SLO compensation initiative**:[99]

- Are certified, lateral entry, Teach for America, or Visiting International Faculty (VIF) teachers;
- Administrators and teachers agree s/he can complete all SLO components;
- Earn "Proficient" or above on each overall standard of the *NC Teacher Evaluation Process* summary evaluation;
- Are in attendance at least 85% of the interval specified on the SLO as verified by the SLO Attainment document;
- Attend required training in the SLO process authorized by TIF-LEAP staff;
- Submit SLO(s) and have the SLO(s) approved by an administrator prior to the deadlines for submission; and

[97] "Teacher Incentive Fund – Leadership for Educators' Advanced Performance." Charlotte-Mecklenburg Schools. http://www.cms.k12.nc.us/cmsdepartments/Tif-Leap/Pages/default.aspx

[98] "Eligibility for Value-Added Participation." Charlotte-Mecklenburg Schools. http://www.cms.k12.nc.us/cmsdepartments/Tif-Leap/Pages/EligibilityforValue-AddedParticipation.aspx

[99] *Ibid.*

- Provide data or portfolio materials to demonstrate that at least 75% of students (or higher if so stated) specified in their SLO have achieved, or exceeded the growth expectations.[100]

[100] "Student Learning Objectives." Charlotte-Mecklenburg Schools. http://www.cms.k12.nc.us/cmsdepartments/Tif-Leap/Pages/StudentLearningObjectives.aspx

Project Evaluation Form

Hanover Research is committed to providing a work product that meets or exceeds member expectations. In keeping with that goal, we would like to hear your opinions regarding our reports. Feedback is critically important and serves as the strongest mechanism by which we tailor our research to your organization. When you have had a chance to evaluate this report, please take a moment to fill out the following questionnaire.

http://www.hanoverresearch.com/evaluation/index.php

Caveat

The publisher and authors have used their best efforts in preparing this brief. The publisher and authors make no representations or warranties with respect to the accuracy or completeness of the contents of this brief and specifically disclaim any implied warranties of fitness for a particular purpose. There are no warranties which extend beyond the descriptions contained in this paragraph. No warranty may be created or extended by representatives of Hanover Research or its marketing materials. The accuracy and completeness of the information provided herein and the opinions stated herein are not guaranteed or warranted to produce any particular results, and the advice and strategies contained herein may not be suitable for every member. Neither the publisher nor the authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages. Moreover, Hanover Research is not engaged in rendering legal, accounting, or other professional services. Members requiring such services are advised to consult an appropriate professional.