Using Multiple Forms of Data in Principal Evaluations

An Overview With Examples

Integrated Leadership Development Initiative

WestEd⊕
WestEd.org

The Integrated Leadership Development Initiative (ILDI) is a cross-agency partnership that focuses on collaboratively guiding and supporting leader development and improving conditions of leadership so that there are highly accomplished leaders in every district and school in California. ILDI members include the California Department of Education, the California County Superintendents Educational Services Association, the Association of California School Administrators, the California Commission on Teacher Credentialing, the California Association of Professors of Educational Administration, the Center for the Future of Teaching and Learning, the California Comprehensive Center at WestEd, and the Regional Educational Laboratory West at WestEd.

Nancy M. Sanders, PhD, teaches and conducts state-level policy research about standards, accountability, and professionalization in educational leadership. Sanders directed the coalition of states in developing the national Educational Leadership Policy Standards: ISLLC 2008 (CSSSO, 2008) and writing the ISLLC 2008-Based Performance Expectations and Indicators for Education Leaders (CSSSO, 2008) to guide leadership policies and programs.

Karen Kearney, MA, is the Director of the Leadership Initiative at WestEd, formerly the California School Leadership Academy. Kearney directs all aspects of the initiative, which focuses on the role of principals, superintendents, and teacher leaders in comprehensive school restructuring.

Scott Vince, MPPA, is a Research Associate with WestEd's Comprehensive School Assistance Program. Vince specializes in research relevant to school turnaround, educator evaluation, and labor-management collaboration.

Project Director, Karen Kearney, Leadership Initiative at WestEd.

Editing and Design, WestEd.

This resource can be downloaded free of charge at http://www.schoolsmovingup.net/effectiveprincipals

WestEd, 730 Harrison Street, San Francisco, CA 94107-1242

Using Multiple Forms of Data in Principal Evaluations

An Overview With Examples

Nancy Sanders

Karen Kearney

Scott Vince

Table of Contents

Using Multiple Forms of Data in Principal Evaluations

Principal evaluations should provide information about the most important aspects of principal performance — what principals do; how they do it; and the impact they have on their schools, teaching, and student learning. Expectations for what constitutes principal effectiveness are extensive, and thus make assessing principals complicated. As a guide for districts to improve their principal evaluation systems, *Key Features of a Comprehensive Principal Evaluation System* (2012) presented 13 elements that are critical in establishing research-based, valid, and reliable principal evaluation systems. Key Feature 8 highlights *using multiple forms of data* and is guided by the question: *Does the evaluation system use multiple forms and sources of data (evidence) that include input from a range of people who work with the principal?* Note that the full document, *Key Features of a Comprehensive Principal Evaluation System* (2012), is available online at http://www.schoolsmovingup.net/effectiveprincipals

Recent developments in measures of principal knowledge and performance and of principal evaluation models emphasize the need to collect and use multiple forms of data to capture the scope and complexity of new expectations. This brief explains basic concepts about multiple forms of data that apply to principal evaluation systems. It then provides examples for using multiple forms of data that illustrate a wide range of options used by states, districts, and organizations. The examples demonstrate various choices and decisions that use multiple data sources within context-specific purposes and resources.

Please note that the examples in this brief help to explain concepts involved in using more than one form of data. These examples are not intended as recommendations. The examples are described based on the time of this writing, and their specific details may change as they are implemented and refined. To confirm information, please check the Web sites provided at the end of each example.

Multiple Forms of Data — Basic Concepts That Apply to Principal Evaluation Systems

What are multiple forms of data?

- Multiple *forms of data* generally means information that is expressed in more than one way, such as *scores* on a test; *ratings* of performance; *narrative descriptions* of performance; observer *judgments*; and *artifacts* or *products* of work. The forms can be verbal, numeric, graphic, or combinations. A familiar form of data is a *rating* on a rubric or scale, which could be a label (e.g., satisfactory or unsatisfactory), or a numeric rating (e.g., on a scale of 1–5).

- Multiple *methods* are various ways of gathering and interpreting information. Evaluators might *observe* performances and might *interview* principals or others. Districts can *test* principals and students; *collect* samples of principals' work, such as reports or teacher evaluations; and *survey* principals' peers, teachers, and community members.
- Multiple *measures* or instruments are standardized ways to gather information, such as *tests*, *surveys*, *observation protocols*, *rubrics*, or *checklists*. *Standardized* means that these measures are structured to gather consistent information across people and settings.

When are multiple forms of data needed?

Some aspects of principal performance and outcomes are fairly straightforward to measure. Multiple measures and forms of data are called for, however, when the information needed for decisions cannot be captured by a single measure or score. For example, measuring the effects of principals on student learning is complex because the effects are not direct. Principals' effects come about through their influences on school quality, teacher performance, and student learning. Multiple forms of data may be needed to provide enough technically sound information for decisions to be made. See also Key Feature 9, Technically Sound Information, from *Key Features of a Comprehensive Principal Evaluation System* (2012).

What sources of data can provide multiple forms of data?

Information can come from principals themselves, their supervisors, teachers, students, families, other community members, the state, and local agencies. Gathering information from multiple sources such as principals, teachers, and supervisors, and others is referred to as a 360-degree or full-circle look at principal performance from different perspectives. When teacher quality and community engagement are explicit goals for principals, the perspectives of teachers and community members are important sources of information.

How should multiple forms of data be selected?

District policy may call for data about general areas of principal performance, such as principal effectiveness, without specifying what forms of data to gather. Or district policies might specify forms of data for decision-making, such as scores on a state test or student graduation rates. The choice of data forms depends on district policy and decisions about what information is needed to make defensible, valid decisions about principals. Using the *Key Features of a Comprehensive Principal Evaluation System* (2012) will guide districts in determining what forms of data are most appropriate for their circumstances and context.

