

# Practical Guide to Designing Comprehensive Principal Evaluation Systems



A Tool to Assist in the Development of Principal Evaluation Systems

FEBRUARY 2014

## **Acknowledgments**

We recognize and thank our colleagues at American Institutes for Research (AIR) for their contributions to this guide: Mariann Lemke, Roshni Menon, Melissa Brown-Sims, Jenni Fetters, and Kim Bobola. We also thank Laura Goe (ETS) and Lynn Holdheide (AIR) for their thoughtful advice and encouragement. In addition, we extend our appreciation to our gracious reviewers: Steven Ross (Johns Hopkins University), Steve Kimball (University of Wisconsin-Madison), Mike Schooley (National Association of Elementary School Principals), Tony Milanowski (Westat), Christine Mason (National Association of Elementary School Principals), and Cary Baird (consultant).

See <http://www.gtlcenter.org/tools-publications/online-tools/principal-evaluation> to view the interactive, online version of the *Practical Guide to Designing Comprehensive Principal Evaluation Systems*.

# **Practical Guide to Designing Comprehensive Principal Evaluation Systems**

Revised Edition

---

FEBRUARY 2014

Matthew Clifford, Ph.D.

American Institutes for Research

Ulcca Joshni Hansen, Ph.D., J.D.

Public Education & Business Coalition

Sara Wraight, J.D.

American Institutes for Research

# Contents

Rationale and Structure . . . . .	1
Overview of Components . . . . .	1
Research and Policy Context . . . . .	2
Research on Principal Influence . . . . .	2
Research on Principal Evaluation . . . . .	5
Goals of Principal Evaluation Design . . . . .	7
State Accountability and District Responsibility in Principal Evaluation Systems . . . . .	8
Key State Roles . . . . .	8
Models for State and District Evaluation Systems . . . . .	10
Factors for Stakeholder Consideration . . . . .	12
Development and Implementation of Comprehensive Principal Evaluation Systems . . . . .	16
Component 1a: Specifying Evaluation System Goals . . . . .	16
Component 1b: Defining Principal Effectiveness and Establishing Standards . . . . .	20
Component 2: Securing and Sustaining Stakeholder Investment, and Cultivating a Strategic Communication Plan . . . . .	22
Component 3: Selecting Measures . . . . .	28
Component 4: Determining the Structure of the Evaluation System . . . . .	41
Component 5: Selecting and Training Evaluators . . . . .	47
Component 6: Ensuring Data Integrity and Transparency . . . . .	50
Component 7: Using Principal Evaluation Results . . . . .	53
Component 8: Evaluating the System . . . . .	57
Conclusion and Recommendations . . . . .	59
References . . . . .	60
Appendix A. Glossary of Terms . . . . .	64
Appendix B. Some Principal Evaluation Measures . . . . .	69

## Rationale and Structure

Across the country, states and districts are designing principal evaluation systems as a means of improving leadership, learning, and school performance. Principal evaluation systems hold potential for supporting leaders' learning and sense of accountability for instructional excellence and student performance. Principal evaluation also is an important component of state and district systems of leadership support efforts, especially when newly designed evaluation systems work in conjunction with principal certification, hiring, and professional development systems.

The *Practical Guide to Designing Comprehensive Principal Evaluation Systems* is intended to assist states and districts in developing systems of principal evaluation and support. The guide is informed by research on performance evaluation design and lessons learned through the experience of state and district evaluation designers. It is organized in three sections:

- Research and Policy Context
- State Accountability and District Responsibility in Principal Evaluation Systems
- Development and Implementation of Comprehensive Principal Evaluation Systems

## Overview of Components

The guide discusses the following components as critical to states' and districts' success in redesigning principal evaluation:

- Component 1a: Specifying Evaluation System Goals
- Component 1b: Defining Principal Effectiveness and Establishing Standards
- Component 2: Securing and Sustaining Stakeholder Investment, and Cultivating a Strategic Communication Plan
- Component 3: Selecting Measures
- Component 4: Determining the Structure of the Evaluation System
- Component 5: Selecting and Training Evaluators
- Component 6: Ensuring Data Integrity and Transparency
- Component 7: Using Principal Evaluation Results
- Component 8: Evaluating the System

Each component includes an overview; practical examples; and guiding questions designed to help stakeholders organize their work, design better evaluation systems, and launch new designs within their state or district. This guide complements the *Practical Guide to Designing Comprehensive Teacher Evaluation* (Goe, Holdheide, & Miller, 2014).

This guide should be used as a facilitation tool for conversation among designers, not as a step-by-step approach to redesigning principal evaluation systems. State and district policymakers should address all components of the guide but also should capitalize on local capacity and processes when doing so.

## Design Assumptions

The following assumptions about principal evaluation design have informed the guide:

- Principal evaluation systems should be as comprehensive as possible while also being feasible to implement.
- Principal evaluations should be accurate, fair, and useful.
- Principals' work is more varied than that of teachers, and their influence on student achievement is indirect; therefore, evaluation systems should have multiple measures of performance and impact—including, but not limited to, student achievement or growth.
- Principals' leadership can extend throughout and beyond the school; therefore, evaluation system designers will want to gather multiple stakeholder perspectives on principal performance.

## Policy Assumptions

The following assumptions about the policy context have informed the guide:

- New evaluation systems should engage stakeholders from across the principal career spectrum to ensure that the system is effective and that evidence informs other services.
- States and districts should consider how well the current principal evaluation system works and capitalize on its strengths during redesign.
- Evaluation systems should include input from principals and other stakeholders.
- Policymakers should ensure that teacher and principal evaluation systems are coherent and mutually supportive.
- Efforts to improve principal evaluation systems are informed by federal initiatives, state legislation, professional association perspectives, and foundation-led efforts. New evaluation systems should be aligned with these efforts.
- States may be in various stages of plan development or revision for a statewide system of principal evaluation and support, so the guide allows designers to focus on the components that are most relevant to them.

## Research and Policy Context

Performance evaluation systems should be based on research-based definitions of educator effectiveness. This section of the guide provides research and policy information about defining principal effectiveness and the need to improve principal evaluation. The information is drawn from several research syntheses and studies focusing on school principal effects, the status of principal evaluation in the field, and national policy initiatives. State and district evaluation designers may find this section useful when orienting stakeholders to issues in principal evaluation. In addition, designers are encouraged to review research documents (see “Additional Research” at right) and speak to principals, superintendents, and other stakeholders about the status of principal evaluation in the state or district.

### Research on Principal Influence

Although research on principal leadership impact continues to evolve, it indicates that principals directly and indirectly affect student learning through their leadership practices.

#### ADDITIONAL RESEARCH

See the following websites for additional research on designing evaluation systems:

- ▶ American Educational Research Association  
<http://www.aera.net>
- ▶ National Association of Elementary School Principals  
<http://www.naesp.org>
- ▶ National Association of Secondary School Principals  
<http://www.nassp.org>
- ▶ The Wallace Foundation  
<http://www.wallacefoundation.org>
- ▶ University Council for Educational Administration  
<http://www.ucea.org>
- ▶ What Works Clearinghouse  
<http://ies.ed.gov/ncee/wwc/>

Figure 1 displays principals’ spheres of influence, according to reviewed research. These areas of interest should be considered by policymakers when designing evaluation systems.

## Principal Practice

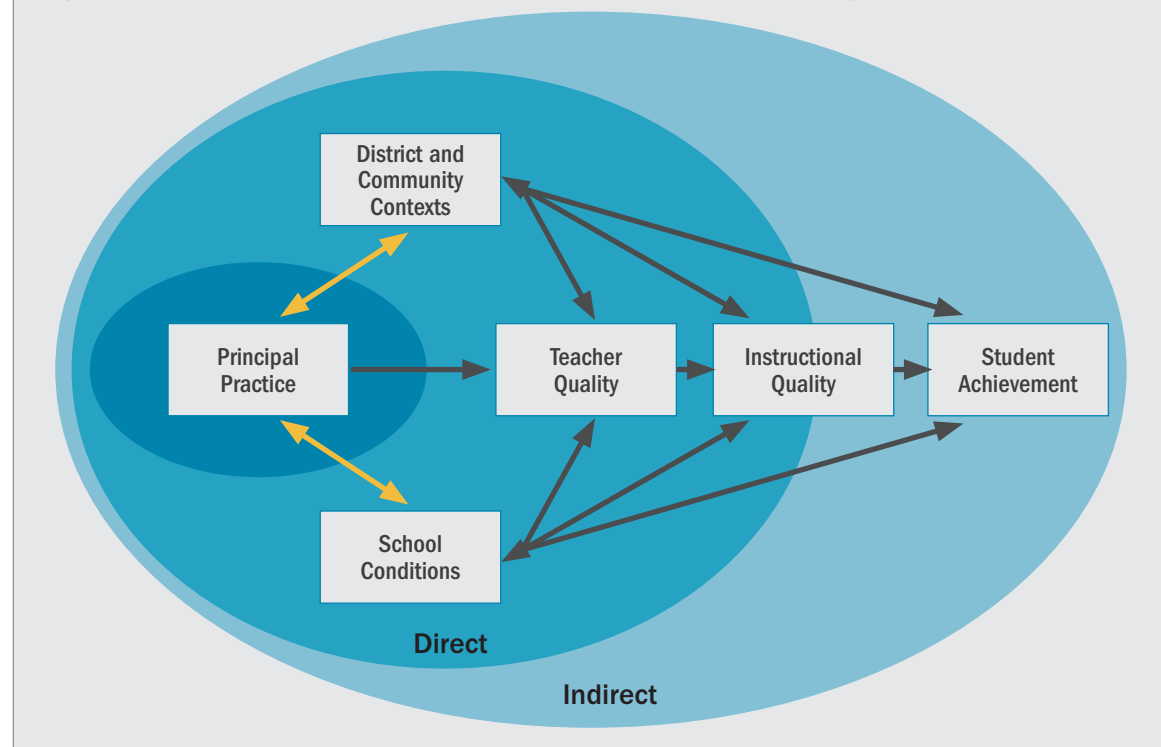
Principals influence student learning and school performance through their practice, which includes knowledge, dispositions, and actions. Although principal effectiveness research is far from definitive (Davis, Kearney, Sanders, Thomas, & Leon, 2011), information about principals' practice forms a reasonable base for principal evaluation and professional development designs.<sup>1</sup>

Researchers have examined studies for evidence of practices that make a difference in schools. Common findings across studies indicate that the following principal practices are associated with student achievement and high-performing schools:

- Creating and sustaining an ambitious, commonly accepted vision and mission for organizational performance
- Engaging deeply with teachers and data on issues of student performance and instructional services quality
- Efficiently managing resources such as human capital, time, and funding
- Creating physically, emotionally, and cognitively safe learning environments for students and staff
- Developing strong and respectful relationships with parents, communities, and businesses to mutually support children's education

<sup>1</sup> Although studies point to practices of effective principals, less empirical work describes how principals do their work and how leadership tasks are distributed so that strong leadership is maintained in schools (Halverson & Clifford, in press; Spillane & Diamond, 2007; Spillane, Halverson, & Diamond, 2004). Understanding how principals conduct their work and how leadership is distributed in schools can provide better insight into the daily work of effective principals and better descriptions of principal practice. Such descriptions are important for the development of evaluation instruments and processes.

Figure 1. Direct and Indirect Influence of Principals on Student Learning



- Acting in a professional and ethical manner (Council of Chief State School Officers, 2008; Marzano, Waters, & McNulty, 2005; Stronge, Richard, & Catano, 2008)

### Direct Influence

By virtue of their position, principals can directly influence school conditions, district and community contexts, teacher quality and distribution, and instructional quality.

In summarizing the research on principal effects, Hallinger and Heck (1998) found that foremost among the ways principals foster school improvement is shaping school goals, school improvement directions, school policies and practices, school structures, and the social and organizational networks within their schools. Similarly, Wahlstrom, Louis, Leithwood, and Anderson (2010) concluded from their meta-analysis of principal effectiveness studies that

principals' influence student achievement by influencing school contexts.

Research also suggests that principals influence *teacher working conditions*, which are defined as teachers' perceptions of the condition of their work and school. Principals' roles in creating positive teacher working conditions include fostering a collegial, trusting, team-based, and supportive school culture; promoting ethical behavior; encouraging data use; and creating strong lines of communication. Ladd (2009) found an association between positive teacher working conditions and student achievement. Principals shape teacher working conditions by acting as school-level human capital managers who may have power to oversee school teacher hiring, placement, evaluation, and professional learning (Kimball, 2011; Milanowski & Kimball, 2010).

Although principals influence school conditions, it is important to note that principals' work also is influenced by school conditions. New principals inherit organizational histories and traditions that they must work through and within in order to bring about meaningful change, and fluctuations in organizational conditions can affect principals' leadership styles or the discretion that principals have to bring about change (Lambert et al., 2004). Principals in turnaround schools, for example, likely need to act quickly and convincingly to improve conditions and achievement (Herman et al., 2008). Other

school contexts may support and inhibit different types of leadership practices.

School principals also influence the district and community contexts of schools and schooling. They oversee the organizational processes that are needed to implement change and to garner the support of the community, parents, teachers, and students in developing district-level policies that regulate relationships between districts and schools (Waters, Marzano, & McNulty, 2003).

Finally, principals also can have a strong and immediate influence on teacher quality, including the distribution of teacher talent. For example, the Retaining Teacher Talent study found that teachers viewed principal quality as a strong factor in their choice to join or leave a school (Public Agenda, 2009). Milanowski et al. (2009) similarly found that principal quality was the most important factor in attracting prospective teachers. Teachers also consider principals as critical factors in their decision to leave the profession (Ingersoll & Smith, 2003).

Working under the supervision of an inspiring and highly competent principal is *exactly* what makes the difference in teachers' openness, even eagerness, to work in challenging school environments (The Wallace Foundation, 2011).

## Indirect Influence

Principals indirectly influence student achievement and instructional quality by creating conditions within schools. Although the influence is indirect, principal effectiveness is defined by these outcomes. Federal and state policies require student growth to be included in principal performance evaluation. Studies on the association between leadership and student achievement suggest that principals have a strong influence on student learning, albeit indirect and not easily measurable. Although many student learning factors have not been fully explained, school leadership is generally recognized as the second most influential school-level factor influencing student achievement, after teacher quality (Hallinger & Heck, 1998; Leithwood, Louis, Anderson, & Wahlstrom, 2004; Murphy & Datnow, 2003; Supovitz & Pogliinco, 2001; Waters et al., 2003). Available studies indicate that principal actions explain between .25 and .34 of the variation in student performance (Leithwood et al., 2004).

Principals also indirectly influence instructional quality by providing resources to teachers and signaling the types of instruction that are acceptable and optimal in the school. Principals can influence instructional quality by providing feedback to teachers; allocating resources to professional development and instructional support; emphasizing the



importance of professional learning communities as a means of reflection and job-embedded professional development; and selecting programs, curriculum, and other instructional resources.

## Research on Principal Evaluation

Principal evaluation has long held promise for improving principal effectiveness, fostering learning and reflection, and increasing accountability for job performance (Orr, 2011). Performance evaluation is particularly important for principals because they report having few opportunities to receive trusted feedback on their work and commonly feel isolated from colleagues due to the rigors of their position (Friedman, 2002). Performance evaluation provides a method for principals to receive feedback on their practices from an evaluator.

Although principal evaluation holds great potential, few research or evaluation studies are currently available on the design or effects of performance evaluation on principals, schools, or students (Clifford & Ross, 2011). Available research studies raise questions about the consistency, fairness, effectiveness, and value of current principal evaluation practices (Condon & Clifford, 2010; Goldring et al., 2009; Heck & Marcoulides, 1996; Portin, Feldman, & Knapp, 2006; Thomas, Holdaway, & Ward, 2000).

Studies indicate that:

- Principals see little value in current evaluation practices.
- Principal evaluations are inconsistently administered.
- Performance evaluation systems and instruments may not be aligned with existing state or national professional standards for practice or standards for personnel evaluation.
- Few widely available principal evaluation instruments have psychometric rigor.

To increase the effectiveness of principal evaluation, state and district policy designers should develop systems that establish explicit expectations for performance and instill confidence and trust in performance ratings and quality of feedback in principals.

## Policy Context

State and federal policies and initiatives have encouraged stakeholders to redesign principal evaluation systems. The American Recovery and Reinvestment Act (ARRA) and the Race to the Top competition encouraged states and districts to develop more rigorous evaluations for high-stakes personnel decisions, including principal retention and compensation. These same federal policies and initiatives provide impetus for teacher evaluation systems improvement.

Most recently, the Elementary and Secondary Education Act (ESEA) flexibility guidance requires states and districts to create principal evaluation systems that:

- Will be used for continual improvement of instruction.
- Meaningfully differentiate performance using at least three levels.
- Use multiple valid measures in determining performance levels, including as a significant factor data on student growth for all students (including English learners and students with disabilities) and other measures of professional practice (which may be gathered through multiple formats and sources, such as observations based on rigorous performance standards and surveys).
- Evaluate principals on a regular basis.
- Provide clear, timely, and useful feedback, including feedback that identifies needs and guides professional development.
- Will be used to inform personnel decisions.

Federal initiatives also require states and districts to create teacher evaluation systems that meet similar criteria, which can facilitate the alignment of teacher and principal evaluation systems.

## State Plans

ESEA flexibility requirements also stipulate that states describe in their plans how they meaningfully engage and solicit input from principals and principal representatives, diverse communities, and other stakeholders. The guidelines encourage states to use multiple methods of communication to actively engage stakeholders from the start and to note specific changes based on input.

