SEP³ Toolkit

STATE EVALUATION OF PRINCIPAL PREPARATION PROGRAMS

FEBRUARY 2016

New Leaders

SEP³ Toolkit | STATE EVALUATION OF PRINCIPAL PREPARATION PROGRAMS

ACKNOWLEDGMENTS

EXPERT ADVISORY GROUP MEMBERS

Mónica Byrne-Jiménez, Hofstra University (New York)

Mary Canole, Council of Chief State School Officers

Stevie Chepko, Council for the Accreditation of Educator Preparation

Matthew Clifford, American Institutes for Research

Shelby Cosner, University of Illinois at Chicago

Brian Dassler, Florida Department of Education

Jacquelyn Davis, George W. Bush Institute

Benjamin Fenton, New Leaders

Susan Gates, RAND Corporation

Mark Gooden, University of Texas-Austin

Jackie Gran, New Leaders

Steven Gross, Temple University (Pennsylvania)

Sara Heyburn, Tennessee Department of Education

Susan Korach, Richie Program, University of Denver (Colorado)

Paul Manna, College of William and Mary

Tricia McManus, Hillsborough County Public Schools (Florida)

Glenn Pethel, Gwinnett County Public Schools (Georgia)

Diana Pounder, University of Central Arkansas

Frances Rabinowitz, Bridgeport Public Schools (Connecticut)

Carol Riley, National Association of Elementary School Principals

Cortney Rowland, National Governors Association

Christopher Ruszkowski, Delaware Department of Education

Erin Swanson, Martin Millennium Academy (North Carolina)

Brenda Turnbull, Policy Studies Associates

David Volrath, Maryland Department of Education

TABLE OF CONTENTS

Guide

SEP³ Toolkit

STATE EVALUATION OF PRINCIPAL PREPARATION PROGRAMS

Guide

FEBRUARY 2016

New Leaders

uc ea — University Council for Educational Administration
Quality Leadership Matters   since 1954

I. INTRODUCTION

Strong school leadership—at the assistant principal, principal and principal manager levels—is critical for improving school performance at scale. Decades of research have established the central role of principals in raising student achievement and, in particular, creating the conditions for teachers to improve their practice.[1]

For principals to be successful at their core work of improving student learning, they need to be well-prepared.[2] While state strategies should address a range of policies and practices that influence the effectiveness of principals, including hiring practices and professional development, initial preparation in state-approved educational administration programs offers a critical moment to set high standards for entry into the role and ensure aspiring leaders have the knowledge and tools they need to be successful.

Currently, many graduates of principal preparation programs are not ready to assume assistant principal or principal positions.[3] While a number of programs have made major improvements and are pioneering powerful, evidence-based practices—such as providing job-embedded opportunities for participants to practice leadership skills and receive feedback[4]—these research-based practices are not widespread across programs. Making them commonplace is central to improving the quality of principal leadership across a state.

A growing number of education leaders and policymakers are working to improve the quality of educational administration programs in their state.[5] This report is designed to inform these efforts. It offers guidance and recommendations on how states can improve principal preparation by increasing the depth and rigor of their principal preparation evaluation process, thus enabling them to accurately assess quality, promote program improvement, and intervene when performance is not satisfactory. The report also reviews contextual factors that states should consider when adapting the recommendations to local conditions, and includes supplementary resources and tools to help states carry out the work. Together, these materials will allow states to undertake an informed and sophisticated approach to the complex work of improving principal preparation.

States are uniquely positioned to influence on the quality of principal preparation programs because most have relevant statutory authority. Specifically, most states grant initial and ongoing approval for principal preparation programs to operate and they issue licenses for individuals to serve as principals. Despite this central role in authorizing principal preparation programs, states lack strong models for assessing the quality of programs to promote improvement. We see two central problems of practice for states and programs:

Why focus on just principal preparation rather than more broadly on educational leadership?

Cultivating leadership for schools includes supporting the development of teacher leaders, assistant principals, principals, and school system leaders. The majority of university-based educational administration programs support leadership development at most or all of these levels. While many ideas addressed here may be relevant to the broader conversation about school leadership, the recommendations set forth here are focused solely on principal preparation because of the important role principals play in improving student learning and because states are responsible for certifying whether an individual is adequately prepared to assume a formal school administration role (i.e. as a principal or assistant principal).

---

1. Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning.* New York, NY: Wallace Foundation.; Louis, K. S., Leithwood, K., Wahlstrom, K., & Anderson, S. (2010). *Learning from leadership: Investigating the links to improved student learning.* New York, NY: The Wallace Foundation.; Marzano, R.J., Waters, T., & McNulty, B. (2005). *School leadership that works: From research to results.* Alexandria, VA: Association for Supervision and Curriculum Development.
2. While this document does not focus on the conditions for effective leadership, that by no means minimizes their importance. See *Great Principals at Scale* for a description of other conditions contributing to effective principal leadership. Ikemoto, G., Taliaferro, L., Fenton, B. & Davis, J. (2014). *Great principals at scale: Creating district conditions that enable all principals to be effective,* New York, NY: New Leaders. http://www.newleaders.org/wp-content/uploads/GPAS_Executive-Summary_Final.pdf.
3. Hull, J. (2012). The principal perspective: Full report. Alexandria, VA: Center for Public Education; Young, M. D., & Brewer, C. (2008). Fear and the preparation of school leaders: The role of ambiguity, anxiety, and power in meaning making. *Educational Policy,* 22(1), 106-129.
4. Darling-Hammond, L. LaPointe, M., Meyerson, D., & Orr , M. T. (2009). *Preparing principals for a changing world.* San Francisco, CA: Jossey-Bass.; Davis, S., & Darling-Hammond, L. (2012). Innovative principal preparation programs: What works and how we know. *Planning and Changing,* 43(1/2), 25-45.
5. The Council of Chief State School Officers. (2012). *Our responsibility, our promise: Transforming educator preparation and entry into the profession.* Washington, DC: CCSSO.

1. Evaluations should be driven by data about program quality and outcomes but state systems for collecting and interpreting data are often too limited to support such an approach (i.e., they lack direct measures of program quality and outcomes and the available data for indirect measures are insufficient or of low quality); and

2. Evaluations should be diagnostic in nature, using program quality and outcomes data to drive inquiries into the sources of successes, improvements, and concerns. However, state systems do not tend to be organized to use data in this way and states often lack the necessary capacity for diagnosis and support.

Regardless of where authority is located within state government (e.g., in some states, all of the relevant authority resides with the state board of education and the state education agency, while in others it is held by a professional standards board or is shared across multiple agencies), this guide offers recommendations and guidance for addressing these two problems of practice, To that end, the remaining sections of the guide are structured as follows:

Section II: Core Design Principles—This section outlines a set of design principals related to purpose, professional standards, data collection and use, and the process of review to which all effective program evaluation systems should adhere.

Section III: A Model Two-Stage Process for Program Evaluation—This section presents:

- **A model for the annual reporting of data on program quality and outcomes (Stage 1).** We present a possible design for a public online portal with data on important indicators of quality for each principal preparation program in the state.

- **A model for in-depth review of all programs and targeted review of programs to address concerns (Stage 2).** We suggest specific improvements to current program review practices to foster a stronger focus on continuous improvement; further, we recommend a diagnostic process for understanding and addressing concerning data on indicators of program performance.

- **Recommendations for choosing indicators of program quality and outcomes.** We identify the specific data that states can use to support the models described above.

- **Recommendations for making summative judgments about program effectiveness.** We offer guidance to states on making summative decisions about programs based on in-depth review and, in some cases, targeted review.

Section IV: Conditions—This section describes baseline conditions for effective program review. It also identifies factors states should consider as they assess their own capacities and priorities before planning and implementing changes to their principal preparation program review processes.

Section V: Tools—This section describes supplementary tools that are designed to help states carry out the recommendations in the guide. These include:

- **State Readiness Diagnostic Rubric (Tool A).** This tool enables states to assess their readiness to design and implement a program evaluation system that reflects the recommendations in the guide.

- **Annual Report Indicators, Reporting, and Interpretation of Results (Tool B).** This tool provides states with specifications for an annual report that would give programs, candidates, and districts basic data on inputs, processes, outputs and program graduate outcomes.

- **Handbook for the In-depth Review Process (Tool C).** This tool describes a detailed process for undertaking comprehensive reviews of individual programs.

- **Handbook for the Targeted Review Process (Tool D).** This tool describes a detailed process for undertaking targeted reviews of individual programs.

Section VI: Resources—This section describes supplementary resources that provide background and context for states as they examine and revise practices related to the review of principal preparation programs. These include:

- **Overview of Current Program Review Practices (Resource A).** This resource provides information about current principal preparation program evaluation practices in states.

- **Review of Other In-Depth Program Review Processes (Resource B).** This resource provides states with a description of the four types of reviews that most higher education leadership preparation programs already undergo so that states can design their system with these existing review processes in mind.

- **List of Other Tools and Resources (Resource C).** This resource provides states with links and citations to tools, research, and resources that could inform their work in this area.

The two-stage model described in Section III is an approach that meets all of the design principles in Section II, but it is not intended for "off-the-shelf" application. States inevitably will need and want to adapt the model to address issues specific to their context. As many advisors and reviewers attest, details and methodology matter a great deal with this type of work. We encourage states to adhere to core design principles as much as possible, and to enlist the support of experts in determining the optimal approach to meeting the design principles within their particular context.

How were these materials developed?

These resources were developed by UCEA and New Leaders in deep and iterative collaboration with state leaders, district leaders, principals, researchers, preparation faculty, and representatives of national organizations committed to high quality educational leadership.

We began by reviewing previous reports about principal preparation program evaluation, principal performance evaluation and state capacity. These reports are listed in the reference section of the guide. The review revealed that the field lacks adequate guidance for states on how best to enact their statutory authority to effectively review programs and the tools to help them do so. We concluded that filling this gap would require drawing on both published research and on the expertise of practitioners and other stakeholders.

To gather the relevant perspectives, we assembled an advisory group consisting of state representatives with experience designing or implementing preparation evaluation systems, methodologists with experience evaluating principal preparation programs, representatives of national organizations focused on issues of leadership preparation, principal preparation program leaders with experience evaluating their own or other programs, district leaders with experience evaluating internal or external programs, and principals.

Subsets of the group participated in four webinars based on their areas of expertise. The webinars addressed specific questions related to: (1) state authority and leadership, (2) data considerations, (3) rigor of outcomes and process, and (4) consumer needs and priorities. We used ideas from these webinars to draft an initial guide and supplemental materials. We then convened the advisory group for a two-day design session to review the drafted documents and bring new ideas and solutions to the conversation. From this session, we re-drafted all of the materials and solicited feedback from the group one more time. In this last stage, we also invited a wider range of stakeholders—including state education agency officials, preparation program providers, and methodologists—to comment on the documents. Before finalizing the materials, we asked five experts to provide in-depth reviews and feedback on the content.

Throughout this process, we have worked collaboratively, bringing the perspectives and knowledge from our two respective organizations (and from the many individuals and organizations that provided feedback) to tackle difficult design decisions. We believe the resulting product will aid states in their quest to improve principal preparation.

New Leaders is a national nonprofit that develops transformational school leaders and designs effective leadership policies and practices for school systems across the country. New Leaders has trained over 2,500 SCHOOL leaders nationwide who are currently affecting over 450,000 students. Since its inception, New Leaders has engaged in rigorous internal and external program evaluation, including a current i3 grant in partnership with the RAND Corporation, to inform program improvement and address accountability. New Leaders has also provided training to other preparation programs on how to design and conduct program evaluation.[6]

The University Council for Educational Administration (UCEA) is a consortium of 99 higher education institutions committed to advancing high-quality educational leadership preparation, research, policy and practice for the benefit of schools and children. Over the course of its 60-year history, UCEA programs have produced thousands of building and district level leaders and produced cutting edge research on the practice and preparation of educational leaders. UCEA works collaboratively with higher education and professional organizations to build the knowledge base on effective leadership preparation, to design and utilize preparation program standards, and to build evaluation tools and practices designed to improve the preparation and professional development of educational leaders and professors.[7]

---

6. Neuman-Sheldon, B., Ikemoto, G., Bailey, M., Erdfarb, T., Nerenberg, L., Patterson, N., & Valdez, M. (2014). *Principal preparation program self-evaluation: Lessons learned by New Leaders.* New York, NY: New Leaders.
7. University Council for Educational Administration. (2013). *Developing evaluation evidence: A formative and summative evaluation planner for educational leadership preparation programs.* Charlottesville, VA: Author.

New Leaders

UCEA University Council for Educational Administration
Quality Leadership Matters   since 1954

II. CORE DESIGN PRINCIPLES

Five core design principles inform the development of principal preparation program evaluation systems. Effective evaluation systems:

1. Promote continuous program improvement.

2. Support states in holding programs accountable for effective practices and outcomes.

3. Provide key stakeholders with accurate and useful information.

4. Are sophisticated and nuanced in their approach to data collection, analysis, and use.

5. Adhere to characteristics of high-quality program evaluation.

1. Promote continuous program improvement.

Effective program review systems encourage improvement and innovation in program design and implementation by doing two things: providing programs with specific and actionable feedback about their practices and outcomes, and allowing adequate time for programs to make changes and assess their impact. To provide this level of accurate and actionable feedback, systems must employ program reviewers who have relevant expertise for making appropriate judgments. The reviewers should possess content expertise in leadership, an understanding of adult learning theory and practices, knowledge of current research about effective leadership preparation, and the ability to accurately assess curriculum and pedagogy.

2. Support states in holding programs accountable for effective practices and outcomes.

An evaluation system is a key way for states to hold preparation programs accountable for delivering high quality preparation for aspiring principals. With approximately 1,000 programs currently in operation[8] and new ones emerging on a regular basis, states need to be able to confidently make consequential decisions such as whether to approve a program, when to put a program on an improvement plan and, in the most serious circumstances, when to rescind program approval. States need to understand the limitations of the indicators they track and ensure they have sufficient and valid information for making consequential decisions. States also need sound program ratings, based on a sufficient number of indicators to meaningfully capture performance and improvement over time. Finally, states need a clear process and timeline for intervening when programs demonstrate unacceptable performance.

3. Provide key stakeholders with accurate and useful information.

When key consumers and partners—especially aspiring school leaders and school districts—have good information about key program indicators, they can use that information to make more informed choices. For aspirants, a state evaluation system can provide concrete information about program features and outcomes (e.g., candidate learning and career outcomes)—including side-by-side, apples-to-apples comparisons—thus helping them choose high-quality programs. For districts, the same information can guide decisions concerning formal partnerships with programs and the hiring of graduates. To meet these purposes, effective evaluation systems make high-quality, easily understandable program data available to the public. (See the sidebar describing important considerations about making data publicly available.)

---

8. Cheney et al.

4. Are sophisticated and nuanced in their approach to data collection and use.

This nuanced approach is guided by the following five precepts:

- **Evaluate what matters.** The data system includes the indicators that are most germane to preparation. We define program effectiveness in terms of inputs (especially the rigor of selection into a program), processes (especially the ways in which a program increases aspirants' leadership knowledge and skills), outputs, (especially aspirants' successful placement in roles as principals), and contributions to important graduate outcomes (especially outcomes for students, including academic achievement measures, attainment measures such as graduation, and non-cognitive measures such as engagement and social/emotional growth).

- **Evaluate accurately.** The data system uses the most accurate data available, and interpretations are made cautiously, with awareness of data limitations. Valid and reliable measures of leadership effectiveness are still in the early stages of development but, once confidence in their accuracy is established, could be a part of the review process. The system takes into account limitations related to reliability and validity in determining whether and how much to weigh particular data sources in evaluation.[10]

- **Include data that can be realistically gathered and shared.** Data are feasible to gather, efficient to report, and possible to corroborate with other sources of information. Further, data collection is ongoing and conducted according to a consistent schedule.

What are considerations for states in responsibly sharing data about programs?

Design Principle 3 recommends that some information about programs be made public, but this should not be done until after states have an opportunity to pilot and make improvements to the system. Providing select information to the public serves several purposes: it helps aspirants make informed judgments about the quality of preparation programs available to them; it helps districts determine whether they want to partner with particular programs; and it helps programs become familiar with statewide practices, understand the factors driving their own results, and use that knowledge to make necessary improvements.