Districts might adopt or adapt existing methods and measures, such as leadership tests or observation protocols. They might provide several forms of data, such as total scores,

subscores, or descriptive profiles. Other existing methods and measures, such as professionally developed observation protocols or performance tasks, can be used to gather important forms of data. A careful analysis of what data are needed and what will be adequate for the district's purposes should guide selecting the forms of data. (This may entail the district reexamining the purposes for principal evaluation.)

Using existing methods, measures, or data, such as state data or an off-the-shelf leadership test, can be more efficient than developing measures if the existing measures are adequate and appropriate for the district's purposes. However, validity, reliability, fairness, and other technical qualities are not automatically transferrable from one purpose and context to another. Technical quality must be established for local purposes and context. The district must consider how similar the purpose or context of an off-the-shelf leadership test is to the district's purpose and context. If the purposes and contexts are different, the technical qualities (especially defensibility and validity) must be established in the local context.

How can multiple forms of data be combined for decisions?

The descriptions and charts in this brief illustrate ways in which multiple forms of data are being defined, organized, and combined for decision-making. The charts show the areas selected for evaluating principals and how each area is weighted (how much each form of data counts) toward an overall evaluation judgment. For each area to be measured, districts must determine what forms of data to gather and use. A single form of data might be sufficient for some areas, and multiple forms of data for other areas. For example, teacher performance may mean gathering data about teachers who are judged as highly effective; who have implemented a specific teaching approach at an expected level; or who serve as a mentor for new teachers.

Figure 1. Hypothetical Example: Principal Evaluation Data and Weighting
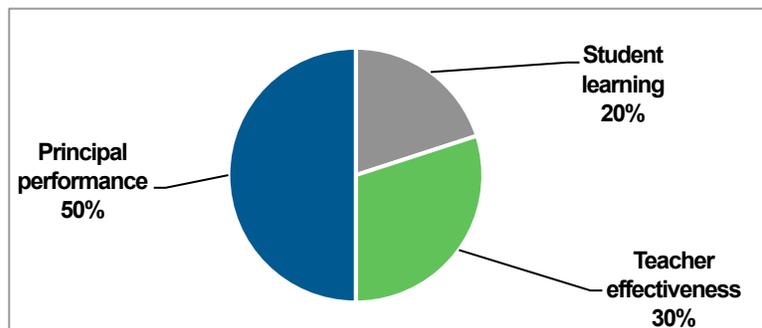

Figure 1 shows a *hypothetical* approach that includes three broad areas in which a district decides to evaluate principal effectiveness: (1) student learning; (2) teacher effectiveness; and (3) principal performance. Each area may provide a different form of data for the evaluation. For example, data about the area of student learning might be in the form of test scores, or a

combination of test scores, graduation rates, and other types of data about district goals for student learning. Data about teacher effectiveness might be in the form of the number of teachers rated as highly qualified, or a combination of data about teacher quality, retention rates, and a value-added measure. Data from the area of principal performance might be in the form of a single qualitative rating by supervisors, or a combination of observations, documentation of activities and effects, and community ratings. When different types of data are gathered in each area, districts can use a single rubric, or they can scale across the areas to combine them, such as, "below expectations," "meeting expectations," or "above expectations" in each area.

Districts should provide detailed descriptions and policies about the areas to be included, the forms of data and measures to be used, and procedures for weighting and combining the areas and forms of data.

1. **Student learning**. A district could use one form of data, such as state achievement test scores, or could combine multiple forms of data, such as state achievement tests (various content areas and levels), graduation rates, attendance, and value-added measures. Data about student learning represents 20 percent of the principal evaluation.

2. **Teacher effectiveness**. A district could use one form of data, such as value-added achievement test scores, or could combine multiple forms of data, such as value-added test scores, the number of teachers rated as effective or highly effective, and teacher retention figures. Data about teacher effectiveness represents 30 percent of the principal evaluation.

3. **Principal performance**. A district could use one form of data, such as a leadership test score, or could combine multiple forms of data, such as a leadership test score, achieving specific goals for the school, and the number of teacher observations completed with timely and specific feedback. Data about principal performance represents 50 percent of the principal evaluation.

Each percentage or weight could be decided based on relative importance to the district's evaluation purposes and the technical soundness of the data for defensible, valid decisions. The examples that follow have well-developed rationales for the data and specify weighting systems. For example, the Arizona framework gives districts flexibility, but it requires that school-level data account for 33 percent to 50 percent of each principal evaluation; other academic performance data add up to 17 percent of the total; and multiple observations of a principal's performance account for 50 percent to 67 percent of the evaluation total.

Comparing the Use of Multiple Forms of Data in Principal Evaluations

Sample measures and forms of data:

Student learning

- Student test scores: reading, mathematics, other content areas;
- Growth rates;
- Value-added measures;
- Targeted student gains (content, grades).

Teacher effectiveness

- Class achievement;
- Targeted student gains;
- Classroom practices;
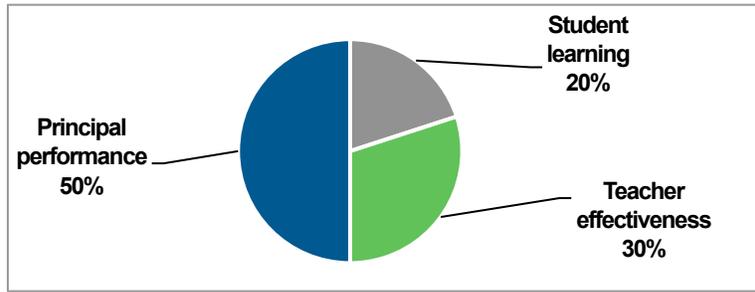- Retention/replacement;
- Teacher evaluation ratings.

School improvement

- Targeted student gains;
- Targeted content gains;
- Quality review;
- Community engagement;
- Parent surveys;
- Teacher surveys;
- School climate measures.

District goals

- Targeted student gains;
- Targeted content gains;
- Public support;
- Administrative support;
- District mandates.