States must describe the process for determining validity and reliability of evaluation measures and how the measures will be applied consistently across school districts. In addition, states must identify measures intended for use in evaluating teachers of nontested grades and subjects. They must include rubrics for training and supporting evaluators in evaluating principals, addressing the education of English learners and students with disabilities. States must provide assurances for data collection and reporting quality and include a method for clearly communicating results to principals.

Each state plan must have processes for reviewing and approving district plans for consistency with state guidelines. The state is responsible for ensuring that districts involve principals in developing, adopting, piloting, and implementing these systems. Further, the state must ensure the use of valid measures and consistent, high-quality implementation across schools within the district (e.g., a process for ensuring inter-rater reliability).

In preparation for these competitions and flexibility requests, many states have passed legislation requiring improvements in evaluation systems. These opportunities have raised awareness of the urgency to enact improved measures of principal effectiveness and support principal growth. Currently, advisory boards, task forces, and multistate consortia are gathering ideas and information to improve evaluation systems.

## State and District Evaluation Design

State and district evaluation design efforts can capitalize on previously developed standards for professional practice and personnel evaluation. National professional standards for principal practice are based on existing research on school principal practice and have been developed through extensive input from practitioners. More than 40 states have passed legislation adopting one or more sets of national professional practices standards, and these standards have been integrated into many preservice and inservice training programs. The following standards may serve as a starting point for additional review and evaluation design:

- **Interstate School Leadership Licensure Consortium (ISLLC) Standards and Indicators.** The ISLLC Standards and Indicators have been produced through extensive review of principal and school effectiveness literature (Council of Chief

State School Officers, 2008). They have been adopted by a majority of states for performance evaluation and preparation purposes (Anthes, 2005; Hale & Moorman, 2003). Those standards can be found online ([http://www.ccsso.org/Documents/2008/Educational\\_Leadership\\_Policy\\_Standards\\_2008.pdf](http://www.ccsso.org/Documents/2008/Educational_Leadership_Policy_Standards_2008.pdf)).

- **National Board for Professional Teaching Standards (NBPTS): Accomplished Principal Standards.** These standards are designed to guide principal development through an extensive review of research literature and expert input. They are intended to guide principal development as instructional leaders and underpin the NBPTS master principal assessment system. Those standards can be found online ([http://www.nbpts.org/sites/default/files/documents/FINAL%20PRINT%20VERSION\\_PRINCIPAL%20STANDARDS.pdf](http://www.nbpts.org/sites/default/files/documents/FINAL%20PRINT%20VERSION_PRINCIPAL%20STANDARDS.pdf)).
- **National Association of Elementary School Principals' Leading Learning Communities: Standards for What Principals Should Know and Be Able to Do.** These standards focus on the role of principals as instructional leaders and participants in learning communities within schools that create conditions for continuously improving student learning. Information on those standards can be found online (<http://www.naesp.org/leading-learning-communities>).

In addition to these nationally recognized, research-based standards for school leaders, other individuals and organizations have created standards for leadership practice to inform state and district evaluation systems. State and district design teams may wish to consult other research-based leadership standards as they develop evaluation systems. For example, master teacher and teacher-leader standards may be informative to principal evaluation design teams as they compare principal and teacher standards. Master teacher standards can be found online (<http://www.nbpts.org>), and teacher-leader standards can be found online (<http://www.teacherleaderstandards.org>).

Research from the fields of human resources and educational human capital management has provided a set of standards to guide design and improvement of personnel evaluation systems. The Joint Committee on Standards for Educational Evaluation's *Personnel Evaluation Standards* (Gullickson, 2009) provides a starting point for policymakers, evaluation designers, and others. A summary of those standards can be found online (<http://www.jcsee.org/personnel-evaluation-standards>).

## Goals of Principal Evaluation Design

Principal evaluation systems should:

- **Be designed with the direct involvement of principals and other constituents.** Engaging leaders in the process builds trust and credibility for new evaluation systems and ensures that the evaluation process is feasible and useful to administrators.
  - **Be educative.** A principal evaluation system should provide useful, valuable, and trustworthy data to advance principals' abilities to be more effective leaders within their schools and communities.
  - **Be connected to district- and state-level principal support systems.** Principal evaluation should be considered one component of a broader approach to leadership development and should support leadership human capital management systems. Data arising from performance evaluation can be used to design professional development and induction systems, shape hiring procedures, improve working conditions, develop incentives, and inform other human resource processes that support leaders (see, for example, Teacher Leadership Exploratory Consortium, 2011).
  - **Be aligned, to the extent practicable, with teacher and other educator performance assessments.** Principals and other educators should be held to the same performance expectations in areas of common work.
- **Be rigorous, fair, and equitable.** The content, instruments, and administration of principal evaluation systems should be legal and ethical; allow for a thorough examination of principal practice; and be valid, reliable, and accurate.
  - **Include multiple rating categories to differentiate performance.** Evaluation should clearly identify principal performance levels.
  - **Gather evidence of performance through multiple measures of practice.** Evaluations should use multiple measures to provide a holistic view of principal performance.
  - **Communicate results to principals consistently and with transparency.** Principal evaluations are powerful to the extent that feedback can be used by principals to improve their work in schools and by district staff to make personnel decisions. Feedback should include all data from evaluations and should be clear, pointed, and actionable.
  - **Include training, support, and evaluation of principal evaluators.** New evaluation systems should be administered with consistency and fidelity, which requires that evaluators are trained, monitored, and supported.

## State Accountability and District Responsibility in Principal Evaluation Systems

Until recent policy changes were enacted, principal evaluation has largely been the responsibility of school districts. States, principal professional associations, and educational foundations have provided school districts with guidance on principal evaluation systems design. As a result of current federal initiatives, states are now increasingly responsible for establishing principal evaluation systems and monitoring principal workforce quality. Given the long history of local autonomy, many states and districts are challenged to create principal evaluation policies that encourage collective responsibility, mutual accountability, and systematic personnel evaluation while providing flexibility to ensure that evaluations reflect the changing dynamics and values of local schools.

This section describes statewide models for design and implementation of principal evaluation systems that have been identified through a literature review and discussions with state-level evaluation design teams. In addition, this section provides an overview of key roles and responsibilities for states and districts in the design and implementation of improved principal evaluation systems.

## Key State Roles

### Interpreting Federal and State Regulations

In response to the Race to the Top competition, federal incentive programs, and ESEA flexibility requirements, many state legislatures have passed new legislation on principal evaluation or examined current principal evaluation policies for compliance with federal reform goals and assurances. Federal and state legislation offers states varying degrees of flexibility to determine how principal evaluation should be designed and implemented and what design decisions can be made by school districts. As such, state departments of education, state-level design task forces, and other entities are responsible for interpreting legislation, designing evaluation processes, and implementing a statewide system of principal evaluation.

Interpreting state and federal legislation is a critical first step in developing a principal evaluation system, and state-level task forces can interpret policies in various ways. In some instances, stakeholders' interpretation can lead to increased variation within a state and can actually harm efforts to implement a consistent program or policy (Berman & McLaughlin, 1976).

State task forces also should recognize that districts will interpret policies as well. Accordingly, states should take proactive steps to help districts understand the spirit and intent of legislation and requirements for compliance and determine the best approach to principal evaluation design and implementation (see Component 2 for guidance on formulating a communication plan).

In addition to clarifying the state-level interpretation of federal and state legislation, state-level task forces can provide school districts and other stakeholders with implementation examples, case studies, and best practices. These examples can offer greater specification to intermediary organizations, school districts, and other entities on how best to implement principal evaluation systems, which is important in ensuring that all understand and operationalize legislation and administrative rules with some fidelity. Although each district has different capabilities and approaches to evaluation, best practice examples can help districts structure principal evaluation and hasten implementation.

## Setting the Design Agenda

Policies often couple principal and teacher evaluation system improvement together in the same policies and implementation timeline. Many states view principal and teacher evaluation as comprising a single educator evaluation system. For example, states and districts participating in the federal Teacher Incentive Fund (TIF) program and ESEA flexibility are required to improve both teacher and principal evaluation systems. Developing rigorous, fair, and equitable performance evaluation systems for principals and teachers helps to ensure that all school-level staff are evaluated annually.

States are responsible for determining the timing and timeline for principal and teacher evaluation system design. The timeline for evaluation systems design is often informed by legislation or federal program requirements, status of the current principal evaluation system, and capacity for design. States need to be familiar with design requirements and their interpretations and waiver or flexibility options.

States also are responsible for creating a coherent educator evaluation system that reflects similarities and differences in teacher and principal practices. Teacher and principal evaluation design processes should consider the unique work of teachers

and principals. The development of unique systems does not mean that principals' and teachers' work is not related or that the two evaluation systems cannot be mutually reinforcing. For example, both teacher and principal standards address "professionalism" and "ethical behavior," so both types of evaluation systems might use the same assessment language and measures for these standards. Similarly, states and districts may include measures of principals' evaluations of teachers as a means of supporting strong teacher evaluation systems.

## Design Approaches

As illustrated in Table 1, states have pursued three approaches for educator evaluation design, and each approach has its strengths and weaknesses.

- **Simultaneous Design.** Principal and teacher evaluation systems are designed at the same time but separately. A single "educator evaluation task force" might be convened to design both systems, or two separate task forces might work in parallel. Subcommittees can share ideas.
- **Principal-First Design.** A principal evaluation system task force is convened for the sole purpose of principal evaluation design prior to launching teacher evaluation system design.

- **Teacher-First Design.** A teacher evaluation system is convened for the sole purpose of teacher evaluation design prior to launching principal evaluation system design.

Available financial or human resources and politics factor into state decisions about the design agenda. No one approach to principal and teacher evaluation systems design is necessarily better than another.

**Table 1. Strengths and Weaknesses of the State Educator Evaluation Design Agenda**

Approach	Strengths	Weaknesses
Simultaneous Design	<p>Coordination of communication plan, implementation, and research timelines.</p> <p>Coordination of evaluation systems launch.</p> <p>Alignment of evaluation timelines within the school year.</p> <p>Conservation of resources because teacher and principal evaluation task forces may meet at the same event site and date.</p>	<p>There may be too much alignment between teacher and principal standards, measurement, and process.</p> <p>Simultaneous implementation of teacher and principal evaluation can overwhelm school districts.</p>
Principal-First Design	<p>Sends a message to teachers and others that evaluation applies to school leaders.</p> <p>Trains principals to be effective evaluators because they have experienced improved performance evaluation.</p> <p>Design and implementation are less demanding on the state and districts.</p>	<p>States and districts may have fewer resources available to design teacher evaluation later.</p> <p>State must support a communication and implementation plan for principal evaluation and then teacher evaluation.</p> <p>Policy may not allow state to design principal and teacher evaluations separately.</p>
Teacher-First Design	<p>Design and implementation are less demanding on the state and districts.</p>	<p>States and districts may have fewer resources available to design principal evaluation later.</p> <p>If teachers are not informed about the evaluation design agenda, they may question whether principals will be evaluated to the same degree.</p> <p>State must support a communication and implementation plan for principal evaluation and then teacher evaluation.</p> <p>Policy may not allow state to design principal and teacher evaluations separately.</p>

## Models for State and District Evaluation Systems

Research suggests that principal evaluation varies among schools, districts, and states and is largely dependent on local contexts for its design and implementation. However, federal guidance and policy have emphasized increased state responsibility for ensuring principal effectiveness and monitoring district principal evaluation practices. Each state must determine the appropriate level of involvement for these tasks and the roles that districts will play in ensuring effectiveness and monitoring. For example, some states may require adoption of a particular evaluation model and logistics (e.g., how often teachers are evaluated), format (e.g., selection of measures), and personnel decisions (e.g., what a rating means in terms of teacher tenure). Others may provide specific direction for adapting guidelines locally and implementing a system.

States' decisions about roles and responsibilities will vary according to state politics, district capacity, state size, goals, and support infrastructure. Decisions also will vary depending on whether or not the state requests ESEA flexibility. Some states, like Tennessee, use a statewide evaluation system and have submitted an ESEA flexibility request. Other states that have submitted an ESEA flexibility request, like New York, are likely to allow districts to choose an evaluation model.

In other states, like Illinois, districts will be allowed to use their own evaluation systems so long as they meet certain requirements.

The following subsections discuss three models for state implementation: state-level evaluation system, elective state-level evaluation system, and district evaluation system with required parameters. Note that this list of options is not exhaustive and that a state may create a hybrid of two or more models. Also, the model adopted for the principal evaluation system may or may not be applied to the teacher evaluation system.

## State-Level Evaluation System

The state-level evaluation system strictly interprets legislation and prescribes the requirements for principal evaluation models. The state determines the components of the evaluation model, the measures to be used, and the administration of evaluations. The state may require that districts use a single evaluation model, as in the case of Tennessee, or use multiple state-approved evaluation models, as is likely the case in Washington.

Tennessee is currently implementing a single, statewide principal evaluation model across all school districts within the state (see “Practical Example: Tennessee Evaluation Model” at right). According to Tennessee task force members, Race to the Top prompted state redesign of principal evaluations. Tennessee’s principal

evaluation design process engaged state-level administrators, district superintendents, school principals and their professional associations, and teachers in the design and implementation of the state model. The state has adopted a single model, which includes value-added measures of student performance as a significant portion of principals’ evaluations.

## Elective State-Level Evaluation System

The elective state-level evaluation system may strictly interpret state and federal legislation and require districts to adopt certain aspects of an evaluation system but allows local discretion on other aspects of the system. For example, state legislation may require that student growth be a significant factor in a principal’s summative performance evaluation but may provide districts latitude in setting the percentage of a principal’s summative score that is based on student growth. The state also may provide districts with flexibility on the standards to be measured by requiring all principal evaluations to address a core set of standards but allowing districts to add standards to reflect district initiatives and values. Colorado, for example, requires districts to adopt seven principal quality standards and associated elements but does not mandate a specific leadership rubric describing performance levels, nor does it prohibit districts from adding standards.

### PRACTICAL EXAMPLE

#### Tennessee Evaluation Model

All Tennessee principals must be evaluated using the state’s model based on the Tennessee Instructional Leadership Standards (TILS). In April 2011, the State Board of Education adopted regulations establishing five levels of principal performance and multiple performance measures with weights:

- ▶ School-level value-added measure (TVASS) (35 percent)
- ▶ Student achievement data (15 percent)
- ▶ Qualitative scores on TILS rubric (includes school climate surveys) (35 percent)
- ▶ Quality of teacher evaluations (15 percent)

Tennessee also requires two annual, on-site observations (announced and unannounced) and provides a list of approved measures for student achievement as well as school climate and working conditions surveys.

Source: Tennessee Department of Education (2011)

In the elective state-level evaluation system model, the state has a major role in establishing a core principal evaluation model and ensuring that districts comply with core elements of the model (see “Practical Example: Colorado’s Elective State-Level Evaluation System” on page 12). The state-level evaluation system model also may allow districts or regions within the state to adjust the model or add to the statewide

## PRACTICAL EXAMPLE

### Colorado's Elective State-Level Evaluation System

In 2010, the Colorado Legislature passed SB 10-191, requiring all districts to adopt new teacher and principal evaluation systems by 2014–15. The legislation established a common definition of principal effectiveness, seven principal quality standards, and the following requirements:

- ▶ Schoolwide student growth scores must account for 50 percent of the final score.
- ▶ Evaluation must occur annually.
- ▶ Results must be used in human resource decisions.
- ▶ Principals ranked “unsatisfactory” must be provided professional development and support to improve.

The Colorado Department of Education has developed a model system for principal evaluation that districts can adopt or adapt. The model system includes rubrics, forms, and guidance on selecting measures. The department has not decided whether the state model will be the “default” model; districts, however, will have the option of developing their own principal evaluation systems that meet state requirements.

Source: Colorado Department of Education (2011)

model. This option allows districts to adapt the statewide model to local contexts and values in ways that maintain the integrity of the statewide model. The option also allows districts to continue to use aspects of their current principal evaluation systems.

## District Evaluation System With Required Parameters

In some cases, a statewide principal evaluation system is impractical and inappropriate. Still, states may wish to provide school districts with guidance on principal evaluation design, compliance with implementation regulations, and state-level priorities. In this case, districts influence district-led development of principal evaluation and other support mechanisms. For example, some states offer districts guidance on principal evaluation design and federal programs, provide access to state-developed rubrics, and identify instruments that may be useful to districts. School districts must determine how state-provided guidance is used to design better evaluation and other professional support systems for principals.

In the district evaluation system model, the state also may review and approve proposed principal evaluation systems prior to implementation. This state role helps to ensure that districts comply with applicable legislation and administrative rules and provides for future state-level audits of

district evaluation systems. Typically, such audits are preceded by published evaluation system criteria or other information so that districts can design evaluation systems in ways that comply with state-level standards.

For example, an Illinois state-level task force has proposed that districts use the state-level model but also allows districts to submit locally developed principal evaluation models for review by a state committee. If the state committee finds that the district's principal evaluation system meets quality criteria, the district can continue using the locally developed system. If the district evaluation system does not meet quality criteria, then the district needs to make changes to the evaluation system or adopt a statewide model. (See “Practical Example: Illinois District Evaluation System Model” on page 13.)