Nevertheless, as with any effort to distill and report data, the challenge lies in the details, and there are reasons to be exceptionally cautious. One concern is that data related to programs can be misleading or misinterpreted. For example, the number of graduates placed as principals can be a valuable indicator of program quality, but it is not useful for comparing programs in urban areas (with dozens of principal vacancies each year) with those in rural areas (with a very low number of annual vacancies).

A second concern is that specific data points might distort decision-making by program leaders. For example, if higher placement data is a priority, programs might preference males during admissions because research shows they tend to be placed more quickly than females. A third concern involves the field's current lack of strong and reliable measures of graduate leaders' impact on student achievement. Student achievement can be influenced by factors such as varied turnover rates, timeframes and sample sizes, thus are not always dependable measures of graduates' effectiveness or the influence of the preparation program.[9]

While these concerns do not outweigh the utility of making data public, they do suggest the need for processes that enable states to understand the limitations of their data and interpret and communicate program results responsibly. Specifically, we recommend that states spend a year or two collecting and analyzing data, and discussing their appropriateness with programs, before reporting them publicly. This trial period would provide states with an opportunity to identify specific challenges and develop solutions to mitigate inappropriate interpretations and perverse incentives. Then, and only then, should states make data public.

---

9. Fuller, E. J., & Hollingworth, L. (2014). A bridge too far? Challenges in evaluating principal effectiveness. *Educational Administration Quarterly*, 50(3): 466-499.

10. It is important to be clear about the meaning and purpose of terms such as validity, reliability, and appropriateness. We use these terms as follows: Valid = Data are accurate and adequate measures of well-defined constructs (i.e., elements of the program and outcomes that are being measured). Reliable = Data sources produce consistent information about the chosen constructs across repeated measures, over time, and for different programs and contexts. We encourage states to use these definitions as guides and to take the time to assess how specific data sources meet them.

NewLeaders

University Council for Educational Administration
Quality Leadership Matters since 1954

- **Consider contextual factors.** Data are means, not ends. In order to make appropriate judgments based on accurate results, states gather additional contextual information. Basic indicator results can be difficult to interpret on their own, but can be the basis of productive investigation into and conversations about program quality and improvement. Analyses of program-related data inform judgments about program status and the need for continued program development.

- **Clearly and transparently communicate how results will be used.** Programs understand which data will be made public, including how and when this will occur. Programs also understand how component parts of the program evaluation will be used to make substantive judgments and decisions about program status.

5. Adhere to characteristics of high-quality program evaluation.

An effective state system of program evaluation reflects what we know about best practices in program evaluation in education. We recommend the Standards for Educational Evaluation as a basis for judging best practices. These standards focus on utility (i.e., the extent to which stakeholders find processes and results valuable), feasibility (i.e., the effectiveness and efficiency of evaluation processes), propriety (i.e., the fairness and appropriateness of evaluation processes and results), accuracy (i.e., the dependability of evaluation results, especially judgments of quality), and accountability (i.e., having adequate documentation to justify results).[11]

---

11. See http://www.jcsee.org/program-evaluation-standards for more details on the standards, which were issued by the Joint Committee on Standards for Educational Evaluation

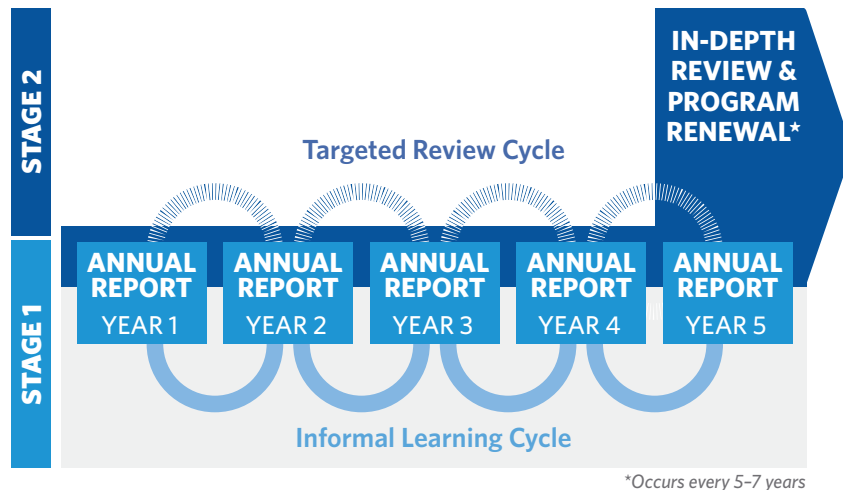## III. A MODEL TWO-STAGE PROCESS FOR PROGRAM EVALUATION

Any number of models could meet the design principles, and states might have different approaches based on preferences, methodological considerations, and/or contextual conditions. In this section, we present one approach for evaluating principal preparation programs that our reviewers agree is particularly well suited to meeting the design principles. This approach involves an iterative, **two-stage process**. The first stage involves collecting data on every program on an annual basis; the second stage involves investigating those data more deeply to promote continuous improvement.

We posit that our recommendations and model meet the design principles outlined in Section II and help answer four fundamental questions about program quality:

1. Is selection into a program rigorous, so that aspirants have the potential to be effective principals?

2. Does the program increase aspirants' leadership knowledge and skills?

3. Do graduates assume roles as principals and assistant principals and are they ready to lead when they do so?

4. Do graduates have a substantial positive impact on student learning and other important school outcomes?

Our recommended iterative, **two-stage** process is illustrated in Figure 1.

Figure 1: Two-Stage Process

Stage 2

Targeted Review Cycle

IN-DEPTH REVIEW & PROGRAM RENEWAL*

Stage 1

ANNUAL REPORT YEAR 1 · ANNUAL REPORT YEAR 2 · ANNUAL REPORT YEAR 3 · ANNUAL REPORT YEAR 4 · ANNUAL REPORT YEAR 5

Informal Learning Cycle

Occurs every 5–7 years

Stage 1 involves the collection of information about all programs in the state and the publication of results in an annual report. Indicators summarized in the report can be a powerful way to convey information to prospective students and other stakeholders. This type of information can also guide an informal learning cycle and serve as an important starting point for program review, improvement and accountability processes.

A note on the tools

The accompanying tools provide more detail on this model, including a sample Annual Report and detailed recommendations for in-depth review processes and summative ratings. It is important to note that these tools are not intended to be used "off the shelf." In the absence of existing state examples that meet the design criteria, states have requested tools that exemplify the design principles and include detailed recommendations for critical decisions, such as what indicators to include and how to use outcome data. While the recommendations presented in the tools were developed with extensive input and feedback from state officials, preparation programs, and methodologists from a range of state contexts, we expect that states will want to adapt them, seeking technical assistance to determine strategies for meeting design principles in ways that fit with local conditions. The local conditions to consider are described in Section IV.

New Leaders

University Council for Educational Administration
Quality Leadership Matters   since 1954

Stage 2 involves an in-depth review of program practices and outcomes. The purpose of the in-depth review is to help states understand how well a program is designed and implemented, support continuous program improvement, surface effective practices that can be shared with other programs, and explore and address any areas of under-performance.

The two stages complement each other in significant ways. For example, while an annual report provides high-level information that allows states to make general comparisons across programs, *the underlying reasons for data results are often difficult to interpret and can be misleading without further analysis and review. As a result, we do not recommend drawing conclusions about program quality based on these indicators alone; rather, the results should be used diagnostically to flag areas of concern that warrant further investigation.* The in-depth review, when conducted by qualified reviewers with sound tools and processes, can provide sufficient information for drawing conclusions and making high-stakes decisions about program approval and renewal.

In-depth reviews happen on a relatively long cycle—every five to seven years. Since states may need to address concerns about program quality and outcomes on a shorter cycle, we also propose that states have an alternative, focused process for deeper investigation of data that raise concerns. To distinguish it from the in-depth review process, we call it a targeted review.

Targeted review involves the identification, diagnosis and improvement of specific areas of under-performance. The basic purpose of a targeted review is to explore and address areas of possible under-performance so that programs can improve quickly or states can take action. Programs may want to independently engage in informal learning cycles similar to the targeted reviews in order to improve practices. However, given limited resources, states should prioritize targeted reviews for programs with significant areas of under-performance.

Figure 2: Annual Cycles Embedded in Two-Stage Process

*Occurs every 5–7 years
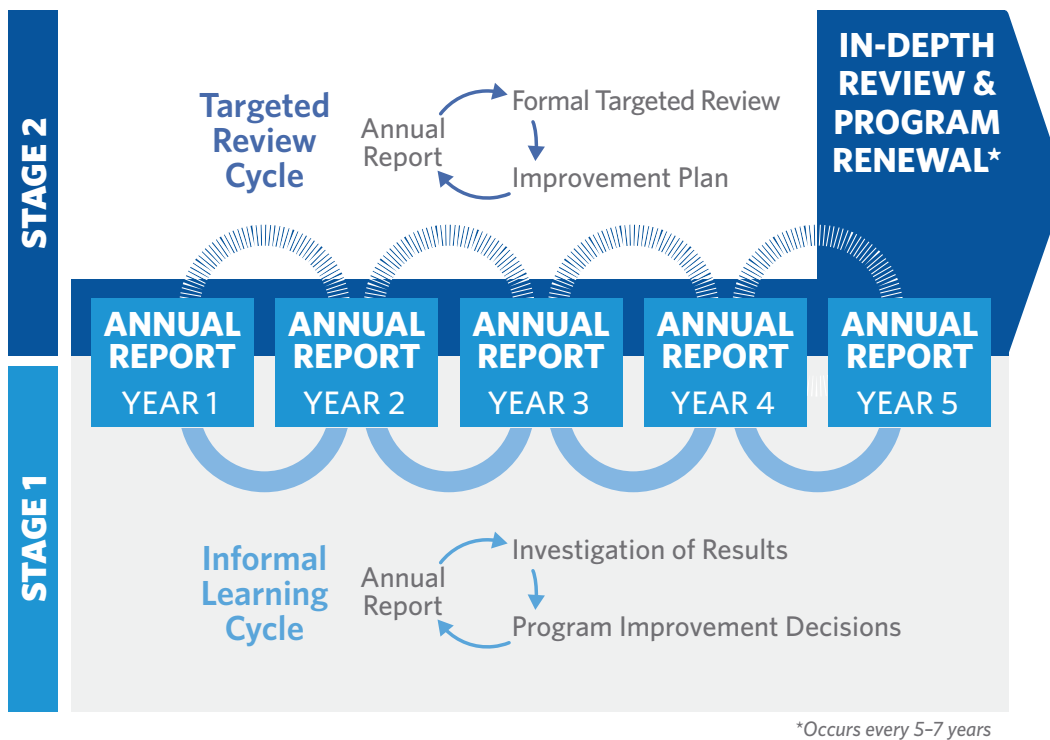
The rest of this section describes the component parts of our model approach, including: (A) annual report, (B) in-depth review, (C) targeted review, (D) indicators, and (E) summative evaluation.

A. Annual Report

We envision the annual report as an online portal with two levels of data: (1) a summary page with basic information for each principal preparation program in the state, and (2) a series of program pages, each with more detailed data for a particular principal preparation program. Tool B (Annual Report Indicators, Reporting, and Interpretation of Results) provides an example of a populated report at each level.

One important purpose of the annual report is to provide data for public review, particularly for aspiring principals and districts. The summary page offers a color-coded visual representation of various programs, highlighting areas of strength and weakness for each of eighteen indicators across programs. At a glance, consumers can ascertain how programs compare to one another across the indicators and whether they meet state-established standards. Meanwhile, the program-specific pages provide context for each program's data, allowing for more nuanced comparisons.

Of course, most indicators included in the annual report are just that, indicators: they are not direct measures of program quality or impact and **should not be used for consequential decisions**, such as program approval, renewal, or funding. At the same time, states can and should use the annual report to determine the timing and focus of in-depth program review (see below). Tool B provides additional details about the purpose and structure of an annual report.

B. In-depth Review

Every program should be reviewed in-depth once every seven years (or more frequently if resources allow). This practice promotes ongoing improvement to recruitment, selection, and program design, and focuses attention on ensuring that graduates are well prepared, are assuming leadership roles, and are having an impact on relevant school and student outcomes.

In most states, such reviews already take place. Indeed, programs typically undergo four types of reviews: (1) accreditation reviews designed to certify that programs conferring degrees meet standards of quality; (2) state reviews designed to monitor quality and help states make consequential decisions about initial and ongoing approval of programs; (3) institutional reviews which are required by the university in which the program is housed and are designed to foster self-study and ongoing improvement; and (4) professional association reviews designed to promote improvement against standards of excellent practice.[12]

We do not recommend that states add a new process. Rather, we recommend systemic improvements to any process that a state currently uses. Specifically, our model of in-depth review makes four upgrades to current practices:

1. In-depth review builds on data patterns identified in the annual report, enabling state personnel (or their designees) to examine the quality of a preparation program in terms of inputs, processes, outputs, and graduate outcomes. Importantly, in-depth review asks additional questions:

   - To what extent are the data in the annual report accurate reflections of program quality?

   - Why is the program meeting or not meeting expectations on a given indicator?

   - Can effective practices be replicated in other settings and/or scaled across the state?

---

12. Note that Resource B (Description of Other In-Depth Program Review Processes) describes each of these review processes in depth. It also describes additional tools (e.g., Quality Measures) that programs can use to self-assess.

New Leaders

University Council for Educational Administration
Quality Leadership Matters   *since 1954*

To answer these questions, the in-depth review includes the collection and review of additional data that can be more readily evaluated as part of an on-site review. (See sub-section C below on indicators for more details.)

2. In-depth review includes the development of a portfolio of practice, the review of that portfolio by a team of experts, and structured site visits. This sequence, which is used by some existing review processes, allows for programs to examine their own practices in all key areas—inputs, processes, outputs, and graduate outcomes—before receiving feedback from external reviewers.

3. In-depth review is conducted by a team of expert practitioners. The team should be made up of at least three professionals, including one state department of education representative, one faculty member from outside the institution, and one school or district administrator. Reviewers should possess relevant expertise for making professional judgments, especially with respect to the indicators driving the review.

4. In-depth review is anchored by rubrics that assess quality across all aspects of program design and implementation. We offer detailed rubrics (that states can use or adapt) drawn from UCEA's Institutional and Program Quality Criteria.[13] They focus on both program features and program outcomes. These program review rubrics were designed to illuminate the difference between program practices that are highly effective, effective, in need of improvement or ineffective for each criterion.

Tool C elaborates on these recommendations and practices, providing a detailed handbook for the implementation of in-depth review.

How do our recommendations fit with other developments related to the evaluation of principal preparation programs?

There are several promising developments coming from the federal government and accreditation agencies. The United States Department of Education (the Department) is calling on states to become more involved by proposing changes in the teacher preparation accountability requirements of Title II of the Higher Education Act (HEA). These changes support intensified monitoring of preparation programs in teacher education. The proposed rules would require states to use federally established guidelines to collect data on individual programs (including information from surveys of graduates and employers, graduate placement and retention rates, and growth in student achievement), rate individual programs (using four rating levels of "low-performing," "at risk," "effective," or "exceptional") and report results on a public state report card. Since the Department does not have statutory authority to include principal preparation programs, the current proposal does not explicitly address principal preparation. However, there are calls to include principal preparation in a revised HEA. The recommendations and tools outlined in this document do not presuppose the passage of federal statutes or finalization of regulations, but were formulated in a way that accounts for possible federal changes.[14]

The Council for the Accreditation of Educator Preparation (CAEP) conducts reviews of principal preparation programs, including those offered through higher education institutions, district- and state-sponsored programs and alternative providers. Principal preparation programs, which are categorized as advanced programs[15] by CAEP, are subject to the following five criteria: (1) candidates develop a deep understanding of the critical concepts and principles of their discipline and the ability to use that knowledge to advance student attainment; (2) partnerships and clinical practice are central features of the program; (3) candidate evaluation is used to inform candidate development throughout the program; (4) providers demonstrate the impact of program completers through a variety of school-level factors and provide indicators of program completer satisfaction with the quality of the program; and (5) providers maintain a quality assurance system comprised of valid data from multiple measures, including those mentioned above. As CAEP continues to make upgrades to its process and standards, states could consider using the CAEP process to supplement or replace the in-depth review process described in this document. States may still want (or be required by federal law) to issue report cards and address low-performing programs.