Hypothetical Example: Principal Evaluation Data and Weighting

Student learning 20%
Teacher effectiveness 30%
Principal performance 50%

New York City

District goals 15%
Achievement 32%
Teacher effectiveness 22%
Principal goals 31%

New Leaders

Teacher retention 10%
Teacher effectiveness 10%
Student outcomes 50%
Leadership action 30%

District of Columbia Public Schools

Other 20%
Achievement 40%
Leadership assessment 30%
School goals 10%
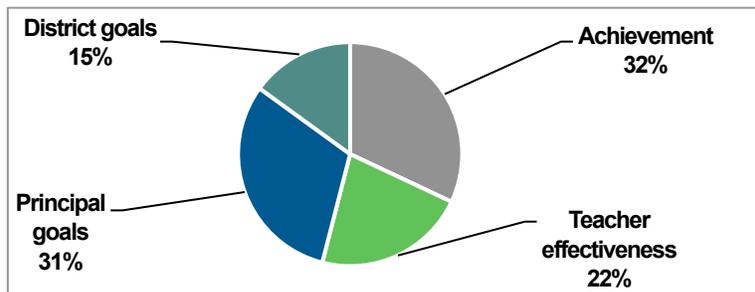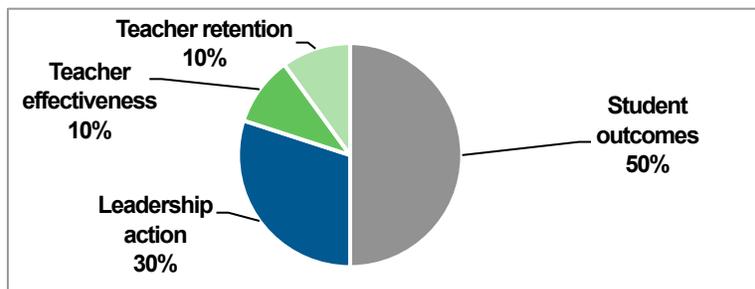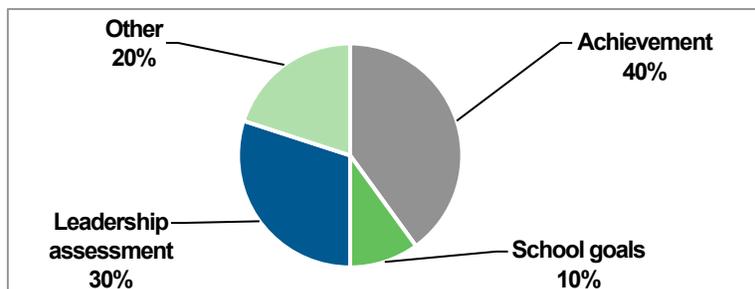
Multiple Forms of Data — Examples From the Field

The following examples represent a range of current options for districts to consider in selecting and using multiple forms and sources of data for principal evaluation. The range of approaches reflects differences in how principal evaluation is conceptualized in different settings, and it also reflects variation in district and state policies, context, and purposes for evaluating principals. The examples include technical documentation that can be useful in evaluating how they might be used in other contexts. The examples do not converge on one specific model for principal evaluation, but they show different considerations for major issues and options.

Reminder: The following are examples of using multiple forms of data, not recommendations for specific approaches. These approaches are changing as they are being implemented and refined. The reference links included in the examples can be used to update information about each source.

Arizona Framework for Measuring Educator Effectiveness

Background: Arizona

The Arizona State Board of Education created the Task Force on Teacher and Principal Evaluation (the Task Force) to develop a new system of teacher and principal evaluation for use beginning in the 2012–13 school year. State law (born from the Race to the Top grant application process) required that the new system include quantitative measures of student academic progress to account for 33 percent to 50 percent of each evaluation. On April 25, 2011, the Board of Education unanimously passed the proposed framework developed by the Task Force.

Key Information: Arizona

The framework promotes flexibility for each local education agency (LEA), as long as four key requirements are satisfied: (1) School-level data must account for 33 percent to 50 percent of each evaluation; (2) Other system/program-level data focusing on specific types of academic performance may be used, but cannot exceed 17 percent of the total (school-level data, combined with system/program-level data, cannot exceed 50 percent of the total evaluation); (3) LEAs must ensure that multiple data elements are used to calculate the student academic portion of each evaluation; and (4) The leadership portion must be based on multiple observations of a principal's performance and should account for 50 percent to 67 percent of the evaluation total.

Indicators: Arizona

School-level data
(33 percent–50 percent of evaluation):

- AIMS (aggregate school or grade level);
- Stanford 10 (aggregate school or grade level);
- District-/school-level benchmark assessments;
- AP, IB Cambridge International, and ACT Quality Core; and
- Other valid and reliable data.

System/Program-level data (17 percent maximum, not to exceed 50 percent when combined with school-level data):

- Survey data;
- Grade-level data;
- Subject-area data;
- Program data; and
- Other valid and reliable data.

Instructional Leadership (50–67 percent of evaluation):

- LEAs may develop their own rubrics as long as they are based on national standards that have been approved by the Board of Education.