## Factors for Stakeholder Consideration

Stakeholders might consider the following factors in selecting a particular model:

- ESEA flexibility requirements as applicable
- Grant requirements as applicable, such as Race to the Top, School Improvement Grants (SIG), Teacher Incentive Fund (TIF)
- Existing or impending state legislation
- Goals and priorities at the state and district levels



- Traditional, state-level role in district practice
- Principal professional association guidelines
- Number and diversity of districts within a state
- Variation in job descriptions of principals in the state
- Capacity for long-term support of principal evaluation design and implementation
- Training or certification of staff needed to implement the system with fidelity and which organizations will provide training
- Stakeholder support for principal evaluation system improvement
- Teachers' and administrators' preferences for certain types of measures
- Prevalence of accepted, rigorous professional standards at the district and local levels

**Note:** Race to the Top, ARRA, and ESEA flexibility indicate that total district-level control with no state-level involvement or accountability is not supported at the federal level.

As the preceding text suggests, no best approach to principal design and implementation exists. State and district design teams must determine the appropriate course of action in light of state or district history, capacity, legislation, administrative rule, and tradition. Table 2 summarizes the strengths and weaknesses of the models presented in this section.

## PRACTICAL EXAMPLE

### Illinois District Evaluation System Model

By 2012–13, all districts in Illinois must evaluate their principals according to new requirements passed by the Legislature in 2010. The state provides a model principal evaluation system; districts, however, have the option to develop their own models and submit them for state approval. The Illinois State Board of Education has proposed the following requirements for all approved models:

- ▶ Student growth must be a “significant factor” in every evaluation.
- ▶ Evaluation of principal practice must account for 50 percent of a principal’s final score.
- ▶ Student growth must be measured using data from two assessment types.
- ▶ Annual evaluation must include two formal observations or site visits.
- ▶ There are four levels of performance.

Unlike the Illinois teacher evaluation model, the state does not require districts to use the state’s default model for student growth for principal evaluation.

*Source:* Illinois State Board of Education (2012)

**Table 2. Strengths and Weaknesses of Principal Evaluation Models**

Model	Strengths	Weaknesses
State-Level Evaluation System	<p><b>Design</b></p> <ul style="list-style-type: none"> <li>■ Sets statewide measures and dimensions</li> <li>■ Allows for coherence between state-level frameworks and measures</li> </ul> <p><b>Evaluator Training</b></p> <ul style="list-style-type: none"> <li>■ Provides conditions for standardized, statewide evaluator training and certification</li> <li>■ Allows for comparison of evaluator severity and reliability</li> </ul> <p><b>Data Collection</b></p> <ul style="list-style-type: none"> <li>■ Facilitates standardized data collection process and timeline</li> <li>■ Increases ability to change system from year to year</li> </ul> <p><b>Use</b></p> <ul style="list-style-type: none"> <li>■ Facilitates the determination of statewide system efficacy and impact</li> <li>■ Eases statewide use of data for principal preparation program design</li> <li>■ Eases statewide coordination of principal professional development programs</li> </ul>	<p><b>Design</b></p> <ul style="list-style-type: none"> <li>■ Does not easily accommodate local leadership context (e.g., goals, mission, vision, school status)</li> <li>■ Diminishes local ownership</li> </ul> <p><b>Data Collection</b></p> <ul style="list-style-type: none"> <li>■ Does not accommodate variance in district human and financial resources to consistently evaluate principals</li> </ul>
Elective State-Level Evaluation System	<p><b>Design</b></p> <ul style="list-style-type: none"> <li>■ Provides for some flexibility on design</li> <li>■ Allows for some continuation of local evaluation designs</li> <li>■ Allows for some accommodation of local contexts (e.g., goals, mission, vision, school status)</li> <li>■ Increases local ownership</li> <li>■ Allows for coherence with state-level frameworks and measures</li> </ul> <p><b>Evaluator Training</b></p> <ul style="list-style-type: none"> <li>■ Provides conditions for some standardized evaluator training and certification</li> <li>■ Allows for comparison of evaluator severity and reliability for some components</li> <li>■ Facilitates data collection</li> <li>■ Provides for standardized data collection on some components</li> </ul> <p><b>Use</b></p> <ul style="list-style-type: none"> <li>■ Facilitates evaluation of statewide performance evaluation system efficacy and impact on common aspects of the evaluation system</li> <li>■ Provides for some use of data for principal preparation and professional development program designs</li> </ul>	<p><b>Design</b></p> <ul style="list-style-type: none"> <li>■ Requires states and districts to expend resources on systems design</li> </ul> <p><b>Evaluator Training</b></p> <ul style="list-style-type: none"> <li>■ Requires states and districts to support evaluators</li> <li>■ Possibly does not allow for state certification of evaluators</li> </ul> <p><b>Data Collection</b></p> <ul style="list-style-type: none"> <li>■ Requires dual file management systems</li> <li>■ Diminishes monitoring of state-level compliance</li> </ul> <p><b>Use</b></p> <ul style="list-style-type: none"> <li>■ Makes aggregating state-level data more challenging</li> <li>■ Makes coordinating principal professional development programs at state level more difficult</li> <li>■ Complicates administration of the statewide performance evaluation system</li> </ul>

Model	Strengths	Weaknesses
District Evaluation System With Required Parameters	<p><b>Design</b></p> <ul style="list-style-type: none"> <li>■ Increases local ownership</li> <li>■ Provides for local flexibility</li> <li>■ Allows for continuation of local evaluation designs</li> <li>■ Allows for accommodation of local contexts (e.g., goals, mission, vision, school status)</li> </ul> <p><b>Use</b></p> <ul style="list-style-type: none"> <li>■ Facilitates evaluation of statewide performance evaluation system efficacy and impact on common aspects of the evaluation system</li> </ul>	<p><b>Design</b></p> <ul style="list-style-type: none"> <li>■ Requires some mechanism for assuring alignment and coherence with state frameworks and measures</li> <li>■ Requires district-level reliability, validity measurement</li> <li>■ May not appear fair to principals because evaluation requirements may differ</li> </ul> <p><b>Evaluator Training</b></p> <ul style="list-style-type: none"> <li>■ Requires districts or regions to train and support evaluators</li> <li>■ Requires districts or regions to determine rater reliability and severity</li> </ul> <p><b>Data Collection</b></p> <ul style="list-style-type: none"> <li>■ Does not necessarily provide for data collection coherence or timelines</li> </ul> <p><b>Use</b></p> <ul style="list-style-type: none"> <li>■ Makes aggregating data challenging, sometimes impossible</li> <li>■ Complicates administration of the statewide performance evaluation system</li> </ul>

## Development and Implementation of Comprehensive Principal Evaluation Systems

The following subsections describe essential components and critical phases of the principal evaluation system design process:

- Component 1a: Specifying Evaluation System Goals
- Component 1b: Defining Principal Effectiveness and Establishing Standards
- Component 2: Securing and Sustaining Stakeholder Investment, and Cultivating a Strategic Communication Plan
- Component 3: Selecting Measures
- Component 4: Determining the Structure of the Evaluation System
- Component 5: Selecting and Training Evaluators
- Component 6: Ensuring Data Integrity and Transparency
- Component 7: Using Principal Evaluation Results
- Component 8: Evaluating the System

Each subsection discusses the importance of the component and includes a series of questions to guide principal evaluation design. Components and questions were identified by the authors through their work with state and district principal and teacher evaluation system design committees.

### COMPONENT 1a

#### Specifying Evaluation System Goals

Specifying the goals of a principal evaluation system is a critical first step. Explicit, well-articulated goals will drive the evaluation design. They provide the foundation for developing and maintaining a comprehensive principal evaluation system because they offer guidance to designers on what the evaluation system should and should not do. In addition, clear system goals help stakeholders gain an understanding of the evaluation system and provide researchers a basis for evaluating system performance.

Although federal and state legislation provide some guidance on principal evaluation system goals, ESEA flexibility requires that states do the following:

- Develop coherent and comprehensive systems that support continuous improvement.
- Customize the systems to the needs of the state, its districts, its schools, and its students.
- Improve educational outcomes, close achievement gaps, increase equity, and improve the quality of instruction.

Discussions of principal evaluation goals may be informed by the ESEA flexibility core policies.

In other circumstances, system designers are often left to define system goals on their own. In-depth conversation and agreement among stakeholders are critical to the design effort. Each designer likely brings his or her opinions about personnel evaluation and principal performance to the table, and these opinions shape decisions about standards, measures, and implementation. Explicitly stated goals add clarity to the group process.

State-level committees often recognize that the intent of principal evaluation is to improve the quality of teaching and learning, but additional system goals can be articulated to show a connection between principal evaluation and the ultimate goal of better instruction and student progress. In Wisconsin and other states, evaluation designers have crafted a theory of action that draws connections between the principal evaluation system and improvements in principals' work, school climate, community relations, teacher quality, instruction, and student learning.

#### Principal Evaluation Goals

The following goals for principal evaluations are based on research and interactions of the Center on Great Teachers and Leaders with states and school districts. (The first two goals relate to either formative or summative evaluation; see sidebar on page 17.)

States and districts may emphasize one or more of these goals:

- **Improve principal practice (formative).** Principal evaluation systems provide credible evidence and actionable feedback on school principal performance, which can be used by principals to improve their practice. The evaluations measure principal effectiveness and are intended to inform professional development improvement and growth.
- **Inform decisions about principal competency (summative).** Principal evaluation systems provide district leaders with evidence of principal performance, which can be used for decisions about job retention, advancement, and compensation.
- **Articulate state or district goals.** Principal evaluation systems define state and district educational improvement priorities through the selection and weighting of competencies.
- **Support teacher growth and evaluation.** Principals can play a pivotal role in evaluating teachers and creating conditions amenable to teacher learning. Principal evaluation systems can reinforce the importance of principals' roles in teacher accountability and professional learning as well as compliance with teacher evaluation practices.

- **Present a coherent vision of educator professional responsibilities.** Many districts and states view principal and teacher evaluations as supporting a common set of educator knowledge, skills, and attitudes, while recognizing differences between professional classifications.

States and districts may emphasize one or more of these goals, but selection of goals informs evaluation system design. For example, if improvement of principal practice is emphasized, the principal evaluation system should include methods of connecting evaluation results to principal professional development planning or decisions about professional development offerings in the state or district. If the goal is more high stakes, the principal evaluation system should establish the psychometric rigor of evaluation measures to ensure that the system is technically and legally defensible.

Principal evaluation system goals can be established by drawing upon the opinions of the design team but also may be informed by other sources. For example, Hazelwood School District (Missouri) conducted a districtwide survey and focus groups with school principals to get stakeholder input on goals selection. State-level teams also may review current state and district initiatives and programs when selecting system goals because doing so supports systemwide coherence and support.

### FORMATIVE VERSUS SUMMATIVE EVALUATION: WHAT IS THE DIFFERENCE?

When speaking about performance evaluations for principals or other educators, formative and summative purposes are highlighted. A single assessment may be used for both formative and summative purposes.

A *formative* evaluation measures competency, and results are used to inform future actions. For example, formative performance assessments may be used to inform principal professional development plans.

A *summative* evaluation informs decisions about overall competence and does not provide opportunities for improvement or remediation after completion.

---

Stakeholders might consider the guiding questions for Component 1a as they work to develop the overall vision and goal of the evaluation system.

## Guiding Questions for Component 1a

### Specifying Evaluation System Goals

#### SYSTEM GOALS AND PURPOSES

1. Have the goals and purposes of the evaluation system been determined?

#### GUIDING QUESTIONS

- What purposes will the evaluation system address (e.g., improved principal practice, competency decisions, articulating state or district goals, support teacher evaluation, establish a coherent vision)?
- What types of effects will the improved principal evaluation system achieve (e.g., improved leadership practices, school conditions, instructional quality, student achievement)?
- What do school principals, superintendents, and others within the state believe should be the goals of principal evaluation and how pervasive are these goals?
- What educational policies, programs, and initiatives may be influenced by principal evaluation design (e.g., school improvement planning, principal certification)?

NOTES

#### GOAL DEFINITION

2. Are the goals explicit, well-defined, and clearly articulated for stakeholders?

#### GUIDING QUESTIONS

- To what degree are goals stated in measurable terms (e.g., learning improvement, closing achievement gaps)?
- To what degree are goals written to represent the opinions and perspectives of multiple stakeholder groups in clear, concise language that is accessible by all?
- To what degree are the relationships between principal evaluation system goals clearly articulated?
- Are the system goals acceptable to stakeholders?

## GOAL ALIGNMENT

3. Have the evaluation goals been aligned to the state strategic plan, the principal evaluation system design communication plan, principal preparation or professional development initiatives, and pertinent school improvement initiatives?



## GUIDING QUESTIONS

- How can principal evaluation system goals align with other initiatives to create more coherence among human capital support systems for school leaders in this state?
- How can principal evaluations align with teacher evaluations so that educator evaluation is more coherent and the two systems are mutually supportive?
- To what degree will districts have flexibility and input in state-level goals and designs?

## COMPONENT 1b

### Defining Principal Effectiveness and Establishing Standards

After the goals and purposes of an evaluation system are established, states and districts should align goals to principal professional standards. The task often begins by defining the term *effective principal*. This definition may differ from the definition of the term *principal quality*, which tends to focus on training, knowledge, or attitudes held by a principal. Principal effectiveness focuses on principal practices and achieved outcomes.

After the term *effective principal* has been defined, professional standards can be aligned to that definition. Many states have adopted professional standards for use in principal certification, hiring, and evaluation. Standards are the basis for definitions of desired performance, and the rating scale by which principal performance can be assessed. (See the Glossary in Appendix A for a definition of *effective principal* and other terms related to principal evaluation.)

National principal professional standards have been painstakingly developed through extensive review processes by principal professional associations and other associations and adopted by states into

certification program review, certification and accreditation requirements, and other legislation or administrative rules. However, principal evaluation systems often are not aligned with state or national professional standards (Goldring et al., 2009). Race to the Top, School Improvement Grants, and other federal initiatives require educator evaluation measures to be aligned with standards of professional practice. States and districts should refer to these standards.

Although many standards for principal practice are available through national policy associations and research organizations, there are few standards to guide principal performance evaluation. Existing standards may not be written in observable or measurable terms—a necessity for principal evaluation—or may not cover the wide breadth of principals' work. States and districts must critically review standards to ensure that:

- Selected standards align with the definition of principal effectiveness.
- Essential or “core” standards to principals' work are addressed by the evaluation system.
- Standards and indicators are written in observable and measurable terms.

In addition to principal professional standards, states and districts may review teacher, teacher leader, and other educator standards for alignment with principal standards. Such a review is important when the evaluation system designers' goal is to facilitate a coherent vision of educator professional practice or principal support of teacher evaluation and learning.

For example, both teacher and principal standards address professional practices and ethics, and designers could examine these standards for alignment. Although evaluation system designers may be tempted to adapt or change standards, alterations to standards language should be made with caution. Standards have been painstakingly written and vetted, but the standards also must be written in observable and measurable terms to facilitate performance assessment.

---

Stakeholders might consider the guiding questions for Component 1b as they define principal effectiveness and develop or revise principal standards.



### Defining Principal Effectiveness and Establishing Standards

#### DEFINITION OF EFFECTIVE PRINCIPAL

1. Has the state defined what constitutes an effective principal?

#### GUIDING QUESTIONS

- Is the state's definition of an effective principal or a highly effective principal consistent with accepted definitions of principal effectiveness?
- Does the definition of principal effectiveness include language about the growth of students or student populations that have historically underperformed on national, state, or local tests?
- How, if at all, will the definition of principal effectiveness reflect differences in organizational level (i.e., elementary, middle, high), school context, or previous school performance?
- Is the definition of effectiveness observable and measurable?
- Will the definition of effectiveness account for professional practice, school performance, teacher support and performance, and community perspectives on leadership, in addition to student achievement?
- How compatible is the definition of principal effectiveness with the state or district definition of teacher, teacher leader, or other educator effectiveness?

NOTES

#### PRINCIPAL STANDARDS

2. Has the state established principal standards in law, statute, or rule?

#### GUIDING QUESTIONS

- Has the state or district adopted principal standards?
- Are the state standards aligned with the definition of principal effectiveness?
- Which standards are considered essential and will be adopted into principal evaluation design?
- Are the adopted standards observable and measurable, or will indicators need to be articulated? To what degree are principal standards accepted by professional associations, principal preparation programs, and other pertinent entities in the state?
- How, if at all, are principal standards aligned with teacher standards so that they mutually support educator effectiveness?
- Are the standards free of "high inference" language or jargon that makes them prone to misinterpretation?
- Have indicators been developed and operationalized into at least four levels of performance, or must the committee do this work?

## COMPONENT 2

### Securing and Sustaining Stakeholder Investment, and Cultivating a Strategic Communication Plan

#### The Importance of Stakeholder Investment

Evaluation systems are much more likely to be accepted, successfully implemented, and sustained if stakeholders are included in the design process. Stakeholder involvement throughout the design, implementation, assessment, and revision of principal evaluation systems increases the likelihood that the system is perceived as responsive, useful, and fair. In addition to building buy-in, involving stakeholders can significantly improve the quality of the product created through the incorporation of their diverse ideas and knowledge. States submitting ESEA flexibility requests are required to describe how they have meaningfully engaged and solicited input from principals, principal representatives, and diverse communities. State flexibility requests are strengthened when they note specific changes based on feedback.

Stakeholder engagement early in the process provides an opportunity to build awareness about the need and reason for the desired change. The process and the outcome will

benefit when stakeholders come to a conclusion that change is needed after they have received balanced information about the deficits of the system.

A stakeholder group or steering committee could include the following:

- Principals and principal association representatives
- Teachers and other school personnel and their association representatives
- School board members
- Superintendents and human resource directors
- Principal preparation program leaders
- Parents
- Students
- Business and community leaders

Some of these stakeholders may have higher priority than others, and their relative value can shift depending on the stage of the design process. In the case of principal evaluation system reform, it is imperative to have principals and teachers at the table throughout the process. Involving educators in the initial stages of development and throughout the implementation process will likely increase educators' collaboration, support, and promotion of state and district efforts and will lead to a system that works in practice.