13. Young, M. D., Orr, M. T., & Tucker, P. D. (2012). *University Council for Educational Administration (UCEA) Institutional and Program Quality Criteria: Guidance for masters and doctoral programs in educational leadership*. Charlottesville, VA: UCEA.

14. Note that UCEA authored a policy statement responding to the draft regulations. That statement can be found here: http://www.ucea.org/resource/marshaling-and-using-resources/

15. CAEP (2014). *Standards for advanced programs*. Washington, DC. Accessible online: https://caepnet.files.wordpress.com/2014/08/caep_standards_for_advanced_programs1.pdf

C. Targeted Review

While every program should use data patterns in the annual report to support their own improvement efforts, states have limited capacity to support improvement in all programs every year. We thus recommend that states prioritize efforts to improve programs whose annual report data raise the greatest concerns. These programs should receive a targeted review to confirm or disconfirm the concerns raised by the annual report.

Like in-depth review, targeted review builds on the annual report data and asks additional questions of programs:

- To what extent are the data in the annual report accurate reflections of program quality?
- Why is the program meeting or not meeting expectations on a given indicator?
- What should be done to improve program performance?

The process for targeted in-depth review proceeds as follows:

1. State leaders identify the particular indicator(s) of concern needing deeper investigation and assign a review team.

2. Program leaders assemble additional data relevant to the indicator(s) of concern and assemble knowledgeable program staff and stakeholders to contribute to the review. Since the review is targeted and focused on developing an in-depth understanding of the indicator(s) of concern, the data collected should be likewise focused and specific.

3. The on-site review team analyzes available data, develops conclusions, and identifies recommended action steps for program improvement. It is critical to employ a rigorous process for analyzing data on indicators of concern and developing sound conclusions and action steps. We recommend two approaches to analysis—root-cause analysis and gap analysis—that are detailed in Tool D.

4. State-designated reviewers complete a draft summary report identifying action steps for the program and the state. The program has an opportunity to respond to any and all descriptions, recommendations and action steps before the report is finalized.

It is crucial to note that reviewers might conclude that the indicators triggering the review do not in fact reflect problems of program quality. In that case, steps 3 and 4 above might not require actions steps for the program.

Like with in-depth review, the review team should be made up of at least three professionals, including one state department of education representative, one faculty member from outside the institution, and one school or district administrator. Reviewers should possess relevant expertise for making professional judgments, especially with respect to the indicators driving the review.

Tool D provides more detail about all aspects of targeted review.

D. Indicators

At the heart of this approach is the idea that program quality should be assessed by examining data in four areas: program inputs, program processes, program outputs, and graduate outcomes.[16] All four areas should be part of the annual report and all four areas should be part of in-depth review.

Figure 3: Categories of Indicators

	Data collected, analyzed and reported annually	Data collected and analyzed during in-depth review & targeted review
Program Inputs	Indicators that reflect the program's ability to recruit and select high-potential aspirants and to diversity the pool of aspiring principals	
Program Processes	Indicators that reflect the quality of learning experiences for aspiring principals	
Program Outputs	Indicators that reflect the success of aspirants in completing a rigorous program and being hired into principal and assistant principal roles	
Graduate Outcomes	Indicators that reflect the impact that program graduates have, both on practices in the schools they lead and on student learning	

Together, the annual report and further review (both in-depth and targeted) allow for a complete diagnosis in any given area that neither can accomplish alone. Consider an example:

An annual report shows disparities among programs in how graduates perceive the relevance of program content. For the vast majority of programs, 80%+ of graduates agree with survey items related to relevance, but for one or more programs, less than 70% of graduates agree with the items. This variation warrants further investigation. A targeted review might unearth the fact that the relatively low scores for a handful of programs stemmed from an issue that has already been addressed (such as differentiating assignments to better address relevance to secondary versus elementary schools), thus allaying concerns about the low scores. On the other hand, a targeted review might find that the program content does not address the needs of aspiring charter school leaders (even though many program participants have residencies in charter schools). This conclusion might spur an action plan for the targeted programs in which they are tasked with making content more relevant for the charter sector.

As this example illustrates, annual reports should not be used alone for high stakes decisions because many of the indicators are not direct measures of program quality and/or could be misleading on their own.

The table below defines the four categories of indicators and provides specific data that could be part of an assessment of program quality. The annual report prioritizes indicators that are important, that can be obtained annually, and that do not require an unreasonable amount of effort or resources to obtain and compile on an annual basis. The below recommended indicators for the annual report align to the design principles, but ultimately the type and number of indicators will likely need to be determined on a state-by-state basis depending on data availability and quality. The in-depth review indicators listed are not recommended for the annual report because they are less likely to be available across all programs, require significant resources to obtain, compile and analyze on an annual basis, and/or include artifacts that a state is unlikely to have capacity to review for every program every year. Tool B provides rationale for each indicator as well as definitions and considerations for interpreting results of these indicators. Tool A supports states in assessing which data are currently available to them.

---

16. McDavid, J. C., & Hawthorn, L. R. L. (2006). *Program evaluation & performance measurement: An introduction to practice.* Thousand Oaks, CA: Sage Publications.

Indicators for Annual Report	ADDITIONAL Indicators and Data for In-Depth Review
Program Inputs: Indicators that reflect the program's ability to recruit and select high-potential aspirants and to diversify the pool of aspiring principals	
1. Admissions rate 2. Teaching experience of admitted candidates 3. Strength of instructional expertise of admitted candidates 4. Demonstrated leadership potential (through experiences leading adults) of admitted candidates	1. Rigor and quality of recruitment and selection processes
Program Processes: Indicators that reflect the quality of learning experiences for aspiring principals	
1. Graduate perceptions of program content 2. Graduate perceptions of quality of faculty 3. Graduate perceptions of quality of peer interactions 4. Use of performance-based assessments 5. Adequacy of internship/residency hours	1. Experience and expertise of instructors 2. Use of a defined competency framework aligned to leadership standards 3. Use of research-based content, curriculum, instructional and assessment practices 4. Implementation of supervised clinical practice with authentic leadership opportunities 5. Presence of collaborative relationships for program enrichment 6. Use of evaluation practices to support improvement
Program Outputs: Indicators that reflect the success of aspirants in completing a rigorous program and being hired into principal and assistant principal roles	
1. Program graduation rate 2. Licensure rate 3. Placement (in school leadership roles) rate 4. Retention (in school leadership roles) rate	1. Measures of graduate knowledge and skills developed through participation in the program 2. Graduate perceptions of readiness for leadership roles based on participation in the program 3. 360° evaluations of program graduates
Graduate Outcomes: Indicators that reflect the impact that program graduates have, both on practices in the schools they lead and on student learning	
Graduate and School Practices 1. Leadership effectiveness of graduates 2. Improvement in school climate in schools led by graduates 3. Improvement in teacher effectiveness in schools led by graduates *Student Outcomes* 1. Improvement in non-cognitive measures of student outcomes in schools led by graduates 2. Student achievement growth in schools led by graduates	*Additional Data and Analyses of Graduate and School Practices, including measures of:* 1. Climate 2. Student engagement 3. Discipline levels 4. Teacher morale *Additional Analyses of Student Outcomes Data (e.g., longitudinal, disaggregated), including:* 1. Improvement in non-cognitive measures of student outcomes in schools led by graduates 2. Student achievement growth in schools led by graduates

1. Program Inputs: Principal preparation programs should have rigorous recruitment and selection processes that result in admitting candidates who have demonstrable effectiveness as educators and potential to be effective principals.

Annual Review: Indicators to assess program inputs include program admissions rates, which demonstrate whether programs are adequately selective,[17] as well as the average number of years of teaching, the instructional expertise, and the demonstrated leadership potential (through experiences leading adults) of admitted candidates. Principals are more successful in leading improvements in teaching and learning when they have been effective classroom teachers,[18] and are more likely to be able to lead the work and learning of others if they have had experience leading grade-level teams or subject area departments. Thus, the percentage of admitted candidates with a record of classroom effectiveness and prior experience as teacher leaders can shed light on the selection progress by assessing whether admitted candidates have sufficient foundation for assuming leadership roles. All of these indicators would need to be reported by programs.

In-depth review: The in-depth review provides an opportunity for closer exploration of the types and quality of experiences that candidates bring, as well as the recruitment and selection processes. Some programs have developed important innovations—including competency-based assessments and recruitment strategies for target populations—that can be shared with other programs.

2. Program Processes: Principal preparation programs implement research-based and research-validated practices, have skilled faculty and/or staff, and create opportunities for authentic leadership practice.

Annual Review: Indicators include perception surveys of graduates and data reported from individual programs related to structures and requirements.[19] Graduates can inform the state through their ratings of the program's rigor and its relevance to their work as principals, the expertise and instructional effectiveness of faculty, and the quality of peer relationships. From programs, the state can learn about the quality of clinical learning experiences and whether they use performance-based assessments to evaluate candidate growth.[20]

In-depth Review: We recommend that the in-depth review of program processes focus on the extent to which a program:

- Has a defined competency framework (aligned to leadership standards and research) that describes the competencies a principal must have to influence school practices and culture in order to drive student achievement and non-cognitive growth.

- Employs conceptually coherent and research-based content, curriculum, instructional, and assessment practices that align to the program's competency framework and incorporate clinical learning strategies.

- Offers supervised clinical practice with authentic opportunities to apply and practice leadership competencies.

- Fosters collaborative relationships with other organizations such as universities and school districts to enrich learning experiences;.

- Engages in regular and rigorous evaluation and enhancement of its practices.

---

17. Admissions rates should be interpreted with caution, though. For example, low rates might simply indicate extensive recruiting relative to available spots rather than rigorous selection, while high rates could be a signal of recruiting that effectively targeted great candidates. However, an extremely high admissions rate (90%+) could be a sign that the rigor bar is too low and would require an explanation.

18. The field is currently attempting to build systems to create measures of teacher effectiveness, and the accuracy and validity of such systems are evolving. The implementation of sound measures is uneven; thus where available, states could consider: the percent of admitted candidates rated effective or above on measures of teacher effectiveness, and the percent of admitted candidates who improved student outcomes based on consistent and methodologically sound measures of aggregated individual student growth.

19. A common survey is necessary in order to compare results across programs and response rates should be collected and used to interpret results.

20. Effective leadership programs provide significant opportunities for participants to practice leadership skills in real-life settings through a residency or internship. Although the number of hours is not an indicator of quality for this program component, it provides important baseline information about the extent to which they exist.

3. Program Outputs: Effective leadership preparation programs are successful in preparing graduates for principal roles.

Annual Review: Recommended indicators include data on candidates' success in completing the programs, passing licensure exams, and being hired into and retained in principal roles. For example:

- The percentage of candidates who complete the program.
- The percentage of program completers who become licensed by the state (representing readiness to assume assistant principalships and principalships).[21]
- The percentage of licensed graduates hired as school leaders (assistant principals and principals) within three years in the state.[22]
- The percentage of licensed graduates retained as school leaders (assistant principals and principals) for three or more years after their initial placement in the state.

In-depth review: The in-depth program review provides an opportunity to analyze patterns in output data including: discrepancies between data about program completion and success on licensure exams, analysis of performance and circumstances of graduates who were not hired into principals' positions within three years, and analysis of performance and circumstances of graduates who were not retained in principal positions for three years.

4. Graduate Outcomes: Effective leadership preparation programs produce graduates who improve student outcomes and schools.

Annual Review: Outcomes indicators fall into two categories of data: (1) indicators of the quality of practices of principals and educators in schools they lead and (2) aggregated individual student growth in schools led by program graduates for three years or more. In the first category, we recommend the annual report include direct indicators of leaders' effectiveness (i.e., the percent of graduates rated effective or above based on leadership effectiveness ratings), leading indicators of student achievement that reflect directly on the practice of principals (i.e., teacher and student surveys of school climate in schools led by program completers, and measures of teacher effectiveness). In the second category are non-cognitive student outcomes in schools lead by program completers (e.g., discipline, attendance). If valid and reliable data are available, we recommend that reported data include the percent of graduates who have a positive effect on value-added or other growth measures of student achievement.

In-depth Review: We recommend in-depth review carefully assesses school contexts (e.g., teacher engagement, student engagement, academic rigor, student and staff attendance, working conditions, school climate, parent perceptions), additional academic student outcomes (credit accumulation, promotion, graduation, achievement), and non-cognitive student outcomes (student attendance, student discipline, engagement). In-depth review is also an opportunity for programs and state personnel to explore the linkages between program features and program outcomes and to use the findings from such an exploration for program improvement.

Do states have the capacity to collect, analyze and use these data?

Many states will need to build additional, more consistent and more accurate data systems (especially related to graduates' employment) in order to fully pursue the approach we describe. Given the mobility of leadership candidates, such a database would need capacity to follow candidates within the state, across both public and private sectors.[23]

In addition to ensuring accurate data systems, states will need to develop the capacity to analyze data in a rigorous manner. It is important for state personnel to have the requisite expertise for analyzing and interpreting indicators of program impact in light of contextual factors that influence program and outcome measures.

There are substantial challenges to creating such systems and capacities, not least the need to make substantial financial investments in systems for data reporting and analysis. These challenges should temper expectations, but they should not stop states from considering their readiness and current capacity for improving existing program evaluation practices. We recommend that states engage in a rigorous self-assessment of their capabilities as a starting point (see Section VI for more details).

---

21. This data point should be a compilation of information reported by the program (e.g., the percent of program completers who do and do not merit endorsement for licensure based on assessment of their performance) and information collected by the state (i.e., licensure exam results).
22. Placement into leadership positions may be delayed due to availability of positions that fit with the specific background and expertise of candidates or general job market in a particular region or state. Some graduates might also need one to two more years of practice in teacher leadership positions before they are ready for promotion into a school leadership position.
23. Ideally, an effective database would be able to follow leaders across state lines. National databases exist in other professions and one is warranted in education.

New Leaders

uc ea University Council for Educational Administration
Quality Leadership Matters *since 1954*

Why should states consider using student achievement to evaluate preparation programs when student achievement outcomes are influenced by so many other things?

It is very difficult to disentangle preparation program effects from other school effects on student achievement.[24] For example, school effects might be influenced by teaching quality, school culture, curricula, and/or resources. However, strong program evaluation examines the entire theory of action of a program, from inputs through outcomes.[25] Even if it's not entirely possible to attribute the final outcomes of student achievement to principals, it can be useful to track achievement as an indicator of a desired outcome of preparation programs. Thus, in the context of examining all components of the evaluation pathway, we recommend using student achievement if and when it is used as part of a two-stage process (as described in this document) and when users understand the limitations of using their state's student achievement measures for this purpose.[26]

This approach is similar to how indicators are used in the medical field. Doctors regularly take patients' blood pressure during each visit. An above-average blood pressure reading should not be used for a high-stakes decision like surgery (after all, someone's blood pressure might be high on a particular day because they are feeling anxious about something), but it can be a useful piece of information when used alongside other information (such as symptoms and/ or family history) to decide when additional information gathering is warranted.

States utilize different methodologies to create school-level measures of student achievement. Value-added measures (VAMs) are becoming a more common measure of student achievement because they isolate the school impact on growth in student achievement growth by controlling for other factors, but the actual methodology of VAMs varies from state to state. For the purpose of evaluating principal preparation programs, states should, at minimum, use measures that examine individual growth (that is, control for prior achievement), maximize the number of students included in the measures, and—if possible—take into account when the principals started their tenure at the school.

Due to the difficulties involved in isolating principal effects from other non-observed school and/or district effects, measures of student achievement should not be used on their own for high-stakes decisions about individual principals. However, several prominent research organizations (e.g., American Educational Research Association, National Research Council, and National Academy of Education)[27] recommend using VAMs to evaluate preparation programs,

particularly if the data are used in combination with other indicators.[28] Despite major limitations of student achievement measures, they can be helpful to identify tail ends of a distribution, making them a useful diagnostic measure.