Possible Weighting Options: Arizona

School-level data 33%
Instructional leadership 50%
System/school-level data 17%

Instructional leadership 50%
School-level data 50%

School-level data 33%
Instructional leadership 67%
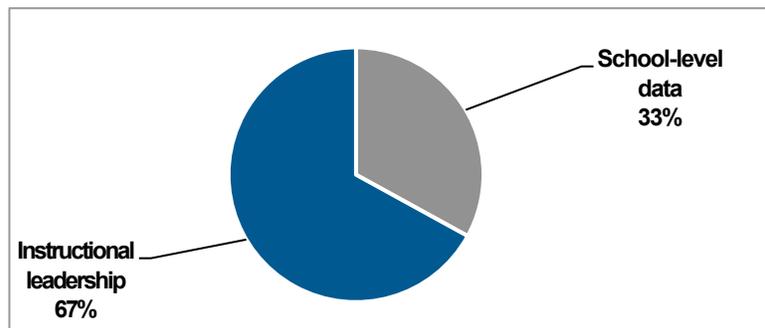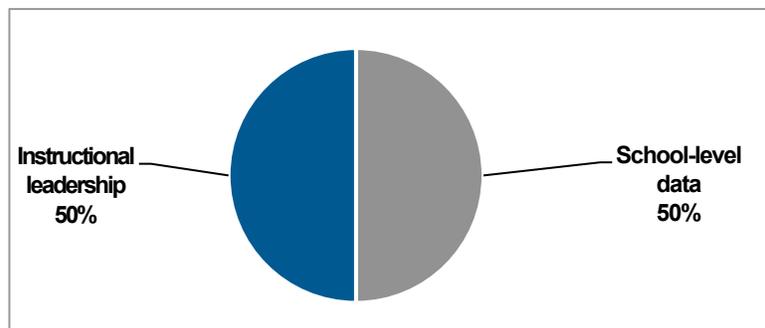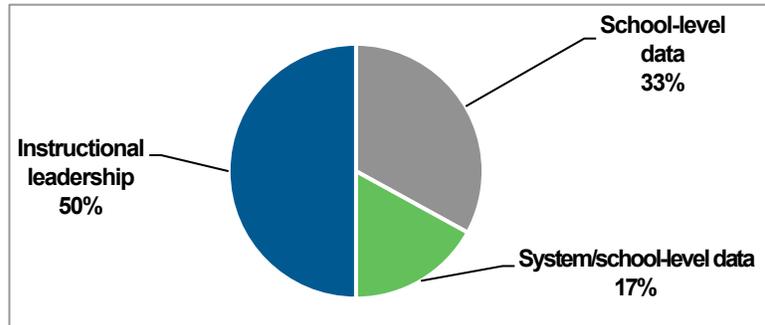
Commentary and Notes: Arizona

The Task Force suggests that student achievement data should not be limited to reading and mathematics, although all student assessment data should be aligned with state standards. The Task Force also recommends that the Interstate School Leaders Licensure Consortium (ISLLC) Professional Administrative Standards should be key features of principal evaluation systems.

Further Information: Arizona

- Arizona Framework for Measuring Educator Effectiveness: http://www.azed.gov/highly-qualified-professionals/teacherprincipal-evaluation/ *(Click on "State Board Adopted Framework" link)*

- Arizona Teacher/Principal Evaluation Web site: http://www.azed.gov/highly-qualified-professionals/teacherprincipal-evaluation/

Colorado Principal Evaluation Framework

Background: Colorado

Colorado's teacher and principal evaluation law, S.B. 10-191, led to the creation of the State Council for Educator Effectiveness (the Council) in January 2010. The Council was charged with tasks that included developing a framework for a fair, rigorous, and transparent system to evaluate teachers and principals. S.B. 10-191 dictates that at least 50 percent of the new evaluation system must be based on student academic growth. In April 2011, the Council submitted its final report and recommendations to the Colorado State Board of Education. A two-year pilot program began in the 2011–12 school year, and the new system will launch statewide in Fall 2013.

Key Information: Colorado

Each district must use the Council's framework for principal evaluation, although districts are allowed flexibility in the weighting decisions of various quality standards and in the selection of supplementary measures to assess those standards. Fifty percent of the principal evaluation must be based on professional practice standards — these standards comprise the first six Colorado Principal Quality Standards. Each of these six standards must account for at least 10 percent of the subscore for the professional practice standards, although one standard can account for as much as 50 percent of the subscore. The remaining 50 percent of the principal evaluation must be based on student academic growth — the newly added seventh standard in the Colorado Principal Quality Standards.

Weighting Display: Colorado
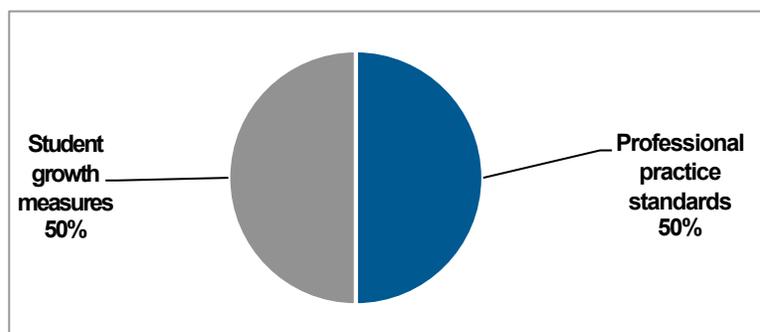
Indicators: Colorado

Professional practice standards (50 percent of evaluation)

Required:
- Teacher and staff perceptions regarding school environment, working conditions, evaluation, and professional supports; and
- Number and percent of teachers in each effectiveness-rating category.

Where appropriate:
- Student/parent/other administrator perception data.

Other:
- Direct observations;
- Evidence of team development;
- Notes from staff meetings;
- School update newsletters;
- Web site content;
- School award structures
- Master school schedule;
- Evidence of community partnerships;
- Parent engagement programs and participation rates;
- 360-degree survey tools;
- School's unified improvement plan;
- Teacher retention data;
- External budget review; and
- School communication plan.

Student academic growth (50 percent of evaluation)

Required:
- Colorado's School Performance Framework:
  o Includes data on student growth and achievement, achievement gaps, graduation rates, etc.; and
- At least one other measure.

Where appropriate:
- Assessments of early literacy/mathematics; and
- Colorado Growth Model.

Other:
- Attendance;
- Promotion rates/transition efforts;
- Students taking higher-level courses;
- High school graduation rates;
- Percentage of students taking courses beyond graduation requirements;
- Percentage of students attending college;
- Student re-engagement rates; and
- Reductions in dropout rates between grades.

Commentary and Notes: Colorado

Although the Council states that districts may select supplementary measures to evaluate principals, districts cannot create supplemental Colorado Principal Quality Standards. The Colorado Department of Education will develop a tool to measure teacher and staff perceptions as they relate to the Colorado Principal Quality Standards.