#### RESOURCE

##### *Communication Framework for Measuring Teacher Quality and Effectiveness: Bringing Coherence to the Conversation*

(<http://www.gtlcenter.org/sites/default/files/docs/NCCTQCommFramework.pdf>)

This framework can be used by regional comprehensive center staff, state education agency personnel, and local education agency personnel to promote effective dialogue about the measurement of educator quality and effectiveness. The framework consists of the following four components: communication planning, goals clarification, educator quality terms, and measurement tools and resources. Although this framework was prepared with teacher evaluation reforms in mind, many of the takeaways are applicable to principal evaluation.

#### Tips for Managing Stakeholder Engagement

Sustaining stakeholder investment often requires that expectations for involvement, level and duration of commitment, and levels of authority be clear. Individual skills, experiences, and interests should be carefully considered when assigning responsibilities and tasks.

Stakeholders and other thought partners could play an integral role in the following tasks:

- Determining system goals and effectiveness definitions
- Informing state or district approaches to design, systemic support, change, and improvement
- Determining the standards and criteria for the system
- Mobilizing support for a redesigned evaluation system
- Seeking feedback and input from practitioners and other groups to ensure that the evaluation system meets expectations for quality and feasibility
- Marketing the system and publicizing the findings emerging from system testing
- Interpreting policy implications
- Investigating and securing federal, state, or private sector funding

## Communication Plan

Communication needs should be considered early in the process. A strategic communication plan detailing steps to inform the broader school community about implementation efforts, results, and future plans may increase the potential for statewide adoption. Misperceptions and opposition can be minimized if the state and districts communicate a clear and consistent message.

A strategic communication plan first identifies the essential messages and audiences. Potential key audiences could include pilot participants, school personnel, families, and the external community. The stakeholder group supporting the planning process can help determine the most effective channel of communication for a particular purpose and target audience. Written, spoken, and electronic communication strategies may include the following:

- Online communications
- Community information nights
- Quarterly memos
- Weekly e-mail updates
- Media relations materials
- Word of mouth
- Events
- Workshops
- Videos
- CDs
- Press releases
- Newsletters

The communication plan for principal evaluation should be well-aligned with the communication plan for teacher evaluation so that stakeholders perceive the systems as compatible and mutually supportive. Enacting similar communication plans for teacher and principal evaluation systems improvement also can be more financially efficient.

Principals' work schedules and preferred methods of communication should be considered when creating a communication strategy. Many school principals report that they work 60 or more hours per week and are connected to multiple, Web-based information sources. Principals also are expected to work in schools with teachers and outside of schools with district staff and community members. A communication plan should be informed by principals' preferred mode of receiving information.

Communication plans should take into account the duration of the process of improving the evaluation system, including its initiation and all implementation phases. For example, communication needs during the design of the system will be different from those during implementation and the process of gathering feedback. Plans should include updates on efforts to build the evaluation system, celebrations of successes as the work moves forward, and recognition of stakeholder contributions.

Communicating success in terms of implementation efforts, changes in educator practice, and student outcomes can be a powerful way to ensure buy-in and secure stakeholder investment. Highlighting successes also reinforces, inspires, and energizes educators. Plans should make the design process transparent for stakeholders; transparency is important for managing politics associated with redesign.

## Considerations for Stakeholder Communication

When developing communication plans, the design committee can anticipate some critical issues related to principal evaluation reform. The following issues frequently emerge in districts and states engaging in these types of reforms and can be addressed through strategic communication planning:

- **Context.** Principals are concerned that the evaluation system does not take into account the unique context of the school or its performance history, which is the basis for their priorities and leadership approaches. Differences in school context may prompt principals to appropriately prioritize certain leadership actions or traits over others; but in doing so, these principals may be “marked down” on the evaluation form. Communications about the new evaluation should explain how these differences will be taken into account, either through the state model or the weighting system.
- **Differentiation.** Principals are wary of a one-size-fits-all approach that might not take into account the differing roles and responsibilities of school leaders at elementary, middle, and high schools, or other types of schools in the public education system. At the state level, the differences between urban and rural contexts are of particular concern.

At the district level, principals may point out the distinctions between elementary and secondary school contexts. Communications about the new system should cover how these differences will be taken into account.

- **Subjectivity.** For any system that includes measures based on individual judgment (e.g., observations, surveys, and interviews), subjectivity will be a concern. Communications should detail the steps that will be taken to make all measures as fair and consistently applied as possible (e.g., evaluator training and system monitoring).
- **Student Outcomes.** In districts and states undertaking evaluation system reform, student achievement outcomes will be considered for the first time or to a higher degree than in the past. Communications should be clear about how these outcomes will be incorporated and their relative weight to other measures. However, with the increased focus on school accountability during the last decade, many principals already may feel as if they are held accountable for student outcomes.
- **Accountability/Authority Balance.** Principals may be concerned about being held accountable for factors that are beyond the reach of their authority (e.g., an evaluation system that holds principals accountable for the actions of teachers in cases in which principals’ have little or no authority in the hiring

and removal of teachers or a system that addresses fiscal responsibility in areas in which principals have little budgetary control). Communications should make clear that principals will be evaluated using fair and appropriate measures that consider the principals’ decision-making authority.

- **Burden.** As principals’ roles and responsibilities evolve with a new focus on instructional leadership, principals are responsible for completing more tasks than ever before. An improved principal evaluation system may be perceived as an increased burden on principal time. In addition, systems engaged in principal evaluation reform often are implementing teacher evaluation reforms that fall on the shoulders of principals. Principals want meaningful, actionable feedback and a fair evaluation without experiencing increased workload. Communication should highlight these concerns for those designing the system.

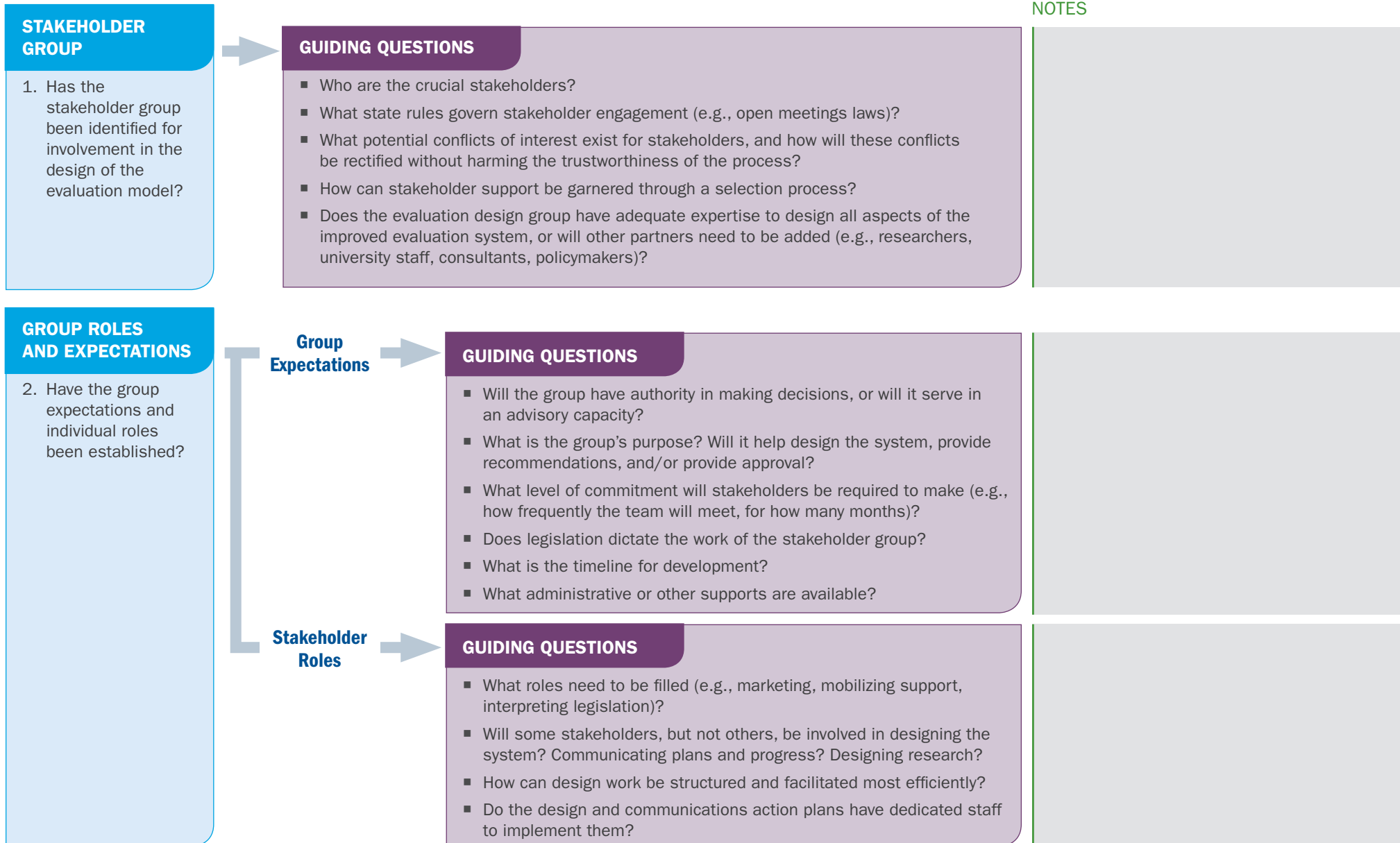
The design committee for a particular state or district should work to identify other issues that may emerge given unique historical or contextual factors.

---

Stakeholders might consider the guiding questions for Component 2 as they develop a strategic communication plan.

## Guiding Questions for Component 2

### Securing and Sustaining Stakeholder Investment and Cultivating a Strategic Communication Plan



**COMMUNICATION PLAN**

3. Does the group have a strategic communication plan to keep the broader school community informed?

**Content** →

**GUIDING QUESTIONS**

- What key messages need to be communicated?
- How will the communication plan gather and address common concerns about principal evaluation system design?
- How will progress on the design, implementation, and success of the evaluation system be shared?
- How will principal evaluation system results (e.g., satisfaction with implementation, fidelity of implementation, increased performance of principals, schools) be communicated, when, and by whom?

**Target Audience** →

**GUIDING QUESTIONS**

- Which target audiences should be kept informed about the development, implementation, and results of efforts related to principal evaluation?
- How will communication efforts be varied according to audience (e.g., board members require more detailed updates than community members)?
- How can existing methods of communication be leveraged?
- Who will be responsible for communicating with constituents and task force members?

**Timing** →

**GUIDING QUESTION**

- Does the plan include communication strategies throughout the development process (e.g., in the beginning, during, and after each phase)?

## FEEDBACK

4. How will feedback be gathered to continuously improve evaluation system design?

Who

### GUIDING QUESTIONS

- From whom does the group wish to solicit feedback?
- At what points in the design process should feedback be solicited?

Methods

### GUIDING QUESTIONS

- What methods will be used to obtain feedback from affected school personnel during the design process (e.g., surveys, focus groups)? How formalized should feedback be?
- What are the indicators of strong system performance?
- How will data on system performance be gathered, represented, and used?
- What resources are currently available to gather information about system design satisfaction and system performance?
- How should feedback be delivered and to whom? How, if at all, will feedback be communicated to stakeholders?
- Will the state and district hire an impartial external evaluator?

Response

### GUIDING QUESTIONS

- How will the group respond to feedback (e.g., Q&A document, FAQ newsletter)?
- Will student outcomes be considered before changes are considered?

## COMPONENT 3

### Selecting Measures

The principal evaluation system purposes and standards should clearly define the types of practices and outcomes that will be assessed by the evaluation system, and measures should be selected accordingly. Measures are the methods that evaluators use to determine principals' levels of performance. Principal evaluation approaches typically include measures of principal practice (i.e., the quality of principals' performance on certain tasks or functions) and outcomes (i.e., anticipated impact on schools, teaching, and students). Selecting or, if need be, developing appropriate measures is essential to evaluation system design.

System design should carefully balance feasibility and fidelity of implementation with validity and reliability issues. Further, an evaluation system can become burdensome for principals, teachers, and evaluators if it attempts to measure too much but can be viewed as invalid if it measures too little. A cumbersome and costly evaluation system will likely face challenges to strong fidelity of implementation.

Current federal definitions of principal effectiveness focus on the use of valid and reliable measures of practice and outcomes. The Race to the Top guidance, for example,

requires states to develop evaluation systems that “differentiate effectiveness using multiple rating categories that take into account data on student growth . . . as a significant factor” (U.S. Department of Education, 2010, p. 34). Race to the Top and Teacher Incentive Fund (TIF) guidance to grantees also stresses the importance of using multiple measures to provide a holistic picture of principal performance, and TIF grantees must include principal observation as one measure of principal performance. ESEA flexibility requires specificity on processes that states use for determining the validity and reliability of the evaluation measures and how those measures will consistently be used across districts.

At this time, research and policy have not suggested a certain number of measures that should comprise a principal evaluation system. Federal regulations for discretionary grant participation (e.g., Race to the Top, TIF, SIG) require that evidence of student learning be a “significant” component of principal evaluation.

States and districts must determine which outcomes and practice measures are most applicable and useful to the purposes of the principal evaluation system. Decisions about outcomes and practice measures should be informed by the degree to which principals

### RESOURCES

#### *Evaluating School Principals (Tips & Tools)*

[http://www.gtlcenter.org/sites/default/files/docs/KeyIssue\\_PrincipalAssessments.pdf](http://www.gtlcenter.org/sites/default/files/docs/KeyIssue_PrincipalAssessments.pdf)

This Tips & Tools document summarizes approaches to principal evaluation design, highlights challenges to evaluation implementation, and identifies state and district examples of strong implementation. Extensive resources and links to programs are provided so that readers can access case examples.

#### *Guide to Evaluation Products*

<http://resource.tqsource.org/GEP/>

This guide can be used by states and districts to explore various evaluation methods and tools that represent the “puzzle pieces” of an evaluation system.

The guide includes detailed descriptions of more than 25 principal evaluation tools that are currently used in districts and states throughout the country. The following information is provided for each tool:

- ▶ Research and resources
- ▶ Appropriate populations for assessment
- ▶ Costs, contact information, and technical support offered



have control over outcomes and research on principal effects. Often, states and districts use the average of all teacher value-added or growth scores in a given school as a factor in principal evaluation, although some policymakers and constituents have raised concerns about the validity of this approach. Some measures of principal outcomes include, but are not limited to, the following:

- Student growth measures
  - Value-added models
  - Student achievement trends
  - Percentage of student learning objectives achieved in a school
  - Locally or regionally used subject-specific test results
- Instructional quality measures
  - Teacher placement indicators (e.g., placement in subject area in which teachers are certified)
  - Teacher retention rates
  - Specific measures of instructional quality
- School performance measures
  - Student behavior measures (e.g., attendance, attrition, behavioral incidents)
  - School climate measures
  - Community participation, interaction, and satisfaction measures

- Progress on school improvement plans
- Progress on school fiscal management plans (as applicable)

Principal practice measures capture the quality of principals' leadership and administrative practices and provide rich data on practice. In the hands of well-trained and experienced principal evaluators, practice measures data can be a source of useful feedback on what principals can do to improve their work, schools, and student learning. Potential principal practice measures include the following:

- Observation instruments (e.g., observations of principal and teacher evaluation practices or data presentations)
- Parent, student, or teacher surveys
- 360-degree surveys
- Portfolios or evidence binders
- Principal professional development plan achievements or evidence of learning

Given the breadth of principals' work, no single measure can provide a holistic picture of principal practice, and each measure has inherent strengths and weaknesses.

Factors in selecting or designing measures should be guided by the following factors:

- Strength of measures

- Application to student populations and leadership contexts
- Human and financial resource capacity

The following subsections briefly describe each of these factors. States and districts should ensure that the design process includes adequate technical expertise and materials to ensure that measures meet the criteria.

## Strength of Measures

All measures have inherent strengths and weaknesses. Validity, reliability, feasibility, utility, and fairness are critical to selecting measures (see "Important Terms for Selection of Measures and Methods" on page 30). Not all measures have sufficient evidence to ensure that they are fair, reliable, research-based, and valid, but the committee should review and retain available research to provide stakeholder evidence of technical soundness. When selecting or designing measures of principal performance, states and districts should have adequate technical expertise to ensure that measures are sufficiently technically defensible and provide actionable feedback.

Student growth measures are particularly concerning to educators, parents, and policymakers and are used in principal evaluation. Federal priorities provide guidance on student growth measures,

stipulating that such measures need to meet the following requirements (Secretary's Priorities for Discretionary Grant Programs, 2010):

- Rigorous
- Between two points in time
- Comparable across classrooms

Student growth measures also must be fair, valid, and reliable for their intended purposes and must include methods for attributing results to individual teachers and

#### IMPORTANT TERMS FOR SELECTION OF MEASURES AND METHODS

**Validity:** A measure that focuses on an assessment's ability to measure what it is intended to measure for prescribed purposes.

**Reliability:** A measure of consistency and stability of a given instrument or rater. Measures are said to be reliable when responses are consistent and stable for each individual who is assessed.

**Feasibility:** A sense that a measure or measures can be implemented as prescribed, given financial, human, or other constraints.

**Utility:** Evidence that a measure provides actionable feedback, which is information that principals can use to make changes in practice.