Aggregating data at the school level introduces particular constraints on what student achievement data can validly and reliably communicate about principal effects. Obvious threats to validity include district-level variables, tenure in role, the time period between program completion to placement, school context, and variations in tests used by different districts. Further, student achievement measures cannot control for unobserved variables related to school or district context. Therefore, the results could insert unfair bias into evaluations (e.g., schools whose graduates enter easier settings could appear to have more successful graduates, and which could incentivize programs not to place graduates in more challenging schools and/or district settings). Based on available research, we believe the concerns regarding bias are legitimate. Even though combining student achievement measures across multiple years decreases the standard error of these scores, the model could still introduce bias for particular contexts.

Given these concerns and other cautions in the research, we recommend that VAM and student growth data be used in the following ways:

Student achievement measures included in the annual reports should:
- not be used alone for high-stakes decisions about program status;
- control for prior student achievement, school context and principal characteristics;
- be used only when a sample of at least 10 schools run by program graduates is available to include in the analysis; and
- be used only for principals who have been in place for three or more years.

Analysis of student achievement measures as part of deeper program review should include:
- comparing results to a matched comparison group;
- identifying the trajectories of student achievement measures prior to the employment of the graduate from the principal preparation program;
- investigating root causes for lower-than-average results; and
- encouraging programs to follow up with graduates to gather their own information about factors that are influencing student outcome results.

24. Grissom, J.A., Kalogrides, D., & Loeb, S. (2015). Using student test scores to measure principal performance. Educational Evaluation and Policy Analysis 37(1): 3-28; Fuller, E. J., & Hollingworth, L. (2014). A Bridge Too Far? Challenges in Evaluating Principal Effectiveness. *Educational Administration Quarterly*, 50(3), 466-499.; Chiang, H., Lipscomb, S., & Gill, B. (2014). Is school value-added indicative of principal quality? Princeton, NJ: Mathematica Policy Research. www.mathematica-mpr.com/publications/pdfs/education/value-added_principal_quality.pdf.

25. McDavid, J.C., & Hawthorn, L.R.L. (2006). *Program evaluation & performance measurement: An introduction to practice*. Thousand Oaks, CA: Sage.

26. Burkhauser, S., Pierson, A., Gates, S. M., & Hamilton, L. S. (2012). *Addressing challenges in evaluating school principal improvement efforts*. Santa Monica, CA: RAND Corporation; Council for the Accreditation of Educator Preparation (2015). *CAEP evidence guide*. Washington, DC: Author; Feuer, M. J., Floden, R. E., Chudowsky, N., and Ahn, J. (2013). *Evaluation of teacher preparation programs: Purposes, methods, and policy options*. Washington, DC: National Academy of Education.

27. National Research Council and National Academy of Education (2010). *Getting value out of value-added: Report of a workshop*. Washington, DC: The National Academies Press; American Educational Research Association (2014, September 11). *AERA-NAED hold successful briefing on teacher evaluation*. Retrieved Jun 12, 2015 from http://www.aera.net/EducationResearch/ResearchandthePublicGood/PublicBriefings/AERANAEDHoldSuccessfulBriefingonTeacherEval/tabid/10902/Default.aspx.

28. The American Statistical Association (2014) states that it can be appropriate to use VAMs "to evaluate effects of policies or teacher training programs by comparing the average VAM scores of teachers from different programs. In these uses, the VAM scores partially adjust for the differing backgrounds of the students, and averaging the results over different teachers improves the stability of the estimates" (p.7). American Statistical Association. (2014). *Using value-added models for educational assessment*. Alexandria, VA: Author.

E. Summative Evaluation

States are responsible for approving principal preparation programs and renewing that approval based on periodic review.[29] In order to make consequential decisions about programs—especially decisions to revoke approval for under-performing programs, but also decisions to share practices from one program as models for other programs—states need to first make sound, data-driven summative judgments about programs. Then they need clear processes and decision rules for acting on those judgments. We refer to this responsibility as summative evaluation.

State contexts and approval authorities differ significantly, thus a single model of summative evaluation is difficult to construct. Instead, we offer these recommendations related to summative evaluation:

Arrive at summative judgments only after annual reporting AND on-site review. Alone, annual report data are insufficient to inform summative judgments about programs. As noted above, the two-stage process provides states with ample opportunity to confirm (or disconfirm) data conclusions and arrive at defensible, fair judgments.

Create enough levels of performance to make the distinctions meaningful. We suggest four levels of performance as adequate differentiation and as a sound basis for taking actions relevant to any program's level of performance. Four levels can be defined as:

- Highly effective—Those programs where data on many indicators related to inputs, processes, outputs and graduate outcomes exceed targets or expectations set by the state and where in-depth review does not result in areas of serious concern (e.g., identified as ineffective on our suggested rubric). We recommend that reports for highly effective programs be augmented to capture a comprehensive description of the practices that are believed to be contributing to success. If resources are available, we also recommend that the state invest in creating case studies showcasing the practices of these programs and their process for ongoing improvement or convene gatherings of preparation programs to highlight effective practices for the purpose of continuous improvement.

- Effective programs—Those programs where data on indicators related to inputs, processes, outputs and graduate outcomes consistently meets targets or expectations set by the state and where in-depth review does not result in multiple areas of serious concern (e.g., identified as ineffective on our suggested rubric). Since this rating represents the expected level of performance for programs in the state, an effective rating triggers renewal. Further, we recommend the review team's report be augmented to capture discrete practices that are believed to contribute to success and can be shared with other programs.

- Needs Improvement—Those programs where data on indicators related to inputs, processes, outputs and graduate outcomes demonstrate one or more areas of ineffective practice. If data on the annual report suggests that a program is at this level, targeted review is a likely step. If the program does not remedy the areas of concern in response to the targeted review, then the state may suspend program approval until deficiencies are remedied.

- Ineffective—Those programs where data on indicators related to inputs, processes, outputs and graduate outcomes demonstrate numerous areas of ineffective practice. If data on the annual report suggests that a program is at this level, targeted review is a certainty. These programs may have their status suspended until deficiencies are remedied. If the state deems the problems to be so severe that improvement is unlikely, then the program status is revoked.

---

29. This guide does not include specific recommendations about processes or criteria for the initial approval of new programs, which is another opportunity to hold a high quality bar for preparation. For guidance on that, see *Change Agents*. New Leaders (2013). *Change agents: How states can develop effective school leaders*. New York, NY: New Leaders http://www.newleaders.org/newsreports/publications/change-agents/.)

Be clear on how the component parts of program evaluation contribute to a summative judgment. The rubrics in Tool C provide reviewers with a template for assigning ratings for each major category of program evaluation—inputs, processes, outputs, and graduate outcomes—based on data from the annual report and further review.

The design principles call for all four categories to contribute to a summative judgment based on annual reporting and in-depth review. However, states will need to determine the relative weight of the indicators measured. As states develop more sophisticated systems for tracking the placement of graduates into principal roles and the impact that they have on student learning and other important outcomes, these measures should take a more prominent role in summative judgments about preparation programs. Until such systems are in place, we recommend that states maintain higher weights on those indicators for which they have valid and reliable data.

Summative judgments might need to be adjusted based on **targeted** review. Assume, for example, that data compiled within the annual report for a program raise concerns about program processes (i.e., there appear to be limited opportunities for aspiring leaders to practice leadership activities in a practicum or residency) and program outputs (i.e., the program has a low rate of graduates being placed in roles as principals and assistant principals); assume also that the targeted review confirms concerns raised by the data. In this case, the program's summative rating should be adjusted downward. However, if the program improves in response to an action plan related to the areas of concern, we recommend that ratings be adjusted upwards to reflect that improvement.

New Leaders

IV. STATE CONDITIONS

Implementing a new, better system of evaluating principal preparation programs is complex work. For many states, this work will be new and difficult and will require significant commitment of time, expertise and resources.

We recommend that states start this work by diagnosing a set of existing conditions and proceed based on what that diagnosis reveals. Three considerations are particularly important:

1. The nature and extent of **authority** for state agencies to approve programs.

2. The extent to which leadership and leadership preparation are **priorities** in the state educational improvement agenda.

3. The extent to which the SEA possesses crucial **technical capabilities**, especially related to data collection and analysis, and substantive program review.

Resource A provides an overview of how preparation programs are currently reviewed by states. This resource can be used to assist in planning as well as in assessing consideration #1 above. Tool A is a Readiness Diagnostic designed to assist in the assessment of conditions #2 and #3 above, helping states distinguish whether conditions are prohibitive, workable, or ideal.

A. State Authority for Principal Preparation

States first need to be clear on where authority for approving principal preparation programs resides and how that authority is currently employed. Many states do have the authority to approve programs, but the nature of that authority and the state entity that possesses the authority can vary. In some cases, authority resides with the state board of education, with functional responsibility residing with the state education agency. In other cases, the state board shares authority with a professional standards board or another public entity.

States also vary in the manner in which they exercise their authority. Twenty-seven states and Washington, D.C. require initial program oversight with documentation and site visits. Twenty-three states require ongoing state reviews at specified intervals, so that reviews are not limited to initial program approval.[30]

Further, the types of data required for reviews vary. Most states that review programs focus their reviews on program standards and processes,[31] and most collect information on the number of preparation program graduates. But few collect information on the performance of graduates or the programs from which they graduate. Nor do most states have accurate longitudinal data on principal job placement, retention rates by principal preparation programs, principal job effectiveness (as measured by performance evaluation ratings) by principal preparation program, or principal job effectiveness (as measured by school outcomes or student achievement) by principal preparation programs.[32]

B. Leadership as a State Priority

The next consideration for states is the extent to which principal leadership, and more particularly principal preparation, is a priority for state leaders. Within education, state leaders may be focused on a wide array of issues, ranging from the adequacy of state funding, to the content of student learning standards and assessments, to the quality of teacher evaluation systems. State leaders (i.e., governors, legislators, state board members, chief state school officers, associations and others) also may differ on the relative priority of issues, let alone particular solutions to those issues.

---

30. Anderson, E., & Reynolds, A. (accepted). State of state policies for principal preparation program approval and candidate licensure. Journal of Research in Educational Administration.

31. Ibid.

32. Briggs, K., Rhines Cheney, G., Davis, J., & Moll, K. (2013). Operating in the dark: What outdated state policies and data gaps mean for effective school leadership. Dallas, TX: The George W. Bush Institute.

New Leaders

University Council for Educational Administration
Quality Leadership Matters   since 1954

We recommend that state leaders consider whether there is agreement—including among key stakeholders such as deans of colleges of education and principals' associations—that principal leadership is among the state's major priorities. Ideally, these leaders and stakeholders share an understanding of how improved principal leadership will contribute to improved educational outcomes. If state political leaders and relevant agencies have a shared commitment to improving principal preparation programs, there are a number of steps that can be taken such as the modification of existing policies that support program review processes and the targeting of funds to support implementation of the guidance offered in this document. And finally, meaningful improvement is more likely if the state education agency is seen as a supportive partner interested in improvement and innovation, not just compliance.

C. Technical Capabilities

The third consideration for states is the extent to which they have, or can build, a set of technical capabilities crucial to effective design and implementation of a system of program review. These capabilities include:

1. **Data and data system requirements.** In order to implement a system that meets the design criteria described in this guide, a state needs a robust annual data system that includes important data and enables tracking over time. The system we recommend would include data on student performance and data on individual educators, including their role (e.g., teacher, principal, assistant principal, other school leader, district leader), preparation program, licensure status, and effectiveness ratings from an educator evaluation system. Ideally, the data system also includes information such as school-level data that can be connected back to the leader and her/his preparation program (e.g., student demographics, attendance rates, graduation rates, and achievement data, as well as other indicators).

   Without these capabilities, we recommend a state start small. For example, states could limit the evaluation to available data (e.g., program input data collected and submitted by programs) and fund and incentivize programs to collect and report on their own output and impact data. However, data collected in this way should be interpreted with caution and not made public due to concerns about verification. Meanwhile, the state could invest in building a more robust data system.

2. **Data compilation and analysis capacity.** In order to implement a system that meets the design criteria described in this guide, a state would need substantial capacity to compile, clean and analyze data. This capacity is both a resource consideration, in that the state needs to fund the analytical capability, as well as an expertise consideration. Ideally, those conducting the analyses have experience in preparation program evaluation work, particularly with respect to principal preparation program evaluation.

   If the state does not have these resources, it might consider seeking partnerships with research institutions or consortia with data analysis capabilities.

3. **Review process capacity.** In order to implement a system that meets the design criteria described in this guide, a state would likely need to invest resources in program review. In particular, the state may need to train and maintain a cadre of reviewers for in-depth reviews.

   If resources are limited, the state might consider limiting the number of programs requiring in-depth review on an annual basis and/or partnering with approved professional associations to conduct the in-depth reviews.

Each of these capabilities is outlined in further detail in Tool A: State Readiness Diagnostic Rubric.

V. TOOLS AND RESOURCES

Here we describe the supplementary tools and resources designed to support effective implementation of the approach described in this guide. The tools should be adapted to meet local context needs.

Tool/Resource	Purpose(s)	Description
Tool A: State Readiness Diagnostic Rubric	Enable states to assess the extent to which they are ready to implement the recommendations provided in the guide.	Rubric detailing when conditions for this work are prohibitive, workable, and ideal. Conditions addressed include the focus alignment and positioning of state leadership, as well as technical capabilities of the state education agency
Tool B: Program Indicators, Rubric and Report	Provide states with specifications for an annual report that would give states consistent information to consider during in-depth review and decide when to initiate a targeted review. Also provides candidates and districts basic data on programs.	Suggested design for an online platform with data and information for each preparation program in the state including: the status of the program and when it is due for review, and annual data points for multiple measures in each area (inputs, processes, outputs and program graduate outcomes).
Tool C: Handbook for the In-Depth Review Process	Provides states and programs with clear processes and tools to support effective enactment of periodic in-depth reviews of programs.	Detailed process guide for undertaking reviews, including sections on: data, process, the review team, and rubrics to assess programs.
Tool D: Handbook for the Targeted Review Process	Provides states and programs with clear processes and tools to support effective enactment of targeted reviews in response to concerning data.	Detailed process guide for undertaking reviews, including sections on: purposes, measures, process, and reviewer credentialing.
Resource A: Overview of How Preparation Programs are Currently Reviewed by States	Provide states with information to compare their current principal preparation program evaluation system to recommendations proposed in this guide.	State-by-state summary tables of what each state requires for program approval and oversight, including the data that each state requires and an analysis of gaps between what data are currently collected versus data recommended for annual review.
Resource B: Description of Other In-Depth Program Review Processes	Provides states with background on the four types of reviews that higher education leadership preparation programs are likely to experience so they can design their system with alignment to these systems in mind.	Narrative description of what each type of review entails, how often they typically occur, and common challenges. The final section describes factors preparation providers consider to be beneficial sources of change.
Resource C: List of Other Tools and Resources	Provides states with links and reference citations to tools, research, and resources created beyond this project.	Bibliographic list of tools and resources, including descriptions and directions for accessing them.

SEP³ Toolkit | STATE EVALUATION OF PRINCIPAL PREPARATION PROGRAMS

Tool A:
Readiness Assessment Rubric and Process Recommendations

I. INTRODUCTION

Implementing a better system of evaluating principal preparation programs is complex work, and it requires that states have certain conditions and capabilities already in place. Before undertaking the work of designing and implementing a new evaluation system, we recommend that states assess their capacity to implement the recommendations in the guide. This tool is designed to help with that assessment of readiness. It has two parts: (1) a readiness assessment rubric and (2) process recommendations for completing the assessment.

The readiness assessment rubric includes information in two broad areas:

A. **Focus, alignment, and positioning of state leadership:** The extent to which state leadership has prioritized school leadership—and specifically school leader preparation—in the state's educational improvement agenda, and the extent to which the state education agency (SEA) is positioned to be an effective resource for local education agencies and leadership preparation programs.

B. **Technical capabilities of the state education agency:** The extent to which the SEA has crucial capabilities needed to support a new evaluation system, especially those related to data collection and the analysis and substantive review of programs.