Further Information: Colorado

- Colorado Principal Quality Standards:
  http://www.cde.state.co.us/EducatorEffectiveness/Partner-SCEE-Resources.asp
  (Click on "Colorado Standards for Principals" link)

- District and School Performance Frameworks Web site:
  http://www.cde.state.co.us/Accountability/PerformanceFrameworks.asp

- Report & Recommendations, State Council for Educator Effectiveness:
  http://www.cde.state.co.us/EducatorEffectiveness/Partner-SCEE.asp
  (Click on "Report Full Text PDF" link)

- State Council for Educator Effectiveness Web site:
  http://www.cde.state.co.us/EducatorEffectiveness/Partner-SCEE.asp

Florida Personnel Evaluation System (Student Success Act)

Background: Florida

In March 2011, the state of Florida passed a new law, the Student Success Act (Senate Bill 736). The Student Success Act requires that at least 50 percent of evaluation scores for instructional staff and administrators be based on student learning growth. The provisions of the Student Success Act supersede the prior evaluation system expectations developed under the state's Race to the Top memorandum of understanding.

Key Information: Florida

The Student Success Act directs that personnel evaluations "be based on sound educational principles and contemporary research in effective educational practices." The law categorizes and defines three components of evaluations: (1) performance of students; (2) instructional leadership; and (3) professional and job responsibilities. At least 50 percent of the evaluation must be based on the performance of students, and the remaining 50 percent can be based on instructional leadership and professional and job responsibilities.

Possible Weighting Options: Florida
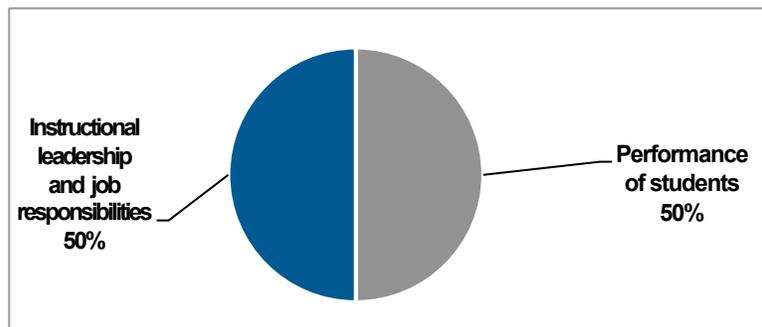
Indicators: Florida

Performance of students (50 percent minimum of evaluation)
- Three years of student data;
- Statewide assessments:
  - Districts must use Commissioner-approved growth formula beginning in 2011–12; and
- District assessments (for subjects and grade levels not measured by statewide assessments — by 2014–15).

Instructional leadership (remaining/50 percent, combined with professional and job responsibilities, of evaluation)
- Performance measures related to effectiveness of the school's teachers;
- Recruitment and retention of effective/highly effective teachers;

- Increases in the percentage of effective/highly effective teachers;
- Appropriate use of evaluation criteria;
- Input from parents; and
- Input from instructional personnel.

Professional and job responsibilities (remaining/50 percent, combined with instructional leadership, of evaluation)
- Job responsibilities, as indicated by the State Board of Education; and
- Additional job responsibilities indicated by the local district.

Commentary and Notes: Florida

The percentage of the evaluation based on student performance may be lowered to 40 percent for those schools with less than three years of student data.

The Florida Department of Education contracted with American Institutes for Research to develop its value-added model. The model will be analyzed over the next four years to determine appropriate adjustments.

The materials on the Florida Department of Education's Web site do not specify the percentage weighting of instructional leadership versus professional job responsibilities.

Further Information: Florida

- Florida Department of Education Review and Approval Checklist for RTTT Principal Evaluation Systems 6-1-2011:
  http://www.fldoe.org/arra/TeacherEvaluationSystems.asp
  (Click on "Review and Approval Checklist for RTTT Principal Evaluation Systems" link)

- Florida Department of Education Teacher and Principal Evaluation Systems Web site:
  http://www.fldoe.org/arra/TeacherEvaluationSystems.asp

- Student Success Act: Overview of SB 736 Legislation Pertaining to Personnel Evaluation Systems March 2011:
  http://www.fldoe.org/arra/pdf/april13tac.pdf

- Summary of CS/CS/SB 736 — Educational Personnel Web site:
  http://www.flsenate.gov/Committees/BillSummaries/2011/html/0736ED

Maryland Educator Evaluation System

Background: Maryland

The Maryland legislature passed the Education Reform Act of 2010, resulting in the creation of the Maryland Council for Educator Effectiveness (the Council). The Council met throughout 2010 and early 2011 before approving guidelines for new evaluation systems for teachers and principals. The resulting principal evaluation system is weighted 50 percent on qualitative measures of professional practice and 50 percent on quantitative student growth measures. Seven jurisdictions will pilot the new systems for twelve months beginning in Fall 2011.

Key Information: Maryland

The Council emphasizes that professional development is the core of its principal evaluation system. The qualitative measures (professional practice) will account for 50 percent of the evaluation and will include the eight outcomes contained in the Maryland Instructional Leadership Framework; however, individual LEAs may add local priorities as needed. The quantitative measures of the evaluation are based on student growth measures, and will be divided between state growth measures (30 percent) and local growth measures (20 percent). Growth measures will comprise "menu items" available to LEAs. These menu items will not be fully developed until the conclusion of the forthcoming pilot program.

Weighting Display: Maryland
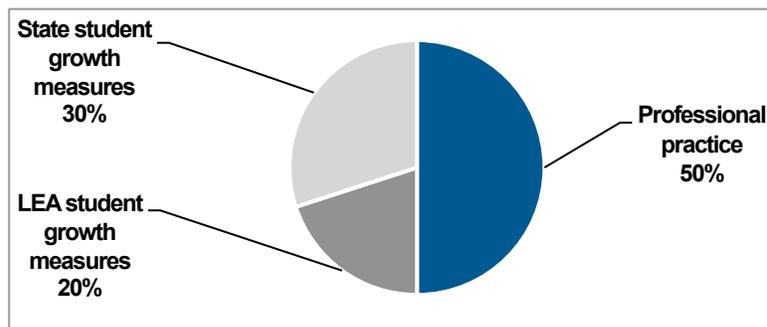
Indicators: Maryland

Professional practice (50 percent of evaluation)
- Maryland Instructional Leadership Framework (25 percent); and
- Other metrics chosen by each LEA depending on local priorities (25 percent).