**Fairness:** Evaluation measures and methods should be consistently administered to principals (in a given population) by trained staff and held to similar standards.

principals (Herman, Heritage, & Goldschmidt, 2011). ESEA flexibility requires that state plans include measures that the state intends to use to evaluate teachers of nontested grades and subjects. (Appendix B provides an overview of measures including descriptions, research base, strengths, and cautions.)

### Application of Measures to All Student Populations and Leadership Contexts

A measure's fairness, in part, is dependent on its applicability in all of the leadership and learning contexts for which it is designed. The ability of a measure to be applied to student learning and leadership situations can ensure fidelity in principal evaluation system implementation and capacity to yield valid and useful results.

For example, rural school districts may be challenged to implement all measures of the principal evaluation system. In some rural districts, the school superintendent is also a school principal and, therefore, cannot evaluate himself or herself. Rural districts also may lack the financial and human resources to implement a system with fidelity and adequately maintain system data. Likewise, observations of instructional leadership that focus on principals' hands-on approach to guiding teachers may not be applicable to large high schools, where responsibility for teacher feedback and support is widely distributed among assistant principals or department chairs.

The application of student growth measures to all students and contexts also should be considered. Currently, many states and districts use the average value-added score or growth measures for all teachers in a given school as a factor in principal evaluation. Certain measures of student learning are not appropriate or useful for all students and learning contexts.

For example, certain measures are not appropriate for use with teachers of students with learning disabilities, gifted students, or English learners. Holdheide, Goe, Croft, and Reschly (2010) address the following specific challenges in evaluating teachers of at-risk populations and measuring student growth in these populations:

- Statewide assessment results may be unavailable (e.g., students working toward alternative standards) or not viable.
- Learning trajectories may be different for students with disabilities and English learners.
- The "ceiling effect" for gifted students may prevent adequate measurement of student growth.
- Attribution of student growth when multiple teachers are responsible for instruction and observation of teacher practice with multiple teachers in the classroom can be complicated.

Many states and districts aggregate results to provide a school-level score for principal evaluation, and this process addresses some of the previously noted concerns. States and districts should proceed with caution when selecting measures and seek independent consultants or researchers to provide more information about the application of measures in all contexts.

For example, states and districts should consider how well measures apply to all student and teaching contexts when opting to aggregate test scores or other measures for principal evaluation. Once chosen, states and districts should clearly specify how measures should be used during principal evaluation and support evaluators in the interpretation and use of results.

States and districts also should consider potential consequences of measures selection. Because not all subject areas are tested, for example, principals might believe that only tested subjects count for evaluation purposes and therefore more time and energy should be allocated to improvement of performance in those subject areas.

## Human and Resource Capacity Strengths and Limitations

Each measure has associated costs—both for purchase and for administration—that should be factored into the principal evaluation system design process. Principal evaluation should be thorough, but some measures require more financial and human resources than others. For example, portfolio reviews often require multiple, trained raters to score each portfolio and a method for retaining records over time. Adopting measures without regard to demands placed on teachers, principals, data managers, parents, and superintendents will likely result in poor compliance or fidelity to system requirements, which detracts from fairness, reliability, validity, and utility.

Selection of measures also should consider ongoing evaluator training in assessing human and financial requirements. Many measures, such as observation forms or school walk-throughs, require people to be trained as astute observers of practice. Such measures typically require an initial training to ensure reliability and validity as well as

additional rater supports to maintain or improve accuracy. Some states, such as Iowa, have developed evaluator certification programs, which provide initial and follow-up training to evaluators on principal and teacher evaluation measures.

---

In the process of selecting or contemplating particular measures, stakeholders might consider the guiding questions for Component 3 for each measure.

## Guiding Questions for Component 3

### Selecting Measures

#### GUIDING FACTORS IN MEASURE SELECTION

- 1. Did stakeholders consider all the recommended factors in selecting measures?

Evaluation System's Purpose

#### GUIDING QUESTIONS

- How well does the selected measure align with the evaluation system's purposes and definition of principal effectiveness?
- Can the measure yield data to monitor the evaluation system?
- Does the selected measure assist the state or district to meet pertinent federal, state, or other guidelines for principal evaluation?

NOTES

Strength of Measures

#### GUIDING QUESTIONS

- What is the strength of evidence that the measure is fair, valid, reliable, feasible, and useful for all of the contexts of intended use?
- What processes are in place (or need to be) to ensure the fidelity of the measure?
- How do selected multiple measures complement each other to strengthen the performance evaluation?
- Do the measures overlap so that they are redundant?
- Do the measures contradict each other so that they are misaligned?

Application to All Leadership Contexts

#### GUIDING QUESTIONS

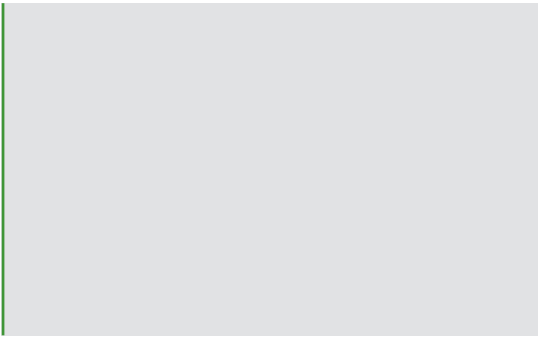
- Is the measure reliable, valid, fair, feasible, and useful for all school leadership contexts?
- How well do student growth measures accurately depict student performance, regardless of context—in particular, in nontested grades and subjects?



**Human and  
Resource  
Capacity**



**GUIDING QUESTIONS**

- What human and resource capacity is necessary to implement the measure reliably and with validity?
  - Are there specific training needs that should be considered?
  - Who will be responsible for maintaining performance data and monitoring system quality?
  - Can resources be pooled between and within districts to implement the measure?
- 

## Guiding Questions for Component 3

### Specific Questions for Measuring Growth in Tested Subjects

#### CONTRIBUTIONS TO STUDENT LEARNING GROWTH

1. Does the state intend to use teachers' contributions to student learning growth (determined using standardized test results) as a factor in principal evaluation (e.g., value-added models and other growth models)?

Plan to Use Other Measures

#### GUIDING QUESTIONS

- Will the other measures be rigorous and comparable across classrooms within a school and across schools?
- How will other measures be used to generate principal evaluation results?
- Is there evidence that the other measures can differentiate among teachers who are helping students learn at high levels and those who are not?
- Will excluding student achievement as a factor be acceptable to the state legislature and the community?
- How will measures be aggregated (e.g., an average of teacher scores) to provide a principal score?

Plan to Use Student Achievement Growth

#### GUIDING QUESTIONS

- Are legislative changes required to implement an evaluation system that includes student growth as a component?
- What types of data will need to be reported?
- Does the state or district currently have human and financial capacity to collect, calculate, and report data with accuracy?
- How will principals be matched to schools, and what decision rules need to be determined to attribute scores to a principal (i.e., for new principals or principals entering a school at mid-year)?
- What types of data will be used in personnel decisions?

#### NOTES

## TESTED SUBJECTS

2. Has a growth model for teachers of tested subjects or principals been selected?

## GUIDING QUESTIONS

- What statistical model of longitudinal student growth will promote the most coherence and alignment with the state's accountability system? Examples: Colorado Growth Model, value-added models
- How will the state or district select potential evaluation models? What technical characteristics does the state or district require?
- Who will be involved in model selection and making decisions about model implementation (e.g., contextual variables to be included, determining exclusion and attribution rules)?
- Who would support or oppose linking teacher and student data? Why? How will these concerns be addressed?
- Will the other measures be rigorous and comparable across classrooms and schools?
- Do these measures meet the federal requirements of rigor: *across two points in time and comparability*?

## PERCENTAGE OF RESULTS BASED ON GROWTH MODEL

3. Has the percentage of principal evaluation results that will be based on the growth model been determined?

## GUIDING QUESTIONS

- Should the percentage differ by the length of a principal's leadership in a school, length of time as a school principal, or other factors (e.g., level of autonomy the principal has in the school, fiscal control)?
- What percentage will be supported by the education community?
- What will the state define as significant?
- Is legislation necessary to determine the percentage?
- Are the assessments reliable and valid to support a significant portion of the evaluation to be based on student progress?

## IDENTIFICATION OF TEACHERS FOR MODEL

4. Have teachers for whom the growth model will be factored into evaluation results been identified?

### GUIDING QUESTIONS

- Will all teachers of tested subjects be included?
- What is the minimum number of students required for a teacher to be evaluated with student growth (e.g., five students per grade or content area)?
- Are there certain student populations in which inclusion in value-added or other growth models may raise validity questions (e.g., students with disabilities, English learners)?
- Can students working toward alternative assessments be included in the growth model?
- How will the state or district choose a model? Will the task force meet with experts? Will the state assessment office investigate options?

## DATA LINKAGE

5. Can student achievement data be accurately linked to schools (data integrity)?

### Data Integrity

### GUIDING QUESTIONS

- What validation process can be established to ensure clean data (e.g., teachers reviewing student lists, administrators monitoring input)?
- Can automatic data validation programs be developed?
- Are there certain student populations in which inclusion in value-added or other growth models is not appropriate (e.g., students with disabilities, English learners)?

### Teaching Context/ Extenuating Circumstances

### GUIDING QUESTIONS

- Have the teacher and principal attribution processes been established for all teaching and leadership situations?
- How will teachers and principals in schools with high student absenteeism rates or highly mobile students be evaluated?
- Has a focus group been held with teachers and principals to determine fair attribution?



## DETERMINATION OF ADEQUATE GROWTH

6. Has a process been established to determine adequate student growth?

## GUIDING QUESTIONS

- How will performance standards be established for principals using student growth, and what will be considered “adequate” or “good”?
- Will a relative or an absolute standard be set (e.g., growth-to-standard or relative growth)?
- Will the standard be based on single-year estimates or estimates combined over time, subjects, or schools (for principals who change schools)?
- How can uncertainty in growth or value-added estimates be taken into account in setting standards or assigning performance levels?
- Who will be involved in setting standards?
- Will the learning trajectory be different for at-risk, special needs, or gifted students?
- Has the “ceiling effect” been addressed?
- Will the use of accommodations affect the measure of student growth?
- Does this measure meet the federal requirements of rigor: *across two points in time and comparability*?

## Guiding Questions for Component 3

### Specific Questions for Alternative Growth Measures in Tested and Nontested Subjects

**MEASURES OTHER THAN STANDARDIZED TESTS**

1. Does the state intend to use measures other than standardized tests to determine student growth (e.g., classroom-based assessments; interim or benchmark assessments; curriculum-based assessments; the Four Ps: projects, portfolios, performances, products)?

**Plan to Use Measures Other Than Standardized Tests but Not Student Achievement Growth**

**GUIDING QUESTIONS**

- Will the other measures be rigorous and comparable across classrooms within a school and across schools?
- How will other measures be used to generate principal evaluation results?
- Is there evidence that the other measures can differentiate among teachers who are helping students learn at high levels and those who are not?
- Will excluding student achievement as a factor be acceptable to the state legislature and the community?

**Plan to Include Student Achievement Growth**

**GUIDING QUESTIONS**

- What would be the challenge of using other measures of growth besides standardized assessment data?
- Will the measures other than standardized tests be rigorous and comparable across classrooms?

NOTES


## IDENTIFICATION OF TEACHERS WHO CONTRIBUTE TO PRINCIPAL EVALUATIONS

2. Have the teachers who meet the criteria for use of measures other than standardized tests been identified?

### GUIDING QUESTIONS

- Will all teachers (in both tested and nontested subjects) be evaluated with alternative growth measures? Only teachers of nontested subjects?
- Which teachers fall under the category of nontested subjects?
- Are there teachers of certain student populations or situations in which standardized test scores are not available or appropriate to utilize?
- Will contributions to student learning growth be measured for related services personnel?

## IDENTIFICATION OF MEASURES

3. Have measures to determine student learning growth been identified?

### Content Standards

### GUIDING QUESTIONS

- Do content standards exist for all grades and subjects?
- Is there a consensus on the key competencies that students should achieve in the content areas?
- Can these content standards be used to guide selection and development of measures?

### Measure Selection

### GUIDING QUESTIONS

- Which stakeholders need to be involved in determining or identifying measures?
- What type of meetings or facilitation will stakeholder groups require to select or develop student measures?
- How will growth in performance subjects (e.g., music, art, physical education) be determined to demonstrate student growth?
- Will the state use classroom-based assessments, interim or benchmark assessments, curriculum-based assessments, and/or the Four Ps (i.e., projects, portfolios, performances, products) as measures?
- Are there existing measures that could be considered (e.g., end-of-course assessments, DIBELS, DRA)?
- Could assessments be developed or purchased?

**RESEARCH**

4. Are there plans to conduct research during implementation to increase confidence in the measures?

**GUIDING QUESTIONS**

- Are federal, state, or private funds available to conduct research?
- How will content validity be tested?
- Can national experts in measurement and assessment be appointed to assist in conducting this research?

## COMPONENT 4

### Determining the Structure of the Evaluation System

The structure of the principal evaluation system contributes to validity of measures and fidelity of implementation. States and districts should clearly communicate the structure of the evaluation system to evaluators, principals, and other stakeholders and create documents that adequately specify the procedure.

State and district principal evaluation designers should create documents that include the following:

- Frequency, order, and timing of the evaluation procedure for all principals
- Any steps of the procedure that fall under the discretion of local evaluators or principals
- The conditions under which evidence collection and evaluation should occur
- The method for scoring and representing principal performance

States and districts report that the most challenging aspect of structuring the principal evaluation system is the determination of evidence levels, weights, and integration.

This section discusses related issues and provides guiding questions for structuring the evaluation system.

### Frequency, Order, and Timing

When designing principal evaluation systems, policymakers should consider the frequency and timing of evaluation to ensure that evaluators, teachers, and principals have the time and attention to critically consider principal performance and complete all aspects of the evaluation. For example, school district testing schedules, professional development days, and other annual schedules will likely impinge on evaluator, principal, and teacher abilities to carefully complete the evaluation forms. Improved evaluation designs will likely require all stakeholders to devote more time to evaluation.

Stakeholder experience with the principal as a school leader also is a concern, which can be addressed by the timing of the evaluation. If policymakers elect to include staff, parent, student, or other surveys in the principal evaluation design, stakeholders must have adequate experience with the principal to allow for an accurate and fair judgment. For example, new staff members need

opportunities to observe and interact with principals in order to make accurate assessments of their performance, just as stakeholders need time to assess new principals' work. Therefore, launching a performance assessment at the beginning of the academic year raises concerns about accuracy, but delaying the performance assessment until, for example, November of each school year provides staff opportunities to form opinions.

When making decisions about the frequency and timing of evaluation, system designers should consider the intended purposes of the evaluation system. National programs (e.g., Race to the Top, TIF, SIG) require grantees to evaluate principals at least twice per year. Such designs might entail one formative and one summative evaluation, but states or districts that set high priorities on formative evaluation may choose to conduct more evaluation cycles so that principals receive frequent feedback on their performance. Similarly, states or districts prioritizing formative evaluation should time the evaluation cycles so that principals have adequate opportunity and access to resources in order to improve their practice.

After all evidence is collected, evaluators need to integrate data into a feedback form. The importance of providing a clear and consistent structure to feedback forms and conversations with evaluators cannot be overemphasized. Principals report that they have few opportunities to receive trusted feedback from colleagues about their practice, and research suggests that feedback is highly valued by organizational leaders and middle managers as a means of developing their work. Without feedback on performance, leaders and managers report that they find it challenging to determine how to improve their work.

## Feedback

Feedback can be powerful, but it also can have a negative effect on personnel if delivered incorrectly. People can lose trust in the evaluation process or the evaluator if feedback is inappropriately structured. The Standards for Personnel Evaluation (Joint Committee on Standards for Educational Evaluation, 2010) indicate that effective feedback forms include the following:

- A clear, concise report of the current assessment by each evaluation area, standard, or domain
- A display of personal growth and/or comparative information (i.e., comparison between the principal and other principals in similar contexts and schools)

- A written narrative that summarizes the evaluation process, findings, feedback, and plans for improvement

Personnel evaluation research indicates that employees find the greatest value in a written narrative and conversation with a trusted, experienced evaluator or supervisor focused on actionable feedback based on data (DeNisi & Kluger, 2000).

## Sources of Evidence

Structuring principal evaluation assessment forms and feedback can be challenging, particularly when evaluation systems involve integration of multiple evidence sources (e.g., surveys, portfolios, observations). In addition to training evaluators (see Component 5) on the provision of effective written and verbal feedback, states and districts may develop the following in order to produce useful feedback forms:

- Clearly defined levels of performance
- Process for establishing weighted standards
- Methods of representing data

## Defined Levels of Performance

In designating the number and description of performance levels, states must ensure that the level designations (e.g., *developing*, *proficient*, *exemplary*) work for principals at different experience levels and determine

whether they should distinguish expected performance for novice principals and more experienced principals. Research suggests that evaluation systems with four or more levels of performance provide workers with more nuanced and actionable feedback for improvement than evaluation systems with two levels (e.g., *present* or *not present*, *yes* or *no*). States and districts should clearly define the distinction between levels of performance by creating rubrics, examples, or other documentation to reduce evaluator and principal misunderstandings of the rating scale.

## Weighted Standards

Principal evaluation systems commonly weight domains or measures to reflect state or district priorities or areas of emphasis for individual principals. Some districts may weight school-level student growth as 40 percent of a principals' total summative score, whereas another district might weight growth at 50 percent of a principals' summative performance evaluation. The weight assigned to measures should reflect the goals and values of the state, district, or principal (depending on the model of evaluation adopted by the state). If, for example, ensuring that principals provide support to teachers in order to improve instruction is a high priority, school climate survey results on that topic may be given a higher weight.