The process recommendations outline how states might use information in these two areas to arrive at conclusions about their readiness to restructure or refine their assessment of leadership preparation programs. Completing this rubric will enable states to determine whether current conditions are ideal, workable or underdeveloped for implementing the recommendations in the guide.

When conditions are ideal, states may move forward with confidence. When conditions are workable in most areas, states may decide to move forward and work on improving conditions at the same time. When conditions are underdeveloped, states would benefit from developing supportive conditions before adopting our relevant recommendations. To move forward when conditions are underdeveloped would invite low-quality implementation and could unintentionally result in poor and potentially negative outcomes.

It is important to note that this is not a scientifically validated instrument. They do not lend themselves well to absolute determinations. Rather, it is a heuristic, allowing states to make sensible judgments about where to start and how fast to proceed.

New Leaders

uc ea University Council for Educational Administration
Quality Leadership Matters   since 1954

II. READINESS ASSESSMENT RUBRIC

A. Focus, Alignment, and Positioning of State Leadership

	Underdeveloped	Workable	Ideal
A1. Commitment to improving school leadership			
State leadership prioritization: Public commitment by state leaders and key stakeholders to improving school leadership	State political leaders (governor, state chief, state board, legislative leaders) rarely discuss school leadership as a way to improve schools. Stakeholders (e.g., associations, prominent local education agency [LEA] leaders, university leaders) have major disagreements on the importance of leadership.	State political leaders communicate about school leaders as one among many issues of concern. Stakeholders have a broad array of perspectives on the importance of leadership.	School leadership is a top agenda item for state political leaders. Stakeholders' perspectives are highly aligned around the importance of school leadership.
Investment in leadership: Visibility of school leadership in state strategic plan and in allocation of resources	The state's strategic plan says very little about strategies to improve school leadership. No discretionary dollars are allocated to improving school leadership, and no effort is made to encourage local investments in leadership.	School leadership is in the strategic plan but is a secondary priority or one on a long list of stated priorities. Investments in leadership are regularly communicated as allowable expenditures in state grant programs (as appropriate).	School leadership is a major focus of the state's strategic plan; the state has a clear understanding of how improved school leadership will contribute to improved educational outcomes. The state has targeted funds (including public and privately sourced funds) to specific efforts to improve school leadership.
A2. Commitment to improving principal preparation			
State leadership alignment: Unified stance of state leaders to improve principal preparation	Lines of authority for improving principal preparation programs are divided or ambiguous. Agencies with responsibility for principal preparation and licensure (e.g., SEA, professional licensing boards) have limited communication and differing priorities.	Lines of authority for improving principal preparation programs are clear. Agencies with responsibility for principal preparation and licensure have goals and strategies that do not conflict, and they communicate regularly.	State political leaders and relevant agencies are unified in a commitment to improving principal preparation programs and agree on the need to rigorously assess the quality of programs, help programs improve, and take action to address underperformance. Agencies with responsibility for principal preparation and licensure have shared goals and are committed to collaboration with each other and with programs (especially when it comes to sharing data).
Policy framework: Policies in place to foster innovation	There is little effort by state political leaders to influence the practices of principal preparation programs.	The state's policy framework allows for innovative principal preparation program design.	The state pursues new partners and encourages universities and other providers to create innovative principal preparation programs.

	Underdeveloped	Workable	Ideal
A3. Commitment to and capacity for continuous improvement[1]			
Collaboration: Perceptions of the SEA as a collaborative partner	LEAs and preparation programs have little interaction with the SEA, viewing the agency as primarily concerned about compliance with statutes and regulations.	LEAs and preparation programs have mixed interaction with the SEA: Some interactions are overly driven by compliance concerns, while other interactions are focused on genuine improvement and making rules work for them.	LEAs and preparation programs view the SEA as a trusted partner committed to continuous improvement; compliance still matters, but the SEA works to make it as seamless as possible.
Communication: SEA systems for communication with partners	Information coming from the SEA to LEAs and preparation programs is either nonexistent or perceived by programs as excessive and disjointed, often sending mixed messages.	Information coming from the SEA to LEAs and preparation programs is perceived by programs as organized, predictable and reasonably clear.	The SEA convenes local partners in ways that foster two-way communication.
Innovation: Perceptions of the SEA as a source of ideas	LEAs and preparation programs do not look to the SEA for new ideas to improve schools and universities.	The SEA serves as an effective information clearinghouse, making innovations in the field visible to LEAs and preparation programs.	The SEA shares data; engages LEAs and preparation programs in conversations about improvement; and offers new learning opportunities, including creative strategies for implementing federal and state policy.
Decision making: Use of evidence in SEA decisions	The SEA offers little explanation or unclear justification for policy changes.	The SEA reports on data used in the design of new policies and articulates the reasons for policy changes.	The SEA transparently shares data, data analysis, and operating theories that underlie policy design and implementation decisions.
Expertise: Knowledge and skills to manage change process for leadership work	LEAs and preparation programs view SEA leadership as having limited understanding of core leadership issues and as being unresponsive or unhelpful in managing the process of large-scale change.	SEA leadership communicates a solid understanding of the connections between leadership and student outcomes, as well as the adaptive challenges associated with large-scale change.	SEA leadership communicates a strong understanding of—and solutions for—the adaptive challenges associated with large-scale change. SEA leadership is deeply involved in national and statewide conversations about the practice and impact of school leaders.

---

1. Determined through anonymous surveys of program leaders.

New Leaders

B. Technical Capabilities of the State Education Agency

	Underdeveloped	Workable	Ideal
B1. Data and data system requirements			
Program data system: System that collects program data (e.g., number of applicants, clinical hours required, 100-word description) from preparation programs	Data are available in isolated locations without an overarching system for integrating the different sources or linking the data points.	A program data system is in place, but it may not include all data points needed for the SEA's annual report; some data may be missing, inaccurate, or lack comparability. Systems support might be needed to design new tools or interfaces to collect needed information from multiple sources and/or agencies. Substantial budgeting would be required for staff time to request missing data, monitor data completion, and build necessary data systems. Time is allocated to ensure data integrity.	A program data system is in place and includes all fields/variables needed for the SEA's annual report. The system enables consistent reporting and data aggregation. Data are complete and accurate. Programs use common definitions of indicators, making the data comparable across programs. The system is not overly burdensome for programs, districts, or school partners.
Placement data systems: Systems that track individual educators and their annual placement role (teacher, principal, assistant principal, other school leader, district leader)	Data are available in isolated locations without an overarching system for integrating the different sources or linking the data points.	Placement data systems exist and are coordinated but have lots of inaccuracies and missing data. Budgeting would be required for staff time to request missing data and monitor data. Time is allocated to clean data.	Placement data systems are complete and accurate.
Unique identifiers for program participants: Identifiers that link data from preparation programs, licensure status, placement data systems, and effectiveness ratings from educator evaluation system	It is not possible to link individuals across data systems (for programs, licensure, placements, school outcomes).	Unique identifiers do not exist, but it is possible to link two or more data systems, and the SEA has the capacity to do this. Budgeting would be required for junior analyst time to link data systems.	Unique state-level identifiers are in place to link individuals to all of the data required by the evaluation system.
Comparable survey data: Common survey administered to program graduates that gathers their perceptions of program process indicators	Graduates of most programs are not surveyed, or the response rates are too low to make results meaningful.	Surveys of program graduates exist, and response rates are reasonable, but the surveys differ, preventing comparison of data across programs.	A common survey is administered to all graduates in the state with reasonable response rates, enabling comparison of data across programs.
Measures of teacher and leader effectiveness: Ratings of individual teachers and principals on the state performance evaluation system	Measures do not exist or do not have any variability.	Measures exist and have some variability but lack validity and reliability. The SEA does not share results publicly and does not provide caveats that caution users on interpretation.	Measures exist, have variability, and have been found to be both reliable and valid. The SEA has the capacity to use measures in contextually appropriate ways. The SEA ensures that any public release of data meets federal and state privacy guidelines.
Measures of student learning gains: Student achievement scores across grade levels in core subject areas	Measures are not based on individual student growth from year to year.	Consistent and methodologically sound measures of individual student growth, including proper controls for student- and school-level variables, exist, but they are not comprehensive across grade levels and subject areas. Measures exist, but ns are small (less than 10 individuals) for most programs.	Consistent and methodologically sound measures of individual student growth, including proper controls for student- and school-level variables, exist. These measures allow for assessment of school leaders' influence on student learning after three years at a school site. Adequate consideration is given to bias against high-needs schools.

	Underdeveloped	**Workable**	**Ideal**
B2. Data compilation and analysis capacity			
Monitoring data reporting completion and accuracy: Requires staffing to ensure the submission and accuracy of data from preparation programs and other data sources	No staff or resources exist.	Staff assignments and/or resources could be prioritized for data monitoring.	Staff and/or resources are already assigned to data monitoring.
Creating and publishing annual reports: Requires technical skill for website/report design and senior analytical skill to make methodological decisions	No staff or resources exist.	Staff assignments and/or resources could be prioritized for data reporting.	Staff and/or resources are already assigned to data reporting.
Creating and implementing methodology for summative rating: Requires specialized assessment and statistical skill	No staff or resources exist.	Staff assignments and/or resources could be prioritized for data analysis/methodology.	Staff and/or resources are already assigned to data analysis/methodology.
B3. Review process capabilities			
Staffing: Requires specialized leadership experience and skills	There is no SEA staff committed to leadership preparation, or those responsible have multiple other roles.	There are staff members at the SEA focused on school leadership, including preparation, but they have limited backgrounds in school leadership or adult leadership.	There are staff members at the SEA focused on school leadership, including preparation, and they are deeply credible with leaders and preparation providers in the state.
Management and training of reviewers: Requires specialized review process capabilities	No staff or resources exist.	The state has a reasonably adequate pool of high-quality, credible reviewers but does not have a track record of systematically vetting them for leadership expertise or training them for inter-rater reliability. The state does not have a strong track record of outsourcing functions and maintaining quality.	The state has a robust pool of high- quality, credible reviewers who have been (or could be) trained for inter-rater reliability and normed to provide useful feedback to programs. or The state has a strong track record of outsourcing functions and maintaining quality. This allows for bringing in an established process (e.g., review by the Educational Leadership Constituent Council).
Implementation of reviews: Requires financial and human resources	No staff or resources exist.	A review process exists, but it is not sufficient for quality, in-depth review of all programs flagged.	Sufficient resources exist to carry out in-depth reviews for all programs flagged as needing it, and for conducting periodic reviews of all programs.

III. PROCESS RECOMMENDATIONS

The readiness assessment rubric can be used in more than one way. A state working to build political support for an evidence-based approach to assessing the quality of principal preparation programs may want a formal process to engage stakeholders in completing the rubric and agreeing on next steps for the work. Meanwhile, a state already committed to an evidence-based approach to assessing the quality of principal preparation programs may want the SEA to simply undertake an internal diagnosis of conditions in order to surface critical gaps and needed resources.

For a more extensive process, we recommend these general steps:

1. **Create a vision for the work.** In order to demonstrate executive-level commitment to an open and honest process of assessing the state's readiness for implementing a better system of evaluating principal preparation programs, it can be helpful to write a purpose statement outlining why the work is important and how it connects to the state's broader vision of leadership. The state's strategic plan for education is an important resource for this step.

2. **Create a project plan.** In order to ensure that the right people will be engaged and will have access to authentic information, it can be helpful to craft a project plan that includes roles and responsibilities and to assemble available data to conduct the readiness assessment.

3. **Convene stakeholders.** In order to build trust in and commitment to the process, it can be helpful to convene leaders from universities, preparation programs, administrator associations, districts and schools. The purpose of such a convening is to share the goals and work plan, ask for authentic feedback, and ask for a commitment to participating in the process.

4. **Conduct the assessment**. This is the heart of the work: gathering data, making sense of it, surfacing and discussing important substantive issues and agreeing on rubric ratings.

5. **Set action steps.** With the assessment complete, state leaders and stakeholders need to make decisions about their readiness and identify areas of focus that are consistent with the conclusions from the readiness assessment. This is also an opportunity to establish a new work plan for the implementation phase, including strategies for addressing any areas of weakness that need to be remedied in the short term.

For a more targeted approach within an SEA, the critical steps are numbers 4 and 5 above, as well as some amount of stakeholder engagement (step 3). Note, however, that some categories of the rubric require information from sources outside of the SEA (e.g., perceptions held by LEA leaders and program leaders), so some level of external engagement is helpful regardless of the scope of the analysis.

Tool B:
Annual Report Indicators, Reporting, and Interpretation of Results

I. INTRODUCTION

As described in the guide, annual reports ideally provide data on a basic and comparable set of indicators. This tool does three things related to annual reports:

1. It recommends a set of indicators for inclusion in an annual report. For each indicator, we provide a description, purpose and rationale, and methodological considerations.

2. It mocks up an example annual report.

3. It highlights important considerations in interpreting results, including that:

   a. data is imperfect, and therefore indicators should not be used in isolation for high-stakes decisions;

   b. the evaluation approach assumes relationships between indicators and preparation program quality;

   c. annual reports are important to use as part of a two-stage process, where results are used to determine the need for in-depth reviews of program quality;

   d. baseline data, resources, policy context, and stakeholder input are important for target setting; and

   e. opportunities for programs to explain results can help allay concerns related to public reporting.

II. RECOMMENDED INDICATORS FOR ANNUAL REPORTS

The following chart defines the recommended indicators and provides a rationale for each, as well as information regarding how the indicators might (or might not) reflect program quality. The proposed indicators adhere to the guide's design principles for annual report data collection and use—measuring what matters most, using indicators that can be measured accurately and consistently across programs, and using indicators that can realistically be gathered and reported on an annual basis.

The indicators reflect a research-based theory of action regarding the program characteristics that improve leadership practice, and how leadership practice then influences school conditions and practices that ultimately improve student outcomes.[1] This framework is described in detail in a 2013 publication by the University Council for Educational Administration's Center for the Evaluation of Educational Leadership Preparation and Practice.[2] We further narrowed the indicators based on assumptions regarding which data could be gathered most accurately, consistently, and feasibly across programs. The advisory group members provided significant input regarding which indicators were most important and feasible to include on the annual report. We include four categories of indicators: inputs, processes, outputs, and graduate outcomes. For the process indicators, graduate survey perception data are the most feasible type of data to use for comparing programs. In order to enable cross-program comparisons of survey results, we encourage states to select a survey and administer it statewide or coordinate its administration by all preparation programs. We recommend that states use the INSPIRE survey suite because these surveys are aligned with research on effective leadership preparation, have been validated, and can provide nationally normed comparisons (see Resource C for additional information).

Many of the indicators have advantages and disadvantages that states should consider when setting targets and interpreting results. These methodological considerations are described in the chart below. States might also want to consult methodologists to understand the pros and cons of their states' data and indicators before deciding how to use, define, and interpret them. Furthermore, we would recommend that states examine the usefulness of their data on an ongoing basis by (a) tracking indicators over time, (b) disaggregating results by program graduate subgroups, and (c) making trend data publicly available after a period of three years.

1. Clifford, M., Behrstock-Sherratt, E., & Fetters, J. (2012). *The ripple effect: A synthesis of research on principal influence to inform performance evaluation design*. Washington, DC: American Institute for Research; Orr, M. T., Young, M. D., & Rorrer, A. K. (2013). *Developing evaluation evidence: A formative and summative evaluation planner for educational leadership preparation programs*. Charlottesville, VA: National Center for the Evaluation of Educational Leadership Preparation and Practice; Neuman-Sheldon, B., Ikemoto, G. S., Bailey, M., Erdfarb, T., Nerenberg, L., Patterson, N., & Valdez, M. (2014). *Principal preparation program self-evaluation: Lessons learned by New Leaders*. New York, NY: New Leaders.
2. Orr, M. T, Young, M. D., & Rorrer, A. K. (2013).