Student growth measures (50 percent of evaluation)
- LEA-growth measures (20 percent):
  - Menu items will be available to LEAs after the pilot year; and

- State growth measures (30 percent):
  - State assessments.

Commentary and Notes: Maryland

None.

Further Information: Maryland

- Maryland Council for Educator Effectiveness Web site:
  http://www.marylandpublicschools.org/MSDE/programs/race_to_the_top/MCEE

New Leaders (NL) Principal Evaluation System

Background: New Leaders

New Leaders (NL) produced a policy paper, *Evaluating Principals: Balancing Accountability with Professional Growth* (2010), to assist policymakers in the development of improved principal evaluation systems. NL's review of the literature found that current principal evaluation systems do not focus on useful points, are based on unclear performance standards, are limited in terms of the rigor of their design, and are limited in the rigor of their implementation. NL recommends that student outcomes and teacher effectiveness outcomes account for 70 percent of the evaluation, and leadership actions account for 30 percent. The system should be informed by principals and other experts and adapted according to future developments in the understanding of practices that contribute to student achievement.

Key Information: New Leaders

NL proposes that principals be measured against three standards: (1) increased success levels for all students (student achievement outcomes); (2) increased levels of effective teachers on staff (teacher effectiveness outcomes); and (3) critical leadership actions that drive student achievement.

Regarding student achievement outcomes, NL focuses on two measures of student success: (1) assessment results in core academic subjects; and (2) other academic measures of college readiness. NL recommends that growth measures constitute at least 50 percent of the student outcomes used in the evaluation (with the remainder focusing on attainment).

Regarding teacher effectiveness outcomes, NL proposes two distinct measures of a principal's impact: (1) growth in the percentage of teachers making "effective" gains in student achievement outcomes; and (2) improvement in the differential retention of effective teachers (this can be applied only in instances where the principal has authority over teacher hiring and the organization of professional development).

NL has developed standards for principal evaluation (based on its review of existing standards) and an accompanying rubric based on six domains of leadership actions. An evaluation (leadership actions) based on this rubric accounts for the remaining 30 percent of the NL principal evaluation system.

Weighting Display: New Leaders

Teachers' gains in student outcomes
10%

Differential retention of effective teachers
10%

Leadership actions
30%

Student achievement outcomes
50%
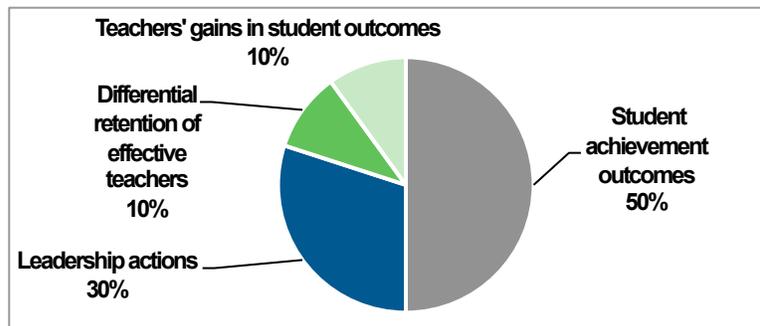
Indicators: New Leaders

Student achievement outcomes (50 percent of evaluation)
- State tests in English Language Arts, mathematics, and other academic subjects;
- Grade-to-grade progression;
- Credit accumulation;
- Graduation rates by cohorts of students;
- Quality and rigor of high school diplomas; and
- Rates of college acceptance/persistence.

Teacher effectiveness outcomes (20 percent of evaluation)
- Growth in the percentage of teachers making "effective" gains in student outcomes (10 percent); and
- Improvement in the differential retention of effective teachers (10 percent).

Leadership actions (30 percent of evaluation)
- Rubric based on six domains of leadership actions: (1) Vision for Results and Equity; (2) Planning and Operations; (3) School Culture; (4) Learning and Teaching; (5) Staff Development and Management; and (6) Personal Leadership and Growth.

Commentary and Notes: New Leaders

NL recommends that the evaluation of principal managers (direct supervisors of principals) be based primarily on student achievement outcomes and principal effectiveness.

Further Information: New Leaders

- Evaluating Principals — Balancing accountability with professional growth:
  http://www.newleaders.org/newsreports/publications/evaluating-principals/
  (Click on the graphic link)

- Principal Leadership Actions: A guide to assessing principal actions that drive dramatic gains in student achievement:
  http://www.newleaders.org/newsreports/publications/evaluating-principals/principal-leadership-actions-rubric/
  (Click on the graphic link)

New York City Principal Performance Review

Background: New York City

In 2005, the New York City Department of Education began efforts to incorporate student achievement data into its Principal Performance Review (NYC PPR). Two years of discussions and pilot programs led to a revised NYC PPR agreement that was signed by principals in 2007 — this agreement includes more autonomy for principals in return for increased accountability for student academic progress.

Key Information: New York City

The NYC PPR comprises the following five areas: (1) accomplishing defined goals and objectives; (2) results from school's Progress Report; (3) results from school's Quality Review; (4) addressing the needs of Special Education and English Language Learner students; and (5) addressing legal requirements and other relevant policies.

Three of the five areas constitute the academic performance portion of the NYC PPR. Goals and objectives account for 31 percent of the evaluation and must relate to student achievement. The school's progress report accounts for 32 percent of the evaluation and focuses on the school's growth in academic performance (60 percent of the progress report), absolute performance (25 percent of the progress report), and school climate (15 percent of the progress report). The school's Quality Review score accounts for 22 percent of the evaluation and is focused on school practices that stem from a principal's actions. The Quality Review is guided by a rubric that comprises five quality statements that determine the organizational readiness of a school to educate students.