When considering how to weight the various measures collected as part of principal evaluation, it is important to remember that all measures are not equally reliable and useful. States may want to determine a measure's strength in comparison with other measures used within the evaluation system when considering the appropriate weighting of measures.

### Methods of Representing Data

After determining levels and weights for standards, states and districts should design a standard form for displaying evaluation results. The form will be disseminated to principals and may be accompanied by supportive data reports that show how results were determined. The form also may display trend or comparative information.

At least three types of forms are currently being used in the field:

- **Scorecards:** A single form displaying a “score” that may be quantitative or qualitative (e.g., *proficient, distinguished*) for each practice, standard, or outcome.
- **Rubrics:** A set of tables with cells that include descriptors of practices or outcomes for each level. Principals’ scores are highlighted on the rubric.
- **Checklist:** A single form that shows whether or not principals met established performance expectations.

Each form is typically followed by a written narrative and presented to principals during a conference between the principal, evaluator, and others. The three types of forms are often used in combination with one another. For example, a scorecard may include a checklist or rubric.

---

Stakeholders might consider the guiding questions for Component 4 as they determine the structure of the evaluation system.

## Guiding Questions for Component 4

### Determining the Structure of the Evaluation System

**MULTIPLE MEASURES**

1. Will the state promote or use multiple measures?



**GUIDING QUESTIONS**

- What do federal and state legislation, professional association documents, and research say about use of single or multiple measures for principal evaluation?
- If a single measure of principal performance is selected, how strong is the evidence base that the single measure is adequate?
- What combination of measures would more accurately capture the breadth of a principal's roles and responsibilities? Which of these measures might the state wish to mandate for all evaluations?
- Will measures vary depending on school context, grade level, or other factors?

NOTES

Blank area for notes corresponding to the Multiple Measures section.

**STRUCTURE**

2. Has the structure of the evaluation system been determined?



**GUIDING QUESTIONS**

- How often will principals be evaluated formatively, and how often will they be evaluated summatively?
- How, if at all, will the frequency of evaluation be differentiated?
- Will formative evaluations include the entire procedure or part of the evaluation procedure?
- Who will be responsible for administering the evaluation system, and how will these evaluators be trained?
- When will data collection and feedback be provided so that all pertinent data are available for review?

Blank area for notes corresponding to the Structure section.



### WEIGHT OF MEASURES

3. Has the state determined the percentage (weight) of each standard or measure in the overall teacher rating?

### GUIDING QUESTIONS

- Will each measure be weighted differently depending on:
  - Its relation to student achievement?
  - Its relation to supporting principals' improvement of practice?
  - Its relation to state and district improvement priorities?
  - Its reliability and validity?
- Will the weight of each measure fluctuate depending on the level of reliability and validity that is proven over time? What process will be used to improve or capture improvements of a measure's reliability or validity over time?
- Will the weight of measures vary depending on school context, grade level, or principal experience level?

### LEVELS OF PROFICIENCY

4. Have the levels of principal proficiency been determined?

### GUIDING QUESTIONS

- How many levels of proficiency can be explicitly defined?
- Can rubrics be developed to ensure fidelity?
- How often can data be generated?
- What implementation limitations should be considered (e.g., how frequently assessments can be conducted)?
- Will baseline data be analyzed prior to making decisions regarding principal proficiency levels?

### FEEDBACK FORM

5. Has the state or district developed a rubric or feedback form?

### GUIDING QUESTIONS

- What degree of flexibility will the state or district allow for reporting evaluation results to principals?
- Will the state or district use a rubric, scorecard, checklist, or other feedback form?
- Will the state or district require evaluators to write a narrative to accompany the feedback form? If so, what should be included in the narrative?

**CONSEQUENCES OF SCORES**

6. How will the evaluation results be used to inform principals' professional development and learning plans? How will the evaluation results be used to inform state or district professional development offerings to principals?

**Meeting or Exceeding Performance Levels**



**GUIDING QUESTIONS**

- Are opportunities for improvement embedded in the evaluation cycle?
- How, if at all, will evaluation results influence monetary or other incentives for principals?
- Will the state or district provide public recognition or advanced certification for master principals or principals who consistently exceed expectations?
- Are the measures technically defensible for personnel and compensation decisions?

**Failure to Meet Acceptable Performance Levels**



**GUIDING QUESTIONS**

- Are opportunities for improvement embedded in the evaluation cycle?
- Are the measures technically defensible for personnel and compensation decisions?
- Will support be provided to assist principals who demonstrate unacceptable performance?
- How much time and assistance, if any, will be provided for a principal to demonstrate improvement before termination is considered?

## COMPONENT 5

### Selecting and Training Evaluators

Implementation of an improved principal evaluation system will be largely dependent on the quality of training and support provided to evaluators. Evaluators—be they superintendents, assistant superintendents, human resource directors, or others—are at least partially responsible for ensuring that evaluation procedures are followed, data are collected with integrity, information is properly interpreted, and actionable feedback is provided. Each evaluator function requires some initial training and ongoing support. When designing the new evaluation system, states and districts should plan to hire or certify new evaluators; monitor evaluator performance; and provide evaluators feedback to promote improvement in implementation fidelity, inter-rater reliability (as applicable), and increased impact.

Selection or hiring of evaluators is dependent upon the evaluation model that the state or district chooses to pursue. Some districts, for example, apportion a percentage of existing staff time to principal evaluation; others hire part-time staff as evaluators. In many small school districts, the superintendent is a school principal, so another person must appraise his or her performance.

An appropriate amount of time should be allocated to principal evaluators to fully complete evaluations as required by the state or district. Whether selected or hired, principal evaluators should have a strong, working knowledge of principals' work and the context of that work (e.g., elementary school, rural school, turnaround school).

When planning for initial and ongoing evaluator training, states and districts should consider existing human capacity strengths and limitations. For example, large investments of time and money for training may not be possible if state and district budgets are tight, and training methods must be sustainable in the long term after grant or other funding has been depleted. Districts may need additional funding flexibility to allocate human resources for training.

The amount and nature of training is dependent on selected measures. For example, value-added measures of student growth would require training related to the technical aspects of the system and data interpretation. Observations or portfolio review would require a substantial investment in training for evaluators to ensure inter-rater reliability as well as training for principals in using self-reflection forms and portfolio assembly procedures. Surveys, which may

or may not be supported by external vendors, typically require local staff to be trained in survey administration and interpretation. Regardless of the measure, evaluators should be trained on the evaluation procedures and provision of actionable feedback to principals.

Some states, such as Iowa, have developed a statewide evaluator certification process that requires all evaluators to successfully complete initial and ongoing training. To be certified, evaluators must be knowledgeable about evaluation procedures and achieve an acceptable level of inter-rater reliability. If evaluators fail to pass initial training or complete ongoing professional development, they are no longer certified to evaluate principals. Other districts have established peer-assisted review meetings for evaluators to review files and provide feedback to improve evaluation practices. Strong initial training, monitoring of evaluator performance, and ongoing feedback and support will likely improve the evaluation system's fidelity of implementation and integrity.

---

Stakeholders might consider the guiding questions for Component 5 during the evaluator selection and training process.

# Guiding Questions for Component 5

## Selecting and Training Evaluators

**PERSONNEL**

1. What level of training is required to administer and interpret evidence of principal performance?



**GUIDING QUESTIONS**

- What types of training do vendors or designers of measures recommend for the administration and interpretation of data?
- What training do school principals need to ensure that they are knowledgeable about the evaluation system and its requirements?
- How much time does training require, and how will training be funded?

NOTES

**TRAINING AND GUIDELINES**

2. Will the state provide training or guidelines on evaluator/reviewer selection and training?



**GUIDING QUESTIONS**

- What criteria will be used to select evaluators or reviewers?
- Who will be eligible to collect evidence and conduct evaluations?
- How will student outcomes or other extant data be managed?
- Will the state require evaluators or reviewers to have experience as a principal at the school level being evaluated?
- How will the state address personnel time limitation for conducting evaluations or reviews?

NOTES



**GUIDING QUESTIONS**

- How will the state ensure implementation fidelity and system integrity?
- Will the state offer specialized training or certification programs for principal evaluation?
- To what extent will the training provide opportunities for guided practice paired with specific feedback to improve reliability?
- Will the state provide examples and explicit guidance in determining levels of proficiency and approval?
- How will the state or district sustain programs to train new evaluators, as needed?

NOTES

## RETRAINING

3. Does the state have a system in place to retrain evaluators or reviewers if the system is not implemented with fidelity?



## GUIDING QUESTIONS

- Will the state monitor evaluator effectiveness?
- If evaluators or reviewers are not implementing the system with fidelity, what mechanisms will be in place to retrain evaluators/reviewers?
- Will evaluators or reviewers be monitored regularly for checks in reliability?
- How will the state or district provide ongoing evaluator training and feedback to ensure that evaluation practice remains strong?
- How will the state or district sustain training programs?

## COMPONENT 6

### Ensuring Data Integrity and Transparency

Evaluation data can inform decisions about individuals' performance and state or district programming. A data infrastructure can collect, validate, interpret, track, and communicate principal performance data to inform stakeholders, guide professional learning decisions, and assess evaluation system quality. In addition, teacher and student performance data will likely inform principal evaluations. Data integrity and transparency are, therefore, imperative to the evaluation system.

The importance of data integrity and transparency cannot be underestimated, given uses of principal performance assessment data. Carefully administered procedures must be in place to ensure data integrity (Watson, Kramer, & Thorn, 2010). Data integrity requires verification and

cleaning of data as well as establishing clear procedures for data collection. For example, determining teacher and principal value-added scores requires that educators review class lists and work assignments to verify student links to teachers and teacher links to principals. Information technology personnel (who know the data and can create mechanisms for data collection) must design a data infrastructure to reflect principal evaluation measures and system purposes. Principals, teachers, and other school personnel should be well-informed about data integrity assurances and appropriate data integrity procedures to ensure accuracy.

Transparency of measures and resulting data also is a key factor in measure selection. Measures that provide real-time feedback, are accessible and easily understood, and have direct application to teacher practice are more likely to have an immediate impact

on teaching and learning. If teachers and administrators are expected to enter information into data portals, ensuring that these portals are user-friendly will be critical as states scale up evaluation efforts.

Data integrity and transparency improve educator evaluation system functions. Design committee members may wish to engage state and district information technology personnel or vendors in early discussions about technology demands. Committee members also might consider how responsibility for data quality is distributed in the state and district and how evaluation systems hold educators responsible for data quality procedures.

---

Stakeholders might consider the guiding questions for Component 6 to ensure data integrity and transparency.

## Guiding Questions for Component 6

### Ensuring Data Integrity and Transparency

#### DATA INFRASTRUCTURE

1. Is the data infrastructure to collect principal evaluation data established?

#### GUIDING QUESTIONS

- Does the state or district have the data infrastructure to link principals to teachers and teachers to individual student data?
- What is the decision rule for linking a principal to school performance, particularly in cases of mid-year principal transfers or new principals?
- Have the critical questions that stakeholders want the evaluation system to answer been identified? Will the data system collect sufficient information to answer them?
- Have information technology personnel been included in discussions of state and district infrastructure demands?
- Do districts have the technology and human capacity to collect data accurately?

NOTES

#### DATA VALIDATION

2. Is there a data validation process to ensure the integrity of the data?

Validation

#### GUIDING QUESTIONS

- What validation process can be established to ensure clean data (e.g., teachers reviewing student lists, administrators monitoring input)?
- Have criteria been established to ensure teacher and student confidentiality?
- Can computerized programs be used or developed for automatic data validation?

Training

#### GUIDING QUESTIONS

- What training will personnel need to ensure accurate data collection?
- Which personnel at the state and district levels will require training to ensure accuracy in data entry and reporting?

## REPORTING

3. Can principal evaluation data be reported (aggregated or disaggregated) to depict results at the state, district, and building levels?

Teacher  
Data

## GUIDING QUESTIONS

- Do teachers, principals, and principal evaluators have access to pertinent data?
- Is there a system whereby teachers or administrators can make changes when errors are found?
- Is the data collection methodology or database easily understood and user-friendly?
- Have principals been trained to extrapolate and use the data to inform teacher practice?
- Are administrators, teachers, and parents (as appropriate) trained in how to use the database and interpret teacher evaluation results?

Student  
Data

## USE OF DATA

4. Is there a plan for how the principal evaluation data will be used?

Data  
Sharing

## GUIDING QUESTIONS

- What level of data is appropriate to share with the principal, without jeopardizing evaluation system integrity or survey respondent confidentiality?
- How frequently, if at all, should principal evaluation data be shared with the education community?
- What principal evaluation data would be relevant, easily understood, and appropriate to share with the education community?
- Who will have access to principal evaluation data?
- How will evaluation results be shared with the community (e.g., website, press releases, town meetings)?

Data  
Use

## GUIDING QUESTIONS

- Will principal evaluation data be used to inform changes in the principal evaluation design?
- Will data be used to identify principals in need of support and target professional learning?
- Will data be used to identify highly effective principals and potential principal mentors?
- Will data be used to identify principals for advanced or master certification?
- Will data be used by states and districts to inform selection of professional development providers or programs?



## COMPONENT 7

### Using Principal Evaluation Results

Data collected from the principal evaluation system hold potential for providing principals feedback, support learning, inform personnel decisions, and facilitate preservice and inservice program planning. States and districts should determine, in advance, how evaluation data will and will not be used because this decision informs data infrastructure and reporting decisions. States and districts should clearly communicate intended uses of data to principals.

States and districts also should consider “decision rules,” or points at which human resource actions should be taken. This section describes issues and raises questions to assist states and districts in creating decision rules about the use of evaluation data.

System designers should critically consider who will have access to principal evaluation data and for what purpose. Some states and districts, for example, may be inclined to publicly release performance assessment results, but doing so may lead to unintended consequences. The National Association of Elementary School Principals strongly opposes the release of principal evaluation

results because making results public could undercut the trust and confidentiality necessary to gather strong data on leadership.

### Decision Rules for Retention, Advancement, and Compensation

If states and districts use evaluation data for retention, progressive discipline, advancement, or compensation decisions, system designers must clearly determine and communicate the assessment results. States and districts will need to determine “cut scores,” which are quantitative or qualitative evidence that performance should trigger a personnel action. Further, states and districts should consider whether all results are weighted equally for personnel decisions and whether single or multiple scores are necessary to prompt action.

### Making Professional Learning Decisions

The use of evaluation results to inform professional development decisions is a valuable function of the evaluation system. So long as data have integrity, evaluation results can be used to identify individual, districtwide, or statewide learning needs and can inform decisions about professional

#### RESOURCE

***Job-Embedded Professional Development: What It Is, Who Is Responsible, and How to Get It Done Well***

<http://www.gtcenter.org/sites/default/files/docs/JEPD%20Issue%20Brief.pdf>

This issue brief provides specific recommendations for states to support high-quality job-embedded professional development (p. 10):

- ▶ “Help build a shared vocabulary.”
- ▶ “Provide technical assistance.”
- ▶ “Monitor implementation.”
- ▶ “Identify successful job-embedded professional development practices within the state.”
- ▶ “Align teacher licensure and relicensure requirements with high-quality job-embedded professional development.”
- ▶ “Build comprehensive data systems to inform decisions.”

development programming. Performance feedback can, for example, result in annual professional development planning decisions for individual principals or could be used at regional or state levels to inform mentoring programs, conference planning, or other professional development programming.

If states and districts intend to use evaluation system data to inform professional development decisions, the following questions might be considered:

- How closely must principals' professional development plans align with evaluation system results?
- Who should have access to individual, district, and state-level data on principal performance?
- How can data be reported to afford better professional development planning decisions?

Just as some states (e.g., Colorado) and districts (e.g., Hillsborough County Public Schools in Florida) hold principals accountable for using evaluation data to inform teacher professional development and retention decisions, design committees may consider how district central office staff are accountable for ensuring that principal evaluation data are used to inform decisions about principal workforce distribution, retention, professional learning, and other human resource functions.

Evaluation system data also may be helpful in evaluating certification and professional development program quality because evaluation data can be used to chart performance needs, professional development participation, growth in practice, and achievement of outcomes. As the evaluation system database matures, these types of reports can be generated.

---

Stakeholders might consider the guiding questions for Component 7 as they contemplate professional development needs.

## Guiding Questions for Component 7

### Using Principal Evaluation Results

#### DECISION RULES

1. Have decision rules for personnel actions using evaluation results been established?

#### GUIDING QUESTIONS

- Does the state intend to align evaluation results to human resource decisions?
- At what point will evaluation results warrant promotion, dismissal, progressive discipline, or other decisions?
- How many evaluation cycles will be used to identify exemplary principals or principals who are in need of improvement?
- To what degree are processes in place to strengthen performance and track growth?
- How will evaluation results be shared with principals?
- How will principals be notified of personnel decisions affecting their career continuation or advancement?

NOTES

#### EVALUATION RESULTS

2. Will principal evaluation results be used to target professional development activities?

#### GUIDING QUESTIONS

- How will performance evaluation data be used to inform professional development choices?
- How effective is principal professional development planning and monitoring?
- To what degree must professional development plans align with evaluation results?
- Will principals identified as ineffective have sufficient opportunities and support to improve before termination is considered?
- Will personnel decisions be defensible if principals were not provided an opportunity and the resources to improve?
- What resources, including time and personnel, are dedicated to teacher improvement?
- How will evaluation systems data inform principal professional development offerings?
- Can evaluation results be used to identify principals for advanced certification or mentoring positions?
- Will the state or district work in collaboration with principal preparation programs to ensure that candidates are prepared with the competencies for which they will be held accountable as they begin leading schools?