New Leaders

uc ea University Council for Educational Administration
Quality Leadership Matters since 1954

Inputs			
Indicator	**Description**	**Purpose and Rationale**	**Methodological Considerations**
Admission Rate	Percentage of applicants who were admitted.	**This indicator attempts to capture the degree to which the program is selective.** High-quality leadership preparation programs have a rigorous selection process.[3] They define the competencies and dispositions required for admission and rigorously screen for these skills, often through performance-based tasks. By setting the bar high, institutions select and invest in program participants who are the most likely to complete the program, obtain placement in a principal position, and lead improvements in their assigned schools.[4]	When setting targets, states should consider that low admission rates might indicate extensive recruiting relative to available spots rather than rigorous selection. Conversely, high rates could indicate recruitment efforts that effectively target well-qualified candidates. However, an extremely high admission rate (90%+) could indicate the need for targeted review. A targeted review might validate original concerns, or it might uncover good reasons for a high admission rate that alleviate concerns.
Teaching Experience	Percentage of candidates with at least two years of teaching experience.	**This indicator attempts to capture the degree to which programs are selecting candidates with sufficient knowledge of and experience with teaching.** While the connection between teaching experience and leadership effectiveness has received limited attention in the research, effective principals practice instructional leadership, which necessitates deep instructional expertise, particularly regarding instruction aligned to the rigor and expectations of college- and career-ready standards.[5] Candidates need at least some experience in the classroom to effectively support teachers in improving instruction.	States should consider that years of experience (quantity) is not a measure of the depth of participants' instructional knowledge (quality). Effective programs may have more direct ways to screen for instructional expertise (see the next indicator).

3. Cheney, G. R., Davis, J., Garrett, K., & Holleran, J. (2010). *A new approach to principal preparation: Innovative programs share their practices and lessons learned*. Fort Worth, TX: Rainwater Leadership Alliance; Darling-Hammond, L., Meyerson, D., LaPointe, M., & Orr, M. T. (2010). *Preparing principals for a changing world: Lessons from effective school leadership programs*. San Francisco, CA: Jossey-Bass; Mitgang, L. (2012). *The making of the principal: Five lessons in leadership training*. New York, NY: The Wallace Foundation.
4. Cheney, G. R., Davis, J., Garrett, K., & Holleran, J. (2010).
5. McCarthy, M. M., & Forsyth, P. B. (2009). An historical review of research and development activities pertaining to the preparation of school leaders. In M. D. Young, G. M. Crow, J. Murphy, & R. T. Ogawa (Eds.), *Handbook of research on the education of school leaders* (pp. 86–128). New York, NY: Routledge.

uc ea University Council for Educational Administration
Quality Leadership Matters   since 1954

New Leaders

| Instructional Expertise | Percentage of admitted candidates who are rated effective or above according to teacher evaluation systems. | **This indicator attempts to capture the degree to which programs are selecting candidates with sufficient instructional expertise.** While the previous indicator captures candidates' amount of experience in the classroom, this indicator attempts to capture the quality of their practice as a classroom teacher. In order for principals to provide teachers with effective feedback to improve their instruction and shepherd a school-wide instructional program, principals need instructional expertise and should have demonstrated success as classroom teachers.[6] | The field is currently attempting to build systems to create meaningful measures of teacher effectiveness, but the accuracy and validity of such systems are evolving, and systems may not be available or appropriate for use in all states. For example, many states lack variability in teacher effectiveness ratings. When this is the case, the instructional expertise indicator may not be useful to include on the annual report.

The utility of this indicator will depend on the accuracy and validity of the measures used to assess teaching effectiveness. States should conduct due diligence in understanding the limitations of their teacher effectiveness systems and be wary of effectiveness rating methodologies that bias against teachers of certain types of students or teachers in schools with certain types of characteristics. For example, states that heavily weight student outcome measures may introduce bias against teachers who work in schools that serve high proportions of low-income, minority, English Language Learner, or special education students. Acceptable explanations might be uncovered during the review process if and when a program is targeted for review. |
| Adult Leadership Experience | Percentage of candidates who have experience leading adults (e.g., department chairs or grade-level chairs). | **This indicator attempts to capture the degree to which programs are selecting candidates with prior adult leadership experience.** Strong leadership preparation programs require previous experience leading adults in some capacity.[7] This prior experience signals that candidates have had authentic opportunities to start developing personal leadership skills that they may not have needed as a classroom teacher but that are critical for effectively leading adults. | Programs may not currently collect data on this indicator, but would be able to do so through their admissions process if states ask for this information. The indicator is likely to require definition: What counts as previous experience leading adults? Targeted review might determine that variance can be explained by the focus of the program and the participants targeted for recruitment (for example, if the program targets master teachers who have not led other adults but who want to become principals).

States should consider tracking this data over time in an effort to determine (a) whether prior experience is related to successful leadership and (b) whether leadership experience is equitably distributed (e.g., by race and gender), which will require that programs report data on this indicator by subgroup. |

6. Hitt, D. H., Tucker, P. D., & Young, M. D. (2012). *The professional pipeline for educational leadership: A white paper to inform the work of the National Policy Board for Educational Administration*. Charlottesville, VA: University Council for Educational Administration.
7. Hitt, D. H., Tucker, P. D., & Young, M. D. (2012).

Processes			
Indicator	**Description**[8]	**Purpose and Rationale**	**Methodological Considerations**
Rigor and Relevance	Percentage of graduates who report that the program content was rigorous and relevant. (Example survey item: "The program content emphasized instructional leadership.")	**This indicator attempts to signal the rigor and relevance of coursework in the program.** Strong leadership preparation programs include research-based content that is aligned with professional standards, and they provide opportunities to apply content within authentic settings.[9]	States should consider examining responses on survey items and flagging programs with responses that are relatively negative for targeted review. States may want to generate crosstabs for different types of schools to determine whether perceptions of rigor and relevance are related to school level (elementary or secondary) or school type (charter, public, or private).
Quality of Faculty/Staff	Percentage of graduates who report that the program had high-quality faculty or staff. (Example survey item: "The faculty/instructors were knowledgeable.")[10]	**This indicator attempts to signal the quality of the program's faculty and staff.** Since leadership is best learned through practice and feedback, quality staff not only have content expertise but also have strong interpersonal skills and provide detailed and useful feedback to program participants.[11]	States should consider using survey items that differentiate types of faculty, such as content faculty, clinical faculty, or principal mentors from their residency site.
Peer Interactions	Percentage of graduates who report that the program provided high-quality peer interactions. (Example survey item: "My interactions with fellow students have had a positive influence on my professional growth.")	**This indicator attempts to capture the degree to which programs provide substantive opportunities for peer interaction.** Strong leadership preparation programs provide cohort structures that enable collaboration, teamwork, and mutual support.[12] When programs provide opportunities for collaborative learning relationships, they can rely on shared experiences and knowledge to solve real-life problems. These structures also build skills related to working in teams and leveraging networks for solutions.	Programs vary in the extent to which they emphasize the cohort experience, which will likely create variability in the survey responses. States should be particularly concerned if large percentages of students disagree that the program provided high-quality peer interactions.

8. We recommend using a common participant survey to gather data for process indicators in order to enable comparison across programs. While surveys cannot measure the indicators directly, a normative comparison of survey results can be helpful in identifying areas for in-depth review. We encourage states to administer surveys immediately following graduation to obtain the highest response rates, and therefore the most accurate information for normative comparison.

9. Darling-Hammond, L., LaPointe, M., Meyerson, D., Orr, M. T., & Cohen, C. (2007). *Preparing school leaders for a changing world: Lessons from exemplary leadership development programs*. Stanford, CA: Stanford University, Stanford Educational Leadership Institute.

10. Ibid.; Darling-Hammond, L., Meyerson, D., LaPointe, M., & Orr, M. T. (2010).

11. Cheney, G. R., Davis, J., Garrett, K., & Holleran, J. (2010).

12. Darling-Hammond, L., LaPointe, M., Meyerson, D., Orr, M. T., & Cohen, C. (2007).

Performance-Based Assessments	Percentage of graduates who report that the program used performance-based assessments. (Example survey item: "My skills and knowledge were assessed based on application in real-life settings.")	**This indicator attempts to provide insight into how programs assess candidate learning and growth.** Performance-based measures are considered to be the most effective in evaluating candidate knowledge and skills.[13] Performance-based assessments are particularly useful for determining whether candidates can apply knowledge to successfully lead adults in changing school and teaching practices.	States should note that programs typically use their own assessments, which vary in quality and validity as measures of participants' knowledge and skills. Nonetheless, it would be worth finding out more about use of performance assessments within programs, the types of performance assessments used, and how they are used to support candidate learning. If a program has relatively small percentages of students reporting that their program used performance-based assessments, this could signal an area for improvement.
Residency/ Internship Hours	Number of required residency/internship hours.	**This indicator attempts to capture the degree to which candidates have an opportunity to apply their knowledge and skills in a supervised clinical setting.** High-quality programs have field-based internships that enable candidates to apply leadership knowledge and skills under the guidance of an expert practitioner.[14] Current national educational leadership preparation standards[15] require a sustained internship experience, defined as a six-month, concentrated (at least 9–12 hours per week) internship that includes authentic field opportunities to practice adult leadership within a school-based environment.[16]	When setting targets, states should consider that the number of hours is not an indicator of internship quality; however, this data point provides important baseline information about the extent to which the internship exists. States may also wish to gather data on (a) candidates' perceptions of the quality of the internship experience and (b) signature experiences embedded in the internship.

Outputs

Indicator	Description	Purpose and Rationale	Methodological Considerations
Graduation Rate	Percentage of program entrants who complete the program and are recommended by the program for certification or licensure.	**This indicator attempts to capture the flow of candidates from admission to graduation.** Graduation rates are particularly useful to prospective applicants and partner districts that are interested in knowing how financial and time investments are likely to pay off in terms of an advanced license or degree.	When setting targets, states should consider natural attrition (which can be affected by program length) and the extent to which programs counsel participants out of the program. In some cases, participants might meet all of the course requirements for program completion, but their performance on program assessments may not warrant recommendation for certification/endorsement or licensure. Therefore, low graduation rates could indicate a rigorous program with a high bar for graduation, or they could reflect poor program quality. Conversely, extremely high graduation rates could indicate a low bar for graduation. In-depth review can provide additional information to support the interpretation of graduation rates. States might consider requiring programs to document the reasons for individuals leaving the program, which could provide valuable information that would help investigate root causes during in-depth review.

13. National Policy Board for Education Administration. (2011). *Educational program leadership standards: 2011 ELCC building level*. Alexandria, VA: Author.
14. Darling-Hammond, L., LaPointe, M., Meyerson, D., Orr, M. T., & Cohen, C. (2007).
15. National Policy Board for Education Administration. (2011).
16. Young, M. D., & Mawhinney, H. B. (Eds.). (2012). *The research base supporting the ELCC standards: Grounding leadership preparation and the Educational Leadership Constituent Council standards in empirical research*. Charlottesville, VA: University Council for Educational Administration.

New Leaders

uc ea University Council for Educational Administration
Quality Leadership Matters since 1954

Licensure Rate	Percentage of program completers who become licensed by the state.	**This indicator attempts to convey the percentage of program completers whom the programs and state agree are ready and willing to serve in school leadership positions.** Although program completers might generally take and pass the state licensure exam, this indicator could help to identify programs with large percentages of completers who either do not take or do not pass the licensure exam.	Targets should be set in light of state licensure exams and cut scores. Furthermore, we encourage states to be mindful of research demonstrating lower performance on licensure exams by minority candidates.
Placement Rate	Percentage of graduates (licensed/endorsed) hired as school leaders (assistant principals and principals) within three years in the state.[17]	**This indicator attempts to signal both program graduates' desire to take on leadership roles and schools' and/or districts' assessments of program graduate quality.** Placement rates are particularly useful to prospective applicants who want to know whether their financial and time investments in the program would help them to obtain a principal position. Placement rates are also useful to partner districts that invest in the partnership to fill principal vacancies.	Targets should take into account that placement into leadership positions may be delayed due to availability of positions that match the specific background and expertise of candidates, or because of the job market in a particular region or state. States might consider the programs' theories of action regarding whether they expect participants to be ready for assistant principal or principal positions immediately upon graduation. Some programs might be designed with the expectation that graduates will need one to two more years of practice in teacher leadership positions before they are ready for promotion into a school leadership position.

For programs that do not expect participants to be ready to assume assistant principal or principal positions upon graduation, the state may need to consider implications for determining which competencies must be demonstrated before earning an administrative license. Meanwhile, if programs are designed to simply further the education of classroom teachers who do not intend to become school leaders, states might want to explore why this is the case and whether the program should be reclassified as a teacher leadership program. If a relatively large percentage of a program's graduates take longer than three years to obtain positions, the state may want to conduct a targeted review to investigate the underlying causes for the low placement rate. When setting targets, we encourage states to consider the extent to which programs are likely to serve multiple states, and to practice flexibility in setting or maintaining targets when occurrences such as recessions or reductions in K-12 funding impact placement rates. |

---

17. This requires the state to (1) collect job placement data, specifically on the role and school where each individual graduate is placed within the state each year, and (2) calculate placement rates.

| Retention Rate | Percentage of cumulative graduates (licensed/ endorsed) retained as school leaders (assistant principals and principals) for three or more years in the state. | **This indicator attempts to capture the effectiveness of the program through the rate at which program graduates are retained within educational leadership positions.** Recent research on graduates of teacher preparation programs suggests that better prepared teachers remain in the profession for longer. While research has not documented this relationship for school leaders, it seems reasonable to expect that this finding would hold true for school leaders as well.[18] Similar to graduation and placement rates, retention rates can be useful to prospective candidates and partner districts interested in knowing whether an investment will pay off in the long term. | States should note that retention rates are often affected by a number of factors, including district needs, school characteristics, working conditions, salary, promotions, quality of district leadership, and local labor market characteristics.[19] States might want to explore these factors during in-depth review. When setting targets and interpreting results, states should also take into account the extent to which programs are likely to serve multiple states, and they should decide whether to include various types of leadership positions in the indicator definition. |

Graduate Outcomes

Graduate and School Practices

Indicator	Description	Purpose and Rationale	Methodological Considerations
Leadership Effectiveness Rate	Percentage of graduates rated effective or above based on teacher and/or leadership effectiveness ratings. (An alternative measure could be percentage of graduates with positive evaluations of leadership practice as assessed by 360° surveys.)	**This indicator attempts to capture the quality of graduates' leadership practice.** The primary objective of principal preparation programs is to provide well-selected participants with the knowledge and skills needed to be effective school leaders. This indicator attempts to signal the extent to which programs are meeting that objective.	Researchers continue to debate the ability of leadership evaluation systems to effectively evaluate the effectiveness of leadership practice.[20] Even strong evaluation systems do not isolate the impact of the preparation program on the leader's practice, which is likely also influenced by their district and state context, opportunities for ongoing development and support, and a variety of other variables.[21]

18. Fuller, E. J., & Hollingworth, L. (in press). Questioning the use of outcome measures to evaluate principal preparation programs. *Educational Assessment, Evaluation and Accountability*; School Leaders Network. (2014). CHURN: The high cost of principal turnover. Hinsdale, MA: Author.
19. Feuer, M. J., Floden, R. E., Chudowsky, N., & Ahn, J. (2013). *Evaluation of teacher preparation programs: Purposes, methods, and policy options*. Washington, DC: National Academy of Education; Goldhaber, D., & Cowan, J. (2014). Excavating the teacher pipeline: Teacher preparation programs and teacher attrition. Journal of Teacher Education, 65(5), 449–462.
20. Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2013). School leaders matter: Measuring the impact of effective principals. *Education Next*, 13(1), 63–69; Chiang, H., Lipscomb, S., & Gill, B. (2014). *Is school value-added indicative of principal quality?* (Working Paper 12). Princeton, NJ: Mathematica Policy Research; Coelli, M., & Green, D. A. (2012). A School principals and student outcomes. *Economics of Education Review*, 31(1), 92–109; Fuller, E. J., & Hollingworth, L. (2014). A bridge too far? Challenges in evaluating principal effectiveness. *Educational Administration Quarterly*, 50(3), 466–499; Grissom, J. A., Kalogrides, D., & Loeb, S. (2015). Using student test scores to measure principal performance. Educational Evaluation and Policy Analysis, 37(1), 3–28.
21. Fuller, E. J., & Hollingworth, L. (in press).