The remaining areas focus on the provision of certain services for Special Education and English Language Learner students (5 percent) and district policies and legal mandates (10 percent).

Weighting Display: New York City
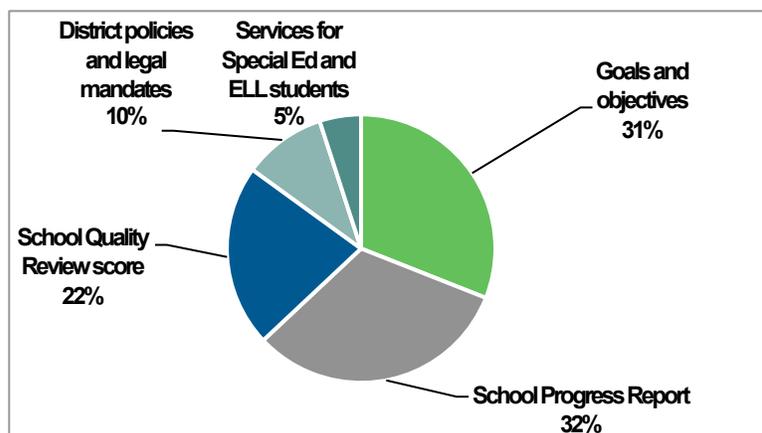
Indicators: New York City

Academic performance (85 percent of evaluation)

- Goals and Objectives (31 percent of academic performance):
    - o Must focus on student achievement; and
    - o Based on data from school progress reports.

- School Progress Report results (32 percent of academic performance):
    - o School's growth in academic performance (60 percent of progress report);
    - o Absolute performance (25 percent of progress report); and
    - o School climate (15 percent of progress report).

- School Quality Review score (22 percent of academic performance).

Services for Special Education and English Language Learner students (5 percent of evaluation)

District policies and legal mandates (10 percent of evaluation)

Commentary and Notes: New York City

Although the NYC PPR manual lists the School Quality Review score as an area under academic performance, it more closely resembles a qualitative leadership assessment. The rubric is completed through a two- or three-day site visit during which interviews and classroom observations are conducted by an external evaluator.

Further Information: New York City

- Evaluating Principals: Balancing accountability with professional growth:
  http://www.newleaders.org/newsreports/publications/evaluating-principals/
  (Click on the graphic link)

- How Four Districts Crafted Innovative Principal Evaluation Systems (case study):
  http://www.wested.org/cs/we/view/rs/1109

- NYC Department of Education 2011-2012 Principal Performance Review (PDF):
  http://www.csa-nyc.org/csa-members/doe-members/other-documents-and-agreements#PPR
  (Click on "Principal's Performance Review and compliance documents" link)

- NYC Department of Education Quality Review Web site:
  http://schools.nyc.gov/Accountability/tools/review/default.htm

- NYC Department of Education Progress Report Web site:
  http://schools.nyc.gov/Accountability/tools/report/default.htm

Tennessee Teacher and Principal Evaluation Policy

Background: Tennessee

Tennessee's First to the Top legislation called for teachers and principals to undergo annual evaluations beginning in the 2011–12 school year. The legislation created the Teacher Evaluation Advisory Council (TEAC) to establish criteria for LEAs to follow in developing their teacher and principal evaluation systems.

Key Information: Tennessee

Approvable LEA evaluation models must include the following three criteria: (1) the primary purpose of the evaluation is to identify and support instruction that leads to high levels of student achievement; (2) evaluations will be used to inform human capital decisions (professional development, hiring, assignments, promotions, tenure, dismissal, compensation); and (3) evaluations will assess principals based on five ratings: (a) significantly above expectations; (b) above expectations; (c) at expectations; (d) below expectations; and (e) significantly below expectations.

Fifty percent of the evaluation must be based on student achievement data (35 percent based on student growth data, and 15 percent based on other measures of student achievement). The remaining 50 percent must be based on a qualitative appraisal instrument.

Weighting Display: Tennessee

Student growth 35%

Measures of student achievement (e.g., state scores) 15%

Qualitative appraisal instrument – leadership 50%
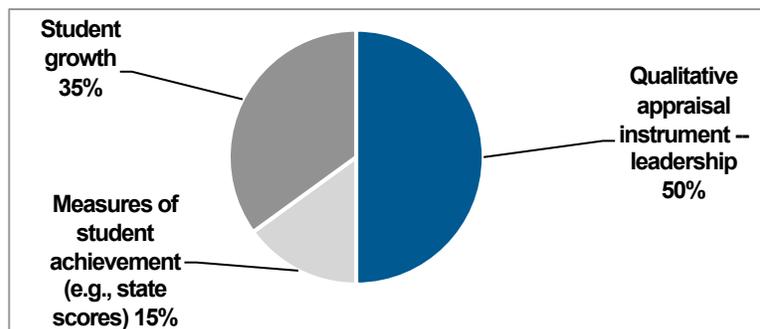
Indicators: Tennessee

Student achievement data (50 percent of evaluation)

- Student growth (35 percent):
  o School-level value-added scores.

- Measures of student achievement (15 percent):
  o State assessments;
  o Tennessee Value-Added Assessment System (TVAAS);
  o National/State off-the-shelf tests;

- o AP/IB/NIC suites of assessments;
- o Graduation rate/CTE Concentrator Graduation Rate;
- o Postsecondary matriculation/ persistence/placement, as determined by the Tennessee Department of Education (TDOE) and Tennessee Higher Education Commission (THEC);
- o Participation in advanced coursework; and
- o Ninth grade promotion and retention rate.

Qualitative appraisal instrument (50 percent of evaluation)

- Approved evaluation model based on the Tennessee Instructional Leadership Standards;
- Review of the quality of principals' teacher evaluations;
- School climate/teaching and learning conditions surveys; and
- At least two onsite observations of principals per year (conducted by the director of schools or a designee).