## EVALUATION OF PROFESSIONAL DEVELOPMENT SYSTEMS

3. Are systems established to evaluate professional learning efforts?

### Evaluating the Training

#### GUIDING QUESTIONS

- What mechanism will be established to ensure that participant feedback is obtained (e.g., training evaluation, follow-up survey)?
- What procedures will be established to ensure that active participation and application are integral parts of the professional development activity?

### Reviewing the Outcomes

#### GUIDING QUESTIONS

- Can the evaluation measure(s) detect principal growth as a result of professional development efforts?
- Can demonstrated principal growth be correlated to improved student achievement?
- What mechanism will be established to follow up with principals to ascertain whether practice has been improved as a result of the professional learning efforts (e.g., follow-up survey or observation)?

### Modifying the Process

#### GUIDING QUESTIONS

- Can the system identify which professional learning opportunities are or are not effective?
- Are changes in the evaluation system necessary to associate principal growth and other outcomes with participation in professional learning activities?
- How will results (e.g., evaluations and outcomes) be used to improve professional development offerings and strategies?

## COMPONENT 8

### Evaluating the System

Research can play an important role in the long-term improvement of principal evaluation systems. Few research and evaluation studies are currently available that test the design and impact of school principal evaluation on principals' practice, school conditions, or student learning (Clifford & Ross, 2011; Davis et al., 2011). The paucity of research on principal evaluation design and the need to "get it right" raises the importance of pilot testing or field testing the principal evaluation system, evaluating system impact, and routinely reassessing and improving system performance.

Systematically evaluating the performance of the evaluation model in terms of its goals and results and modifying its structure, processes, or format accordingly ensures system efficacy and sustainability. State or federal policy and programs may require states to determine the quality of evaluation system implementation and the impact of system implementation on leaders, schools, and students. Such research can ensure that the evaluation system is technically sound, and therefore legally defensible, especially when evaluation results are intended to influence compensation and personnel decisions.

An independent research study also can be effective in gaining stakeholder support for the new evaluation system. Studies can identify the factors that help or hinder system performance. For example, the state and districts will want to know whether:

- Stakeholders value and understand the system.
- Student performance has improved.
- Principal practice has been affected.
- Principal retention or mobility has improved.
- School conditions and instructional quality have improved.
- The system has been implemented with fidelity and integrity.

States have used external and internal review processes to collect and analyze data or a combination of both. Surveys of teachers, administrators, and stakeholders may be valuable for this process.

Ultimately, researchers should work closely with stakeholders to ensure that the design addresses important questions. A state or district may wish to study the following:

- Principal and supervisor satisfaction with the evaluation process

- Fidelity of implementation to core elements of the evaluation system
- Inter-rater reliability on evaluation measures
- Validity studies on evaluation measures
- Impact of evaluation system implementation

Ideally, research studies will involve a comparative component, which allows researchers to examine differences between implementation and nonimplementation sites.

---

Stakeholders might consider the guiding questions for Component 8 to determine the overall effectiveness of the principal evaluation system.

## Guiding Questions for Component 8

### Evaluating the System

#### EVALUATION PROCESS

1. Has a process been developed to systematically evaluate the effectiveness of the principal evaluation model?

#### GUIDING QUESTIONS

- Has the model been piloted, or are there plans to pilot the model prior to statewide or districtwide implementation?
- Is there a plan for securing stakeholder and participant feedback?
- Will research be conducted in conjunction with implementation to provide validation?
- Will research be conducted to determine whether there is correlation between growth model scores and observation ratings?
- How will the state or district ensure that evaluation studies are conducted with integrity?
- Are resources available to conduct an internal or external assessment of the evaluation model?

NOTES

#### EFFECTIVENESS OUTCOMES

2. Have outcomes to determine the overall effectiveness of the principal evaluation system been established?

#### GUIDING QUESTIONS

- Have the stakeholders identified factors that should be considered in determining whether the evaluation system is effective (e.g., participant satisfaction, improved teacher practice, other improved student outcomes)?
- Have explicit benchmarks or targets been established to determine the effectiveness of system implementation?
- How will effectiveness be measured?
- Has the data infrastructure been established to track data over a period of time to determine teacher and student growth?
- In review of baseline data, what would be acceptable performance targets?
- How will fidelity of implementation be measured?
- Will data be collected on principal effectiveness to determine whether effective principals are and remain equally distributed throughout the state in high-performing and low-performing schools?

## Conclusion and Recommendations

Principals are uniquely positioned to influence teacher quality, school performance, and student learning. For this reason, principal evaluation systems hold great promise for providing feedback and self-reflection, which can facilitate leader engagement in professional learning and improved practice. Rigorous and systematic principal evaluation systems also hold promise for modeling the type of evaluation that principals should conduct with teachers.

Cultivating effective principal evaluation systems is challenging, particularly with the dearth of research-based models and measures currently available. In many states, principal evaluation is not widely or systematically practiced, aligned with state or national professional standards, or linked to state or district data infrastructures. State and district design teams, therefore, have the opportunity to develop innovative assessment systems that sponsor better leadership through learning.

Improved principal evaluation systems require states and districts to make a myriad of decisions, from selecting or creating feedback forms to generating new data infrastructures. Most important, though, states and districts can generate trust among stakeholders, which will support collaborative design and instill support for a system that encourages leaders to think deeply with colleagues about improving the achievement of schools and promoting student learning. The new evaluation system not only should hold principals accountable for performance, it also should support principals' continued growth; help educators at all levels of the school system identify strong leadership practices and professional learning opportunities; and encourage leadership that is supportive of students, communities, and schools.

## References

- American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, 123 Stat. 115 (2009). Retrieved from <http://www.gpo.gov/fdsys/pkg/BILLS-111hr1enr/pdf/BILLS-111hr1enr.pdf>
- Anthes, K. (2005). *Leader standards*. Denver, CO: Education Commission of the States. Retrieved from <http://www.ecs.org/clearinghouse/58/19/5819.doc>
- Berman, P., & McLaughlin, M. W. (1976). Implementation of educational innovation. *The Educational Forum*, 40, 345–370.
- Clifford, M., & Ross, S. (2011). *Designing principal evaluation systems: Research to guide decision-making*. Washington, DC: National Association for Elementary School Principals. Retrieved from [https://www.naesp.org/sites/default/files/PrincipalEvaluation\\_ExecutiveSummary.pdf](https://www.naesp.org/sites/default/files/PrincipalEvaluation_ExecutiveSummary.pdf)
- Colorado Department of Education. (2011). *User's guide: Colorado Model Evaluation System for Principals and Assistant Principals*. Denver, CO: Author.
- Condon, C., & Clifford, M. (2010). *Measuring principal performance: How rigorous are commonly used principal performance assessment instruments?* Naperville, IL: American Institutes for Research. Retrieved from [http://www.air.org/sites/default/files/downloads/report/Measuring\\_Principal\\_Performance\\_0.pdf](http://www.air.org/sites/default/files/downloads/report/Measuring_Principal_Performance_0.pdf)
- Council of Chief State School Officers. (2008). *Educational leadership policy standards: ISLLC 2008*. Washington, DC: Author. Retrieved from [http://www.ccsso.org/Documents/2008/Educational\\_Leadership\\_Policy\\_Standards\\_2008.pdf](http://www.ccsso.org/Documents/2008/Educational_Leadership_Policy_Standards_2008.pdf)
- Davis, S., Kearney, K., Sanders, N., Thomas, C., & Leon, R. (2011). *The policies and practices of principal evaluation: A review of the literature*. San Francisco, CA: WestEd. Retrieved from [http://www.wested.org/online\\_pubs/resource1104.pdf](http://www.wested.org/online_pubs/resource1104.pdf)
- DeNisi, A. S., & Kluger, A. N. (2000). Feedback effectiveness: Can 360-degree appraisals be improved? *Academy of Management Executive*, 14(1), 129–139.
- Elementary and Secondary Education Act (No Child Left Behind Act of 2001), Pub. L. No. 107-110, 115 Stat. 1425 (2002). Retrieved from <http://www.ed.gov/policy/elsec/leg/esea02/index.html>
- Friedman, I. (2002). Burnout in school principals: Role related antecedents. *Social Psychology of Education*, 5(3), 229–251.
- Goe, L., Holdheide, L., & Miller, T. (2014). *Practical guide to designing comprehensive teacher evaluation systems*. Washington, DC: Center on Great Teachers and Leaders. Retrieved from <http://www.gtlcenter.org/sites/default/files/docs/practicalGuideEvalSystems.pdf>
- Goldring, E., Cravens, X., Murphy, J., Porter, A., Elliott, S., & Carson, B. (2009). The evaluation of principals: What and how do states and urban districts assess leadership? *Elementary School Journal*, 110(1), 19–39.



- Gullickson, A. R. (2009). *Personnel evaluation standards: How to assess systems for evaluating educators* (2nd ed). San Francisco, CA.
- Hale, E., & Moorman, H. (2003). *Preparing school principals: A national perspective on policy and program innovations*. Washington, DC: Institute for Educational Leadership. Retrieved from <http://www.iel.org/pubs/preparingprincipals.pdf>
- Hallinger, P., & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980–1995. *School Effectiveness and School Improvement*, 9, 157–191.
- Halverson, R., & Clifford, M. (in press). Distributed leadership in high school. *Journal of School Leadership*.
- Heck, R. H., & Marcoulides, G. A. (1996). Principal assessment: Conceptual problem, methodological problem, or both? *Peabody Journal of Education*, 68(1), 124–144.
- Herman, R., Dawson, P., Dee, T., Greene, J., Maynard, R., Redding, S., et al. (2008). *Turning around chronically low-performing schools: A practice guide* (NCEE #2008-4020). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved from [http://ies.ed.gov/ncee/wwc/pdf/practice\\_guides/Turnaround\\_pg\\_04181.pdf](http://ies.ed.gov/ncee/wwc/pdf/practice_guides/Turnaround_pg_04181.pdf)
- Herman, J. L., Hertiage, M., & Goldschmidt, P. (2011). *Developing and selecting assessments for student growth for use in teacher evaluation systems*. Los Angeles, CA: Assessment and Accountability Comprehensive Center. Retrieved from [http://www.cse.ucla.edu/products/policy/shortTermGrowthMeasures\\_v6.pdf](http://www.cse.ucla.edu/products/policy/shortTermGrowthMeasures_v6.pdf)
- Holdheide, L., Goe, L., Croft, A., & Reschly, D. (2010). *Challenges in evaluating special education teachers and English language learner specialists* (Research & Policy Brief). Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from <http://www.gtlcenter.org/sites/default/files/docs/July2010Brief.pdf>
- Illinois State Board of Education. (2012, January). *Education reform in Illinois: Non-regulatory guidance on the Performance Evaluation Reform Act and Senate Bill 7*. Retrieved from [http://www.isbe.net/PERA/pdf/pera\\_guidance.pdf](http://www.isbe.net/PERA/pdf/pera_guidance.pdf)
- Ingersoll, R., & Smith, T. (2003). The wrong solution to the teacher shortage. *Educational Leadership*, 60(8), 30–33.
- Joint Committee on Standards for Educational Evaluation. (2010). *Personnel evaluation standards: Summary of the standards* [Website]. Retrieved from <http://www.jcsee.org/personnel-evaluation-standards>
- Kimball, S. (2011). Strategic talent management for principals. In A. Odden (Ed.), *Strategic management of human capital in public education: Improving instructional practice and student learning in schools* (133–152). New York, NY: Routledge.
- Ladd, H. (2009). *Teachers' perceptions of their working conditions: How predictive of policy-relevant outcomes?* (CALDER Working Paper 33). Washington, DC: Urban Institute. Retrieved from <http://www.urban.org/uploadedpdf/1001440-Teachers-Perceptions.pdf>

- Lambert, L., Walker, D., Zimmerman, D. P., Cooper, J. E., Lambert, M. D., Gardner, M. E., et al. (2004). *The constructivist leader* (2nd ed.), New York, NY: Teachers College Press.
- Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning*. New York, NY: The Wallace Foundation. Retrieved from <http://www.wallacefoundation.org/knowledge-center/school-leadership/key-research/Documents/How-Leadership-Influences-Student-Learning.pdf>
- Marzano, R. J., Waters, T., & McNulty, B. A. (2005). *School leadership that works: From research to results*. Alexandria, VA: ASCD.
- Milanowski, A., & Kimball, S. (2010). The principal as human capital manager: Lessons from the private sector. In R. Curtis & J. Wurtzel (Eds.), *Teaching talent: A visionary framework for human capital in public education*. Cambridge, MA: Harvard Education Press.
- Milanowski, A. T., Longwell-Grice, H., Saffold, F., Jones, J., Schomisch, K., & Odden, A. (2009). Recruiting new teachers to urban school districts: What incentives will work? *International Journal of Educational Policy and Leadership*, 4(8).
- Murphy, J., & Datnow, A. (2003). *Leadership lessons from comprehensive school reform*. San Francisco: Corwin Press.
- National Council for Accreditation of Teacher Education. (2014). *Glossary* [Website]. Retrieved from <http://www.ncate.org/Standards/UnitStandards/Glossary/tabid/477/Default.aspx>
- Orr, M. (2011, September 22–23). *Evaluating leadership preparation program outcomes*: Presented at U.S. Department of Education's School Leadership Program Working Conference "Learning and Leading: Preparing and Supporting School Leaders," Virginia Beach, VA.
- Portin, B. S., Feldman, S., & Knapp, M. S. (2006). *Purposes, uses, and practices of leadership assessment in education*. New York: The Wallace Foundation. Retrieved from <http://depts.washington.edu/ctpmail/PDFs/LAssess-Oct25.pdf>
- Public Agenda. (2009). *Retaining teacher talent survey of teachers: Full survey data*. New York, NY: Author. Retrieved from <http://www.learningpt.org/expertise/educatorquality/genY/FullSurveyData.pdf>
- Secretary's Priorities for Discretionary Grant Programs, 75 Fed. Reg. 47,288 (proposed Aug. 5, 2010). Retrieved from <http://www2.ed.gov/legislation/FedRegister/other/2010-3/080510d.pdf>
- Spillane, J. P., & Diamond, J. B. (2007). *Distributed leadership in practice*. New York, NY: Teachers College Press.
- Spillane, J., Halverson, R., & Diamond, J. (2004). Towards a theory of school leadership practice: Implications of a distributed perspective. *Journal of Curriculum Studies*, 36(1), 3–34. Retrieved February 17, 2012, from <http://ddis.wceruw.org/docs/SpillaneHalversonDiamond2004JCS.pdf>
- Stronge, J., Richard, H., & Catano, N. (2008). *Qualities of effective principals*. Alexandria, VA: Association for Supervision and Curriculum Development.

- Supovitz, J., & Poglinco, S. (2001). *Instructional leadership in a standards-based reform*. Philadelphia, PA: Consortium for Policy Research in Education. Retrieved from [http://www.cpre.org/images/stories/cpre\\_pdfs/AC-02.pdf](http://www.cpre.org/images/stories/cpre_pdfs/AC-02.pdf)
- Teacher Leadership Exploratory Consortium. (2011). *Teacher leader model standards*. Washington, DC: Author. Retrieved from [http://www.teacherleaderstandards.org/standards\\_overview](http://www.teacherleaderstandards.org/standards_overview)
- Tennessee Department of Education. (2011). *Teacher and principal evaluation policy: Final reading item: IV. C*. Nashville, TN: Author. Retrieved from <http://www.tn.gov/sbe/2011Aprilpdfs/IV%20C%20Teacher%20and%20Principal%20Evaluation%20Policy.pdf>
- Thomas, D., Holdaway, E., & Ward, K. (2000). Policies and practices involved in the evaluation of school principals. *Journal of Personnel Evaluation in Education*, 14(3), 215–240.
- U.S. Department of Education. (2010). *Race to the Top application for initial funding* (CFDA Number 84.395A). Washington, DC: Author.
- U.S. Department of Education. (2011). *Elementary and Secondary Education Act (ESEA) flexibility*. Washington, DC: Author. Retrieved from <http://www.ed.gov/esea/flexibility/documents/esea-flexibility.doc>
- Wahlstrom, K. L., Louis, K. S., Leithwood, K., & Anderson, S. E. (2010). *Investigating the links to improved student learning: Executive summary of research findings*. New York, NY: The Wallace Foundation. Retrieved from <http://www.wallacefoundation.org/knowledge-center/school-leadership/key-research/Documents/Investigating-the-Links-to-Improved-Student-Learning-Executive-Summary.pdf>
- The Wallace Foundation. (2011). *Research findings to support effective educational policies: A guide for policymakers* (2nd ed.). New York, NY: Author. Retrieved from <http://www.wallacefoundation.org/knowledge-center/school-leadership/key-research/Documents/Findings-to-Support-Effective-Educational-Policy-Making.pdf>
- The Wallace Foundation. (2012). *The school principal as leader: Guiding schools to better teaching and understanding*. New York, NY: Author. Retrieved from <http://www.wallacefoundation.org/knowledge-center/school-leadership/effective-principal-leadership/Documents/The-School-Principal-as-Leader-Guiding-Schools-to-Better-Teaching-and-Learning.pdf>
- Waters, T., Marzano, R., & McNulty, B. (2003). *Balanced leadership: What 30 years of research tells us about the effect of leadership on student achievement*. Denver, CO: McREL. Retrieved from [http://www.mcrel.org/~media/Files/McREL/Homepage/Products/01\\_99/prod82\\_BalancedLeadership.ashx](http://www.mcrel.org/~media/Files/McREL/Homepage/Products/01_99/prod82_BalancedLeadership.ashx)
- Watson, J., Kramer, S., & Thorn, C. (2010). *Data quality essentials: Guide to implementation*. Washington, DC: Center for Educator Compensation Reform. Retrieved from <http://cecr.ed.gov/pdfs/guide/dataQuality.pdf>

## Appendix A. Glossary of Terms

This glossary contains terminology that often is associated with the development of educator evaluation systems. As states move toward comprehensive evaluation of principals, expectations and intersections of responsibility are of critical importance.