New Leaders

uc ea University Council for Educational Administration
Quality Leadership Matters   since 1954

| School Climate | Percentage of graduates (licensed/endorsed) placed as principals with improved measures of school climate after two years of leadership. (Example items from teacher surveys: "The school has a well-developed process for facilitating ongoing school-wide improvement and long-range planning"; "Students are academically engaged in their coursework"; "The principal builds a collaborative environment.") | **This indicator attempts to capture the influence of school leaders on school climate.** Research has shown that a positive school climate is associated with and predictive of improved student outcomes.[22] Further, some research suggests that effective leadership practices are associated with improved perceptions of school climate.[23] | Leadership practice is one of many factors that affect school climate, and preparation programs are one of many factors that affect leadership practice. Measures of positive school climate for graduates of one program might be lower than those for other programs because the program targets recruitment and placement in challenging schools. Even though preparation programs only indirectly impact school climate, this indicator could still help states determine whether to conduct an in-depth review. |

22. Clifford, M., Menon, R., Gangi, T., Condon, C., & Hornung, K. (2012). *Measuring school climate for gauging principal performance: A review of the validity and reliability of publicly accessible measures*. Washington, DC: American Institutes for Research.
23. Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning*. New York, NY: The Wallace Foundation.

Teacher Quality	Percentage of graduates (licensed/endorsed) placed as principals with improved measures of teacher effectiveness in their school after two years of leadership.	**This indicator attempts to capture the influence of school leaders on the effectiveness of the teaching that takes place within their school.** One prominent role of a principal is to improve the effectiveness of teaching in a school.[24] In fact, principals' impact on student achievement is mostly indirect and occurs through their impact on instruction.	The field is currently attempting to build systems that create meaningful measures of teacher effectiveness. However, such systems are still evolving; thus, the results from their measures of effectiveness may not be available or appropriate for use in all states. For example, many states have adopted measures of teacher effectiveness that heavily weight measures based on changes in student achievement but do not use a statistical approach that controls for factors outside the control of teachers.[25] States should also be careful about interpreting changes in small schools, especially if teacher effectiveness is heavily weighted by estimates of teacher-driven improvements in student outcomes. Research has consistently shown that smaller schools have smaller class sizes. Both small classes and small schools are strongly associated with greater volatility in such measures, and yearly increases or decreases can be misinterpreted as important changes in teacher effectiveness.[26] In sum, the types of measures used for teacher quality are important and should be used with caution. While measures of the quality of individual teachers can be unstable, they can be more useful when aggregated to the school level.[27] They can also be useful in distinguishing the tail ends of the distribution of teacher quality.[28] Therefore, if a preparation program's graduates consistently have lower rates of teacher effectiveness, it is worth investigating the root causes of that pattern.

24. Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004); Louis, K. S., Leithwood, K., Wahlstrom, K. L., & Anderson, S. E. (2010). *Learning from leadership: Investigating the links to improved student learning.* New York, NY: The Wallace Foundation.

25. Collins, C., & Amrein-Beardsley, A. (2013). Putting growth and value-added models on the map: A national overview. *Teachers College Record*, 116(1), 1–34; Ehlert, M., Koedel, C., Parsons, E., & Podgursky, M. (2014). Selecting growth measures for use in school evaluation systems should proportionality matter? *Educational Policy.* Advance online publication. doi:0895904814557593; Fuller, E. J., Hollingworth, L., & Liu, J. (2015). Evaluating state principal evaluation plans across the United States. *Journal of Research on Leadership Education*, 10(3), 164–192.

26. Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., ... Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers* (EPI Briefing Paper No. 278). Washington, DC: Economic Policy Institute; Di Carlo, M. (2015). Sample size and volatility in school accountability systems [Blog post]. Retrieved from http://www.shankerinstitute.org/blog/sample-size-and-volatility-school-accountability-systems

27. Darling-Hammond, L., Audrey Amrein-Beardsly, A., Haertel, E., & Rothstein, J. (2011). *Getting teacher evaluation right: A background paper for policy makers*. Washington, DC: American Educational Research Association.

28. Glazerman, S., Loeb, S., Goldhaber, D., Staiger, D., Raudenbush, S., & Whitehurst, G. (2010). *Evaluating teachers: The important role of value-added*. Washington, DC: Brown Center on Education Policy, Brookings Institution; Otterman, S. (2010, December 26). Hurdles emerge in rising effort to rate teachers. *The New York Times*. Retrieved from http://www.nytimes.com/2010/12/27/nyregion/27teachers.html; Corcoran, S., & Goldhaber, D. (2013). Value added and its uses: Where you stand depends on where you sit. *Education and Financial Policy*, 8(3), 418-434.

Student Outcomes			
Indicator	**Description**	**Purpose and Rationale**	**Methodological Considerations**
Student Achievement Outcomes	Percentage of graduates (licensed/ endorsed) placed as principals with positive gains in student achievement measures (as defined in Tool A) after three years of leading a school in their state (if valid and reliable data are available).	**This indicator attempts to signal the influence of school leaders on student achievement outcomes.** Principal preparation programs prepare graduates to improve student outcomes, particularly student achievement.	Student achievement and non-cognitive outcomes can be influenced by many factors—not only the principal but also teaching quality, school culture, curricula, etc.[29] We recommend using gains in student achievement and noncognitive outcomes over a three-year period during which a principal is likely to influence some of these other school factors (e.g., by improving teacher quality or choosing the curriculum). In setting targets, states should consider that the extent of a principal's influence depends on the autonomy provided to the principal regarding these matters and the level of support he or she receives from the district office. Methods for measuring student achievement are constantly evolving, and states should consider investing in and using the best available methods that most accurately measure the principal's impact on student growth (for example, by including the maximum number of students in the models; controlling for other variables; and, when possible, examining the growth starting with when the principal took the helm of the school). Measures of noncognitive student outcomes vary across contexts. Each state would need to identify the most relevant indicators for its policy context and its available data for this purpose.
Non-cognitive Student Outcomes	Percentage of graduates (licensed/ endorsed) placed as principals with improved measures of noncognitive student outcomes (discipline, attendance, safety, parent ratings) after two years of leadership.	**This indicator attempts to signal the influence of school leaders on students' noncognitive outcomes.** Students need cognitive and noncognitive skills to be successful in college and careers.[30] Principals play an important role in influencing noncognitive outcomes by setting the culture and climate of the school.[31] They often directly influence discipline and attendance policies and/or execution of those policies.	

29. Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2013).
30. Kautz, T., et al. (2013). *Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success*. Paris, France: The Organization for Economic Co-operation and Development.
31. Leithwood, K., & Jantzi, D. (2000). The effects of transformational leadership on organizational conditions and student engagement with school. *Journal of Educational Administration*, 38(2), 112–129.

III. ANNUAL REPORT

A. Overview

This section describes the basic content and structure of an annual report on the indicators described in Section II and provides examples of how the data displays for the report could be designed. The first example shows a summary report with basic information for each principal preparation program in the state. The second example shows a program-level report with more detailed data for an individual principal preparation program.

As noted in the guide, we recommend that states pilot this system for at least one year (and preferably two) before setting targets and at least two years before making results public. The pilot period enables states to collect and analyze data, share data with programs, and make any necessary changes to the system before sharing data publicly. The targeted review process is designed to help states determine whether results on these indicators tend to be useful or not. For example, if targeted reviews reveal that seemingly concerning admission rates are in fact uniformly not concerning, states may decide this indicator should not be included on a public annual report. This process of investigation also provides states with an opportunity to identify specific challenges and develop solutions to mitigate inappropriate interpretations and possible perverse incentives for how programs are implemented.

B. Summary Report
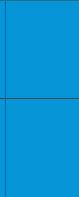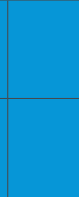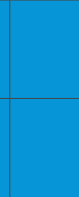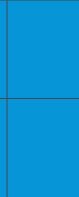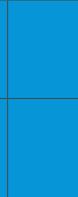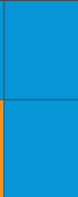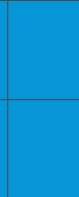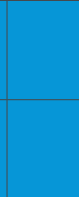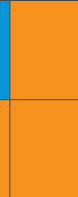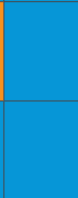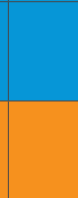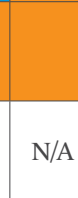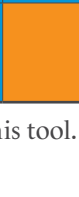
Ideal implementation of the annual report such as the one shown in the table below would enable each cell to be clickable, as follows:

- Clicking on "Western University Master's in Educational Administration" takes the user to the report for that program. The same is true for each program listed in the left column.

- Clicking on "Inputs" takes the user to a report for the program input data for all programs. The same is true for all of the other column headings (i.e., "Processes," "Outputs," and "Graduate Outcomes").

- Clicking on indicators such as "Admission Rate" takes the user to the description and considerations for that indicator (listed above).

As the following example illustrates, the summary report provides a side-by-side comparison of the extent to which each program has met state-set targets across all indicators. This type of reporting enables states to look across all relevant indicators to identify the programs that they want to engage in a targeted review. The program-level reports complement the side-by-side comparison by providing context-specific program information. In the example below, Central University Accelerated Master's Program has met fewer targets than other programs. The program-level report provides information that explains several of the missed targets. The fact that the program is fairly new, for instance, might explain why there is insufficient outcomes data for reporting. However, the large number of concerning results related to program inputs and processes might lead a state to go ahead and prioritize the program for targeted review to further explore the reasons for those results.

C. Summary Annual Report Example

Program Name	Inputs				Process					Outputs				Graduate Outcomes				
														Graduate and School Practices			Student Outcomes	
KEY — ▇ MEETS TARGETS, ▇ DOES NOT MEET TARGETS, na INSUFFICIENT DATA	Admission Rate	Teaching Experience	Instructional Expertise	Adult Leadership Experience	Rigor and Relevance	Quality of Faculty/Staff	Peer Interactions	Performance-Based Assessments	Residency Hours	Graduation Rate	Licensure Rate	Placement Rate	Retention Rate	Leadership Effectiveness Rate	School Climate	Teacher Quality	Non-Cognitive Student Outcomes	Student Achievement
Western University Master's in Educational Administration	Meets	Meets	Meets	Does Not Meet	Meets	Meets	Meets	Meets	Meets	Meets	Meets	Meets	Meets	Meets	Meets	Meets	Does Not Meet	Does Not Meet
Central University Master's in Educational Administration	Meets	Meets	Meets	Meets	Meets	Meets	Meets	Meets	Meets	Meets	Meets	Meets	Meets	Meets	Meets	Meets	Meets	Does Not Meet
Eastern School District Administrator Preparation Program	Does Not Meet	Meets	Meets	Meets	Meets	Meets	Does Not Meet	Meets	Meets	Meets	Meets	Meets	Does Not Meet	Meets	Meets	Meets	Does Not Meet	Does Not Meet
Central University Accelerated Master's Program	Meets	Does Not Meet	Does Not Meet	Does Not Meet	Meets	Does Not Meet	Meets	Meets	Does Not Meet	Does Not Meet	Does Not Meet	Does Not Meet	Meets	N/A	N/A	N/A	N/A	N/A

*Targets are set by the state per discussion in the introduction of this tool.

D. Program-Level Annual Report Example

Name of Institution: *Central University Accelerated Master's Program*
Website: *www.centraluniversity.edu/edadmin*
Program Approval Status: *Approved*
Accreditation Status: *Accredited*
Mission and Brief Description of Program: *The mission of the Central University Accelerated Master's Program is to provide a rigorous, practice-based preparation for the principalship to administrative candidates with high potential. The program is limited to candidates who meet a high bar for entry (including recommendation from their district superintendent) and are assigned to roles where they can practice their leadership skills while taking courses. The program targets master teachers who are ready to take on teacher leadership roles and typically takes two to three years to complete.*

	Indicator	Data*	Target**	Notes from Program	Notes from State
Program Inputs	Admission Rate	60%	60-85%	N/A	Meets expectations according to rubric, which targets 60–85% acceptance rate.
	Teaching Experience	90%	≥90%	N/A	Meets expectations because 90% of participants had at least two years of teaching experience.
	Instructional Expertise	25%	≥95%	We target teachers who are entering teacher leadership roles. Many have only been teaching for two years, so they often do not have an effectiveness rating at the time of application.	Concerning. How does the program know whether the candidates have enough expertise to lead the instructional improvement of other teachers?
	Adult Leadership Experience	30%	≥75%	We target teachers who are entering teacher leader roles, so they have not had prior leadership experience. However, greater than 75% are in leadership roles once the program starts.	Meets expectations with justification.
Program Processes*	Rigor and Relevance				
	Quality of Faculty/Staff				
	Peer Interactions				
	Performance-Based Assessments				
	Residency/Internship Hours				
Program Outputs	Graduation Rate				
	Licensure Rate				
	Placement Rate				
	Retention Rate				
Program Graduate Outcomes — Graduate & School Practices	Leadership Effectiveness Rate				
	School Climate				
	Teacher Quality				
Program Graduate Outcomes — Student Outcomes	Student Achievement Outcomes				
	Student Non-cognitive Outcomes				

> **This report is intended to be completed for each program according to targets set by each state, which would be informed by the availability, quality, and methodological considerations of their data as well as other contextual factors.**

* Data would not be included in public version—only the extent to which the data met or did not meet the target.
** Normative or performance targets are set by the state. Targets in this chart are for illustrative purposes only.
*** This section and the subsequent sections provided to illustrate the template that would be completed for each program.

New Leaders

IV. IMPORTANT CONSIDERATIONS IN INTERPRETING RESULTS

We encourage states to keep the following considerations in mind when deciding how to interpret results from annual reports.

Data are imperfect, and therefore indicators should not be used in isolation for high-stakes decisions

Research has consistently demonstrated the limitations of specific data indicators as measures of quality (including, for example, a recent statement from the American Education Research Association on the use of value-added measures of student achievement[32]). Put simply, single types of data on their own are often imperfect.

But imperfect data can provide important and useful information when used appropriately. For example, it is well documented in the medical field that mammograms have a 50–60% false-positive rate and a 20% false-negative rate. Given the costs and risks of mammograms (such as exposure to radiation and stress from false-positive tests), leading medical organizations differ in their recommendations about the age at which women should start receiving mammograms and how often they should be administered. However, no one in the medical field recommends discontinuing mammogram testing altogether. To the contrary, mammograms are a critical diagnostic tool that generates data that doctors examine in conjunction with other data to make decisions about further testing and treatment. No doctor would ever use mammogram data alone to recommend surgery or make other make high stakes decisions.

What is true for physicians reading a mammogram result is also true for state education leaders looking at the graduation rate of a principal preparation program or the growth in student achievement in schools led by a program's graduates. By themselves, these indicators offer only limited insights, but combined with a deeper professional review, they can help state officials arrive at a full and accurate picture of program quality. That fuller picture can be the basis for states to make consequential decisions about program approval and can be the impetus for continuous improvement of all programs.

The evaluation approach assumes relationships between indicators and preparation program quality

The indicators reflect a theory of action that assumes that preparation programs influence the quality of leadership practice, which in turn influences teacher quality and school climate, and these school characteristics ultimately impact student outcomes.[33] While there is limited research connecting principal preparation programs to all of the indicators, there is a strong research base connecting each of the direct relationships within this theory of action. For example, research has shown that a positive school climate is associated with and predictive of improved student outcomes.[34] In addition, some research concludes that effective leadership practices are associated with improved perceptions of school climate.[35] Therefore, there is a research base to support assumptions that preparation programs can have some indirect influence on our recommended indicators. We recommend additional research to test the assumptions in our theory of action.

Annual reports are important to use as part of a two-stage process, where results on report indicators are used to determine the need for in-depth reviews of program quality

We cannot stress enough the importance of using annual report indicators diagnostically to determine whether programs require targeted review. Similar to how doctors would never decide to perform surgery based on mammogram results alone, states should never use dashboard indicator data alone to make high-stakes decisions about preparation programs.