Commentary and Notes: Tennessee

The principal evaluation guidelines in this document apply to principals and school administrators who spend at least 50 percent of their time on administrative functions. Principals who rate in the top three quintiles of student growth may use their student growth scores for the entire 50 percent student achievement portion, thus omitting the "measures of student achievement" portion.

Further Information: Tennessee

- Tennessee First to the Top — Teacher and Principal Evaluation Web site: http://www.tn.gov/firsttothetop/programs-committee.html

- Tennessee State Board of Education. Teacher and Principal Evaluation Policy: http://www.tn.gov/firsttothetop/docs/IV_C_Teacher_and_Principal_Evaluation_Policy.pdf

## Washington D.C. Public Schools Effectiveness Assessment System for School-Based Personnel (School Leaders): IMPACT

Background: Washington D.C.

District of Columbia Public Schools (DCPS) introduced IMPACT in 2009 as part of a reform effort to establish a comprehensive system of guidance and support for teachers and school-based staff. Revised for the 2010–11 school year, IMPACT now extends to school leaders and provides a Leadership Framework Assessment and clear performance expectations.

Key Information: Washington D.C.

The IMPACT principal evaluation system comprises two main elements: (1) student achievement; and (2) leadership outcomes. Each element accounts for 50 percent of the overall evaluation. The student achievement element comprises three distinct measures: (1) District of Columbia Comprehensive Assessment System (DC CAS) proficiency goals; (2) school-specific student achievement goals; and (3) the school's value-added score. The leadership outcomes element comprises four distinct measures: (1) Leadership Framework Assessments midyear and end-of-year scores; (2) Special Education Compliance; (3) Teacher Retention; and (4) Family Engagement.

Specified Weighting: Washington D.C.
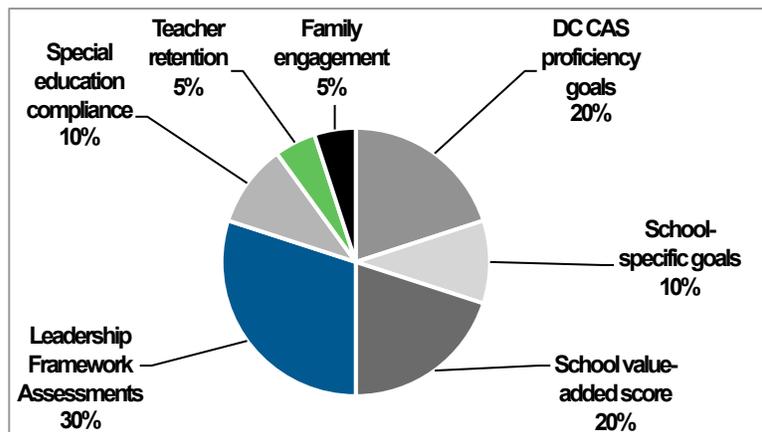
Indicators: Washington D.C.

Student achievement (50 percent of evaluation)

- DC CAS Goals (20 percent of evaluation score):
  - Proficiency goal for reading; and
  - Proficiency goal for mathematics.

- School-specific achievement goals — select two (10 percent)
  - Graduation rates;

- o   DC CAS performance for a particular subgroup;
- o   Increases in the percentage of students achieving advanced level; and
- o   Promotion rates from ninth grade to tenth grade.

- School value-added score (20 percent):
  - o   Uses the same score as the school value-added score used in staff members' IMPACT evaluations.

Leadership outcomes (50 percent of evaluation)

- Leadership Framework Assessments (30 percent):
  - o   Midyear assessment; and
  - o   End-of-year assessment.

- Special Education Compliance (10 percent):
  - o   School Individual Educational Plans (IEP) timeliness;
  - o   Required-action timeliness;
  - o   Eligibility timeliness; and
  - o   IEP documentation.

- Teacher Retention (5 percent):
  - o   Percent of teachers retained who have an IMPACT score of 300 or higher.

- Family Engagement (5 percent):
  - o   Responses to the DCPS stakeholder survey.

Commentary and Notes: Washington D.C.

The 2010–11 revision of the IMPACT system was informed by focus groups, conversations, and feedback sessions.

The Leadership Framework Assessment comprises three domains of effective school leadership: (1) instructional leadership; (2) organizational leadership; and (3) leadership for increased effectiveness. A four-level rating scale and rubric are used by superintendents to rate principals in each domain.

Further Information: Washington D.C.

- About DC Comprehensive Assessment System (DC CAS) Web site:
  http://dcps.dc.gov/DCPS/dccas

- IMPACT — The DCPS Effectiveness Assessment System for School-Based Personnel
  Web site:
  http://www.dc.gov/DCPS/impact

- IMPACT—The District of Columbia Public Schools Effectiveness Assessment System for School-Based Personnel 2011–2012 (Guidebook for Principals): http://www.dc.gov/DCPS/Learn+About+Schools/School+Leadership/IMPACT+%28Performance+Assessment%29
  (Click on "2011–12 DCPS School Leader IMPACT Guidebook: Principals" link)

- Leadership Framework Web site: http://www.dc.gov/DCPS/Learn+About+Schools/School+Leadership/Leadership+Framework

Conclusion

Leadership is complex, and thus measuring leadership effectiveness demands using multiple forms of data for defensible, valid district decisions. This brief describes basic concepts for using multiple forms of data in principal evaluation systems, including information about student learning. While research and professional standards show that there is strong agreement that using more than one form of data and data source is good practice, there is less consensus on the methods that districts select and use in principal evaluations. This is demonstrated by the examples from states, districts, and initiatives included in this brief. It is important to follow these efforts as they are revised and implemented to see how the concept of using multiple forms of data plays out in practice.

Key questions to keep in mind are: How can collecting multiple types of data be integrated into other district data collection systems for efficiency and to inform the day-to-day work of principals? Do multiple forms of data provide more accurate and useful principal evaluations than single measures? Do collecting and using multiple forms of data inform and improve principals' performance, teacher effectiveness, and student learning?