The glossary is divided into three sections. The first section pertains to principal evaluation and contains a listing of general terminology and definitions for various ways of measuring performance. The second section addresses common terminology and definitions for performance measures for both teacher and principal evaluations. The third section defines technical aspects of both teacher and principal performance evaluation. Sources are cited in instances in which the definition has a primary source.

### Section 1: Principal Evaluation

#### General Terminology

**effective principal:** “Principal whose students, overall and for each subgroup, achieve acceptable rates (e.g., at least one grade level in an academic year) of student growth.” States, local education agencies, or schools “must include multiple measures, provided that principal effectiveness is evaluated, in significant part, by student growth.... Supplemental measures may include, for example, high school graduation rates and college enrollment rates, as well as evidence of providing supportive teaching and learning conditions, strong instructional leadership, and positive family and community engagement.” (U.S. Department of Education, 2010, p. 7)

**highly effective principal:** “Principal whose students, overall and for each subgroup, achieve high rates (e.g., one and one-half grade levels in an academic year) of student growth.” States, local education agencies, or schools “must include multiple measures, provided that principal effectiveness is evaluated, in significant part, by student growth.... Supplemental measures may include, for example, high school graduation rates; college enrollment rates; evidence of providing supportive teaching and learning conditions, strong instructional leadership, and positive family and community engagement; or evidence of attracting, developing, and retaining high numbers of effective teachers.” (U.S. Department of Education, 2010, p. 8)

#### Principal Performance Measures

**principal observations:** Used by the superintendent, or his or her designee, to measure observable principal behaviors, actions, or practices within a principal practice framework. Evaluators use these observations to make consistent judgments of principals’ practice. High-quality observation instruments are based on standards and contain well-specified rubrics that delineate consistent assessment criteria for each standard of practice.

**leadership artifacts:** Artifacts used to analyze principal behaviors, actions, and practices. Often, they relate to the “technical core” of schooling—what is required to improve the quality of teaching and learning. They include, for example, a vision statement, a schoolwide learning improvement plan, climate survey results, principal analyses of teachers’ growth and development in relation to a schoolwide improvement plan, tracking of teacher professional development needs, classroom instruction observations, evidence of the principal hiring carefully, and evidence that the principal views “data as a means not only to pinpoint problems but to understand their nature and causes.” (The Wallace Foundation, 2012, p. 12)

**multiple measures of principal performance:** The various measures of principal effectiveness that include multiple measures of student learning and measures of traditional practices. They include, for example, high school graduation rates and college enrollment rates. They also may include a measure of progress on an individual, school, or district performance goal; feedback from teachers or other stakeholder groups; an assessment of the quality of the principal's evaluation of teachers; evidence of the principal's leadership for implementing a rigorous curriculum; and evidence of the principal's leadership for high-quality instruction. Although multiple measures of principal performance are recommended, this evidence "will likely need to be weighted and represented in ways that reflect leadership standards and priorities." (Clifford & Ross, 2011, p. 6)

**student growth:** According to U.S. Department of Education regulations, a principal's students must demonstrate high rates of student growth overall and for each subgroup. Effectiveness is determined (in significant part) using aggregate rates of student growth. However, there is no federal requirement that each student in the principal's school must demonstrate a high rate of student growth individually. (U.S. Department of Education, 2010)

**working conditions** (also teaching conditions, school conditions): Sometimes used as a measure of principal performance, *working conditions* refers to the conditions in which learning occurs and may include amenities, physical environment, stress and noise levels, and degree of safety or danger.

## Section 2: Educator Evaluation

### General Terminology

**educator growth and development system:** A comprehensive performance management system that incorporates multiple measures of both educator evaluation and student learning and has the intent of improving the knowledge, skills, dispositions—that is, positive behaviors characterized by "professional attitudes, values, and beliefs demonstrated through both verbal and non-verbal behaviors as educators interact with students, families, colleagues, and communities"—as well as the practices of professional educators. Beyond a simple evaluation system, an educator growth and development system is connected closely to other key aspects of the educator continuum (e.g., induction, professional development). (National Council for Accreditation of Teacher Education, 2014)

**simple growth models:** Traditional definitions of growth models indicate that they are statistical models that measure student achievement growth from one year to the next by tracking the same students. This type of model addresses the following question: How much, on average, did students' performance change from one grade to the next? The question can be answered using simple or more complex methods.

**nontested grades and subjects:** The grades and subjects that are not required to be tested under the Elementary and Secondary Education Act (or by state statutes and regulations).

**performance management system:** The entire system that affects a teacher's or principal's career continuum. Although evaluation is a large component of the system, *performance management* refers to the utilization of evaluation data to inform decisions including hiring, tenure, compensation, and dismissal of teachers as well as hiring, compensation (e.g., performance pay), financial incentives or rewards, job selections, school placements, and dismissal of principals.

**portfolios and evidence binders:** A collection of materials that exhibit evidence of educator practice, school activities, and student progress. Portfolios are usually compiled by the teacher or the principal and may include the teachers' instructional artifacts or principals' leadership artifacts, videos of classroom instruction, notes from parents and others, and the educators' analyses of their students' learning in relation to their school improvement plan. Evidence binders often have specific requirements for inclusion and may involve a final educator-led presentation of the work to an evaluation team.

**360-degree evaluation:** A method of gathering information about employee performance from the employee's supervisors, colleagues, supervisees, students, other constituents, and/or the employee himself or herself.

**unique identifier:** Numbers that are assigned to each individual student, teacher, and principal in a school and are matched to data about that student's, teacher's, or principal's performance.

**value-added models (VAMs):** Complex statistical models that attempt to determine the extent to which specific teachers and schools affect student achievement growth over time. These models use at least two years of students' test scores and may take into account other student- and school-level variables, such as family background, poverty, and other contextual factors.

## Educator Performance Measures

**evaluation tools:** Models, rubrics, instruments, and protocols that are used by evaluators to assess educators' performances.

**formative educator evaluation:** Used primarily to provide feedback to improve performance and future actions. Along with summative educator evaluation, it is an integral part of educator staff development and critical in providing "useful, valuable, and trustworthy data and feedback for advancing educators' abilities to be more effective teachers and principals" within their schools and communities. (Clifford & Ross, 2011, p. 4)

**goal-driven professional development plans:** Evaluation instruments that offer educators the opportunity to set their own ambitious but feasible objectives for their professional growth in collaboration with their evaluator or other colleagues. Some instruments require educators to specify the professional development in which they will participate to ensure that their students achieve their growth objectives.

**growth measures:** Assessments of students' improvements in learning from one point in time to another point in time. Growth measures refer to the scores that are developed from a growth model or with regard to academic goals (e.g., student learning objectives).

**growth to proficiency models:** Models that measure whether students are on track to meet standards for proficient and above.

**measures:** Types of instruments or tools used to assess the performance and outcomes of educator practice (e.g., student growth scores, observations, student surveys, analysis of classroom artifacts, student learning objectives).

**measures of collective performance:** The use of measures required by the current provisions of the Elementary and Secondary Education Act and/or other standardized assessments designed to measure the performance of groups of teachers. Measures of collective performance may assess the performance of the school, grade level, instructional department, teams, or other groups of teachers. These measures can take a variety of forms including schoolwide student growth measures, team-based collaborative achievement projects, and shared value-added scores for coteaching situations.

**multiple measures of educator performance:** The various types of assessments of educators' performance—including, for example, classroom observations, student test score data, self-assessments, or student or parent surveys.

**multiple measures of student learning:** The various types of assessments of student learning—including, for example, value-added or growth measures, curriculum-based tests, pretests and posttests, capstone projects, oral presentations, performances, or artistic or other projects.

**performance continuum:** Indicator of progressing levels of performance; generally set on a scale within a measure, such as a rubric.

**practice standards:** The broadest category of performance that describes the behavior and characteristics of an effective educator.

**rubric:** A method for defining and categorizing performance by highlighting important aspects of performance and defining observable and measurable levels of performance along a performance continuum. In personnel performance assessment, rubrics can be used to communicate performance expectations, support self-reflection on practice, and facilitate self-reflection between evaluator and educator.

**school climate surveys:** Questionnaires that ask parents, teachers, and others to rate the principal or the school on an extent scale regarding various aspects of school leadership as well as the extent to which they are satisfied with conditions for student and adult learning.

**summative educator evaluation:** This type of evaluation of educators' practice integrates multiple sources of data for the purpose of making high-stakes personnel decisions. Along with formative educator evaluation, it is an integral part of educator staff development and critical in providing "useful, valuable, and trustworthy data and feedback for advancing educators' abilities to be more effective teachers and principals" within their schools and communities (Clifford & Ross, 2011, p. 9)

**teacher and principal self-assessments:** Surveys, instructional logs, or interviews in which teachers or principals report on their work in the school, the extent to which they are meeting standards, and in some cases, the impact of their practice. Self-assessments may consist of checklists, rating scales, and rubrics; they may require teachers and principals to indicate the frequency of particular practices.

## Section 3: Technical Terms

**fair:** A term used to describe evaluation measures and methods that are impartial in content and consistently administered to educators by trained staff so that they are held to similar standards.

**feasible:** Whether an evaluation measure or method can be developed, implemented, or is reasonable.

**fidelity:** Accuracy and exactness of facts or details on performance measures. Fidelity of implementation requires that evaluators are trained, monitored, and supported.

**inter-rater reliability:** A construct in measurement describing the degree to which different assessors rate the same observed behaviors or other phenomenon the same way.

**reliability:** A measure of the degree to which an instrument measures something consistently. A validated instrument must be evaluated for how reliable the results are across raters and contexts. Discussion of methods for measuring teaching effectiveness often makes reference to rater reliability—whether or not raters have been trained to score reliably. Scoring reliably means being able to do the following: rate consistently with standards, rate consistently with other raters (referred to as inter-rater reliability), and rate consistently across observations and contexts. Ratings should not be influenced by factors such as the time of day, time of year, or subject matter being taught; they should be consistent across observations of the same educator.

**teacher effect:** A teacher's contribution to student performance growth compared with that of the average (or median, or otherwise defined) teacher in the district or the state.

**validity:** The ability of an instrument to measure the attribute that it intends to measure.



# Appendix B. Some Principal Evaluation Measures

Measure	Description	Research	Strengths	Cautions
Classroom Observation	Used to measure observable behaviors or practices of school principals including such aspects as communication; ability to distribute leadership, instructional leadership and management; ability to read and convey performance data; and ability to provide feedback to teachers. Can measure broad, overarching aspects of the day-to-day or context-specific aspects of various school leadership responsibilities that fall under the purview of a school administrator.	There is a lack of research on valid and reliable principal observation protocols.	<ul style="list-style-type: none"> <li>■ Provides rich information about principal behaviors and practices.</li> <li>■ Can be used to evaluate a principal in various contexts.</li> <li>■ Can provide useful information for formative and summative purposes.</li> </ul>	<ul style="list-style-type: none"> <li>■ Careful attention must be paid to choosing or creating a valid and reliable protocol and training and calibrating raters.</li> <li>■ Valid principal observations are scarce. There are not many existing observation protocols that are designed to evaluate or observe principal practice as opposed to teacher practice (e.g., classroom observations).</li> <li>■ Principal observations should go beyond relying on yes-or-no checklists, be used in conjunction with other forms of data (e.g., principal portfolios, 360-degree evaluations), and take into account the principal's position or level of experience as well as the school context in which he or she is working in order to gain a full picture of principal practice.</li> <li>■ Observation protocols should assess the specific behaviors and actions of a principal rather than just personality traits, be tied to a validated rubric, and help inform professional development goals and growth plans.</li> </ul>

Measure	Description	Research	Strengths	Cautions
<p>Parent and Student Surveys</p>	<p>These surveys are used to gather parent and student opinions or judgments about the effectiveness of the principal's practices or the effectiveness of the school in meeting the interests and needs of parents and students. Survey results factor into principal evaluation.</p>	<ul style="list-style-type: none"> <li>■ The use and effect of parent and student surveys for principal evaluation purposes have not been examined in research literature, although many states and districts use these surveys as part of principals' evaluation.</li> <li>■ Several studies have shown that high school, middle school, and elementary student ratings may be as valid as judgments made by college students and other groups and, in some cases, may correlate with measures of student achievement.</li> <li>■ Several studies have shown that parental involvement with the school has an impact on student achievement.</li> </ul>	<ul style="list-style-type: none"> <li>■ Provides the perspective of students and parents or guardians on principal leadership or school conditions.</li> <li>■ Can provide formative information to help principals improve practice in a way that will connect with and impact students.</li> <li>■ Makes use of the perspectives of students, who may be as capable as adult raters at providing accurate ratings.</li> </ul>	<ul style="list-style-type: none"> <li>■ Student and parent ratings have not been validated for use in summative assessment and should not be used as the sole or primary measure of teacher evaluation.</li> <li>■ Students and parents cannot provide information on all roles of the principal.</li> </ul>
<p>School Climate Surveys</p>	<p>These surveys are commonly used to measure the perceived presence of teaching and learning conditions and gauge changes in perceptions over time.</p> <p>They are typically administered annually to educators, staff, students, and possibly parents to gauge the relative presence of certain traits or practices in a school.</p>	<ul style="list-style-type: none"> <li>■ School climate represents a set of organizational traits that research indicates are associated with robust and encouraging outcomes, such as better attendance, higher morale, and increased academic effectiveness.</li> <li>■ Research studies have shown that teachers stay employed longer at schools with positive climate, and this consistency benefits students' academic achievement.</li> </ul>	<ul style="list-style-type: none"> <li>■ Provides a way of measuring direct effects of principal effectiveness related to school-level conditions, such as the ability to influence student learning by working directly with teachers to improve instruction and creating safe, healthy, and effective schools where strong teaching and learning are valued.</li> <li>■ Can provide formative and summative information to help principals improve their practice.</li> <li>■ Based on frequency of administration, can provide data to benchmark change over time.</li> </ul>	<ul style="list-style-type: none"> <li>■ Any survey that forms part of a high-stakes principal performance assessment should be valid and reliable to ensure its accuracy and applicability in measuring principal performance.</li> <li>■ Principal effectiveness is a multifaceted construct, and its assessment might require multiple measures to develop a holistic picture of performance.</li> </ul>

Measure	Description	Research	Strengths	Cautions
360-Degree Surveys	<p>Using a survey format, 360-degree approaches gather and compare perception-based feedback from multiple constituents (e.g., the principal, staff, teachers, parents, students, supervisors) to create an aggregate profile of a principal's performance on specific competencies. This approach, usually paired with mentoring and coaching, is designed specifically to help principals to reflect holistically on their performance through self-assessment and examining feedback from their key constituents.</p> <p>Unlike stand-alone perception surveys, 360-degree surveys include principal self-assessment using a common set of survey questions and topic areas, which allows a principal's perspective to be compared with the perceptions of other constituents. Traditional 360-degree instruments are uniquely designed for each constituent type; it is possible, however, to use stand-alone staff, parent, and student surveys for 360-degree purposes if the questions and topics are similar and the principal uses the survey questions to engage in self-assessment.</p>	<ul style="list-style-type: none"> <li>■ Despite their rising popularity in principal evaluation, rigorous research on the effect of 360-degree surveys on principal performance is lacking.</li> <li>■ Studies of 360-degree approaches in other fields have provided mixed results but suggest that this approach works best when used as part of a coaching model.</li> </ul>	<ul style="list-style-type: none"> <li>■ Provides a wide range of feedback about a principal's performance, usually on a number of important components of leadership across multiple roles.</li> <li>■ Designed to facilitate both broader and deeper principal self-reflection by providing access to more data during the self-assessment process.</li> <li>■ Enables multiple constituents to provide feedback that can easily be compared and that is intended for formative development of the principal.</li> </ul>	<ul style="list-style-type: none"> <li>■ The 360-degree approaches rely on perception-based data and were originally designed to support principal self-reflection and principal coaching; 360-surveys should not be used as a single, stand-alone measure of principal performance.</li> <li>■ The 360-degree surveys work best when incorporated into formative evaluations combined with strong coaching. The 360-survey data should be incorporated into summative evaluations with caution and only as part of the self-assessment component in a broader evaluation model.</li> </ul>

Center on  
**GREAT TEACHERS & LEADERS**

at American Institutes for Research ■

1000 Thomas Jefferson Street NW  
Washington, DC 20007-3835  
877.322.8700

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



AMERICAN INSTITUTES FOR RESEARCH®

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

Copyright © 2014, 2012 American Institutes for Research. All rights reserved.

This work was originally produced in whole or in part by the Center on Great Teachers and Leaders with funds from the U.S. Department of Education under cooperative agreement numbers S283B120021 and S283B050051. The content does not necessarily reflect the position or policy of the Department of Education, nor does mention or visual representation of trade names, commercial products, or organizations imply endorsement by the federal government.

The Center on Great Teachers and Leaders is administered by American Institutes for Research and its partners: the Council of Chief State School Officers and Public Impact.