In order to make appropriate judgments based on accurate results, states will need additional contextual information. For example, a 100% admission rate for a program could indicate that the program does not have rigorous admissions, or it could be a result of targeted recruiting and effective prescreening, such that only strong applicants apply. Therefore, it is important to provide programs with opportunities to explain their results so that states can better interpret those results and draw conclusions. These information exchanges about root causes can be the basis of productive conversations about program quality and improvement.

A two-stage approach to using data to inform decision making is particularly important for incorporating student outcomes results, which are challenging to interpret for a number of reasons. First, student outcomes can be influenced by many factors, including prior student achievement and other school and district factors. Even if a state's value-add methodology is effective in isolating principals' impact on student outcomes, it can be difficult to attribute student outcomes to preparation programs.[36] Nonetheless,

32. AERA Council. (2015). AERA statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher, 44*(8), 448–452.
33. Orr, M. T., Young, M. D., & Rorrer, A. K. (2010). *Developing evaluation evidence: A formative and summative evaluation planner for educational leadership preparation programs.* Charlottesville, VA: UCEA National Center for the Evaluation of Educational Leadership Preparation and Practice.
34. Clifford, M., Menon, R., Gangi, T., Condon, C., & Hornung, K. (2012).
35. Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004).
36. Attribution to preparation programs can be especially difficult when graduates assume other roles (such as school coach, department chair, or assistant principal) for two to five years before assuming a principalship. The skills that program participants gain while in these other roles—and the ongoing support they receive from mentors and the district—are likely to contribute to their ultimate effectiveness as principals.

student outcomes results can be useful for identifying outliers or unusual results (e.g., programs whose graduates consistently have lower student outcomes results) and therefore can be useful diagnostic data for further investigation. Similarly, targeted review of variations in admission rates, placement rates, and all of the other indicators could either allay concerns or lead to identification of areas for improvement.

Baseline data, resources, policy context, and stakeholder input are important for target setting

The annual report portrays program results relative to targets. We recommend that states set performance targets informed by the following:

1. **Baseline data.** We recommend collecting at least one year of data (and preferably two) prior to establishing cut scores. In doing so, we encourage states to examine the normative distribution of results on each indicator. States might also consider the extent to which they expect results to vary from one year to the next. For example, placement rates might be expected to fluctuate depending on the economy or school consolidations.

2. **Available resources.** States may want to set targets so that they help to identify a certain number of preparation programs for targeted review that is feasible given resource constraints.

3. **Policy context.** States may want to consider the policy context, including the extent to which they want to push programs toward more ambitious goals, in deciding targets.

4. **Stakeholder input.** We recommend that states vet their rubric and targets with stakeholders, including leadership preparation programs.

States will want to carefully consider how to balance all of these inputs in setting targets. For example, suppose that in determining a target for licensure pass rates, a state collects baseline data showing an average pass rate of 70% and a range of 62 to 100%. The state also notes that the exam has a national pass rate of 80%. Because the state wants to encourage preparation programs to achieve higher pass rates, it chooses an 80% pass rate as the target.

Opportunities for programs to explain results can help allay concerns related to public reporting

During the review process for these tools, several reviewers expressed concern that inappropriate conclusions might be drawn from public reports, particularly because the data are imperfect and the public is unlikely to be fully aware of all of issues that should be considered in interpreting results. As noted above, we recommend a two-year piloting period during which states collect data on the indicators and explore the usefulness and limitations of the data via the targeted review process. After the pilot period, we recommend that states make the annual reports available publicly. The public versions of these reports would indicate whether programs met targets on input, process, output, and graduate outcome indicators. Information on inputs (such as admission rates and the percentage of candidates with adult leadership experience) and outputs (such as placement and retention rates) could be particularly valuable to applicants deciding where to apply, as well as to districts deciding on program partners. If the piloting period uncovers challenges or complexities in interpreting particular indicators, states may not want to report those indicators publicly.

Since programs might have reasonable cause for atypical results on these indicators, we recommend that the reporting system provide opportunities for programs to publically explain their results. For example, a program might explain unusually low percentages of candidates with prior leadership experience by stating, "We target teachers who are entering teacher leader roles, so they have not had prior leadership experience. However, greater than 75% are in leadership roles once the program starts." Another program might explain placement rates that do not meet targets by pointing out that the rural location of the school means that job openings are not as readily available. Another program might explain student achievement results that do not meet targets by stating, "We do not think the student achievement results accurately reflect the impact of our program because the majority of our graduates lead parochial schools, which are not included in the analysis." The reports could also help users understand the strengths and limitations of the indicators and data they make available.[37]

---

37. Feuer, M. J., Floden, R. E., Chudowsky, N., & Ahn, J. (2013)

Tool C:
Recommendations and Practices for In-Depth Review

I. INTRODUCTION

The model process for review described in the guide suggests that every principal preparation program undergo an in-depth review periodically (at least every seven years) to promote ongoing improvement to all aspects of program design and implementation. The model process also suggests that states review some programs more frequently. See Tool D for a description of these more frequent, targeted reviews.

In-depth reviews already take place. Indeed, programs typically experience up to four types of reviews:

1. accreditation reviews designed to certify that programs conferring degrees meet standards of quality;

2. state reviews (which in some states are combined with accreditation reviews) designed to monitor program quality and help states make consequential decisions about initial and ongoing approval of programs;

3. institutional reviews, which are required by the university in which the program is housed and are designed to foster self-study and ongoing improvement; and

4. professional association reviews designed to promote improvement against standards of excellent practice.

We do not recommend that states add a new process. Rather, we offer recommendations and specific practices that any of these processes could include. The recommendations and practices address the following areas:

- recommended data to consider;

- the process of review;

- the composition of the review team; and

- the rubric used to assess quality across all aspects of program design and implementation.

II. DATA

Many of the indicators included in the annual report are also relevant data to consider in the in-depth review process, and we recommend including a longitudinal review of the Annual Report results. These results, however, are limited to data that can be obtained annually and are reasonable in terms of the amount of effort or resources necessary to obtain and compile them on an annual basis. The following chart provides examples of additional data sources that could be integrated into the In-Depth Review process. These data are relevant and valuable, but they are only included in the In-Depth Review because they take additional resources to obtain, compile and analyze, and/or they are only available for certain programs.

Program Inputs

- The rigor and quality of recruitment and selection processes in admitting candidates with strong leadership potential

Program Processes

- The experience and expertise in educational leadership of instructors in the program
- The use of a defined competency framework—aligned with leadership standards—that describes the competencies a principal must have to influence school practices and culture in order to support student achievement gains
- The use of research-based and conceptually coherent content, curriculum, and instructional and assessment practices that align with the program's competency framework and incorporate problem-based learning strategies
- The implementation of supervised clinical practice with authentic opportunities to apply and practice leadership competencies
- The presence of collaborative relationships with other organizations (e.g., universities, school districts) for program enrichment and development, research, and practicum experiences
- The ongoing use of evaluation practices to support improvement

Program Outputs

- Assessments of graduate knowledge and skills developed through participation in the program
- Graduates' self-perceived readiness for leadership roles based on participation in the program

Program Graduate Outcomes

- 360° evaluations of program graduates, including assessments from teachers, supervisors and program graduates
- Student outcome growth in schools led by program graduates
- improvement in cognitive and non-cognitive measures of student outcomes in schools led by graduates

III. THE REVIEW PROCESS

We recommend careful attention to three process steps: (a) portfolio development, (b) portfolio review, and (c) site visits.

A. Portfolio Development

We recommend that program reviews begin with development of a program portfolio. The review and feedback process is facilitated when the organization of the portfolio is aligned with a program review rubric. The rubric included in this document, which reflects current research on preparation programs and aligns with the Council for the Accreditation of Educator Preparation's current accreditation program reporting requirements, delineates the core areas of the program review and required data sources and artifacts, as well as key questions to pose during the review process.

The portfolio provides narrative information that describes the program, how it is delivered, and how it meets state expectations. The portfolio demonstrates how the program addresses state (or national) standards; collects evidence regarding candidates' professional knowledge and skills, professional dispositions, and accomplishments in relation to state (or national) standards; analyzes the validity and reliability of the results from the evidence; and uses the analysis to improve the quality of the program's work. The narrative explains distance-education formats (if applicable), and distance-education options for completing the program are evaluated separately. In addition to assembling evidence, the narrative highlights what program faculty believe to be areas of strength and areas for improvement. The narrative is not overly burdensome on programs (i.e., it is no more than 25 pages in length).

The portfolio references and provides analyses of the four types of data delineated above: (1) program inputs, (2) program processes, (3) program outputs, and (4) program graduate outcomes. Finally, the portfolio includes artifacts to which the review team can refer during the review process. While some artifacts are specified in the rubric (e.g., course syllabi, internship handbooks, candidate assessment data, alumni surveys), programs may also include other materials that facilitate a robust understanding of the program.

B. Portfolio Review

After receiving the portfolio, the state assigns a three-person team to review the program materials using the program review rubric. The rubric not only facilitates the review of program artifacts and data but also allows reviewers to assess core areas along a developmental scale (ineffective, needs improvement, effective and highly effective). The review of materials provides the review team with a strong understanding of the program and enables the identification of strengths and weaknesses as well as areas for further examination during the site visit review.

In the event that the program-review team members are unable to find essential information or data summaries during their review of the materials, the team leader summarizes their information needs and requests the information from the program. All correspondence, including the request, is copied to relevant state personnel. Deadlines for responding to the request accommodate the timelines of the program review team and the time required for the program to adequately respond. Once the review team has had an opportunity to review the full complement of program materials, complete the program review rubric, and identify key areas for follow-up during the site visit, the team moves to stage three of the process.

C. Site Visit

The third stage of the review process involves a site visit. As noted above, the site visit is ideally conducted in conjunction with the program's institutional, accreditation, or professional association review—and at least one member of the review team joins the institutional, accreditation, or professional association review team on its site visit.

The agenda for the visit is designed to complement the review of the program portfolio and to enable further examination of any areas identified as concerns or weaknesses during the review. The visit includes a combination of observations, interviews, and further data review. For example, during the site visit, reviewers conduct individual interviews or focus groups with program leaders, faculty, current program participants, graduates and employers. Reviewers also have an opportunity to observe one or more classes, a coaching session and other relevant activities.

At the end of the site visit, the visitation team meets with college- and program-level leaders to debrief them on the key findings from the visit.

Immediately following the site visit, team members develop a report that reflects data reviewed during all phases of the review process. The report also identifies specific action steps for both the program and the state. Recommended action steps for the program are designed to support continuous program improvement.

The draft report is shared with both the appropriate state authority and the institution undergoing review. The institution has an opportunity to respond to any and all descriptions, ratings and recommendations before the report is finalized. Based on the institution's feedback and in light of any additional data or information provided, a final version of the report is developed, identifying specific action steps for both the program and the state. In the unlikely event that program leaders believe their program has been misrepresented, avenues are provided by the state for appeal.

IV. THE REVIEW TEAM

Our recommendations for the composition and role of review teams for in-depth review are the same as for the model process we outline in Tool D.

V. RUBRIC

The following program review rubric, which was drawn and modified from the University Council for Educational Administration's Institutional and Program Quality Criteria,[1] focuses on both program features and program outcomes. The rubric is designed to illuminate the difference between program practices that are highly effective, are effective, need improvement, or are ineffective for each criterion. Ratings for each criterion are based on expert judgments regarding the presence of a majority of the practices for each indicator and the collective effect of those practices. In addition to facilitating program review, the rubric, which details expectations for quality leadership preparation, can support program planning and improvement efforts and related conversations.

---

1. Young, M. D., Orr, M. T., & Tucker, P. D. (2012). *University Council for Educational Administration (UCEA) institutional and program quality criteria: Guidance for master's and doctoral programs in educational leadership.* Charlottesville, VA: UCEA.

Program Inputs			
The rigor and quality of recruitment and selection processes in admitting candidates with strong leadership potential			
Highly Effective	Effective	Needs Improvement	Ineffective
• Has a systematic written plan for recruitment and selection of candidates who demonstrate leadership potential; shows how the plan builds on the host institution's recruitment and admission practices. • Uses a defined set of strategies for recruiting applicants, including a variety of media and personal recommendations. • Differentiates strategies to seek applicants who demonstrate different types of leadership potential. • Has clearly articulated selection criteria; and applicants are afforded multiple methods to document academic and leadership potential. • Has balanced assessment of multiple robust sources of evidence on academic and leadership potential in admissions decisions. • Provides an excellent plan for increasing the diversity of students entering the program.	• Has a written plan for recruitment and selection of candidates who demonstrate leadership potential. • Uses a defined set of strategies for attracting and recruiting applicants who demonstrate leadership potential. • Has stated selection criteria and applicants are afforded more than one method to document academic and leadership potential. • Has assessment of two to three sources of evidence on academic and leadership potential in admissions decisions. • Provides an adequate plan for increasing the diversity of students entering the program.	• Has a weak recruitment and admission plan that may or may not be written. • Uses a variety of strategies for attracting and recruiting applicants. • Selection criteria are stated, and applicants are afforded only one method to document academic and leadership potential. • Admission decisions involve a limited assessment of evidence on academic and leadership potential. • Provides a weak plan for increasing the diversity of students entering the program.	• No evidence in this area.

Program Processes			
The experience and expertise in educational leadership of instructors in the program			
Highly Effective	Effective	Needs Improvement	Ineffective
• Program faculty and/or instructors are selected based on at least four of the following five criteria: scholarly expertise, content knowledge, relevant professional experience, demonstrated effectiveness in educational leadership, and evidence of teaching quality. • Faculty members stay informed by reading professional and research journals, newsletters, and blogs in education and leadership. Faculty members frequently share research-and-practice books, articles, and other resources on best practices focused on the essential problems of schooling, leadership, and administration with each other and students. • Course observations and evaluations are regularly reviewed for all courses. • Institution provides frequent opportunities for faculty to enhance their teaching, content knowledge, evaluation and research-utilization skills based on an assessment of faculty needs.	• Program faculty and/or instructors are selected based on at least three of the following five criteria: scholarly expertise, content knowledge, relevant professional experience, demonstrated effectiveness in educational leadership, and evidence of teaching quality. • Faculty members stay informed by reading professional and research journals, newsletters, and blogs in education and leadership. Faculty members sometimes share research-and-practice books, articles, and other resources on best practices focused on the essential problems of schooling, leadership, and administration with each other and students. • Course observations and evaluations are regularly reviewed for most courses. • Institution provides frequent opportunities for faculty to enhance their teaching skills and content knowledge.	• Program faculty and/or instructors are selected based on at least two of the following five criteria: scholarly expertise, content knowledge, relevant professional experience, demonstrated effectiveness in educational leadership, and evidence of teaching quality. • Faculty members stay informed by reading professional and research journals, newsletters, and blogs in education and leadership. Faculty members occasionally share research-and-practice books, articles, and other resources with each other and students. • Course observations and evaluations are occasionally used for some courses. • Some form of professional development for teaching faculty is available.	• No evidence in this area.
The use of a defined competency framework—aligned with leadership standards—that describes the competencies a principal must have to influence school practices and culture in order to drive student achievement gains			
Highly Effective	Effective	Needs Improvement	Ineffective
• Program has a formally articulated theory of action for the course sequence, teaching strategies, learning activities, and assessments. Student outcomes are clearly stated, and program design is aligned with these outcomes. Program faculty have developed a crosswalk of course content, learning activities, and assessments that are aligned with Interstate School Leaders Licensure Consortium (ISLLC) or other leadership standards and elements.	• Course sequence, teaching strategies, learning activities, and assessments are described in materials. Student outcomes are clearly stated, and program design is aligned with these outcomes. Program faculty have developed a crosswalk of course content, learning activities, and assessments that are aligned with ISLLC or other leadership standards and elements.	• Course sequence, teaching strategies, learning activities, and assessments are described in materials. Student outcomes are described. Program faculty have developed a crosswalk of course content that is aligned with ISLLC or other leadership standards and elements.	• No evidence in this area.

New Leaders

Program Processes			

The use of research-based and conceptually coherent content, curriculum, and instructional and assessment practices that align with the program's competency framework and incorporate problem-based learning strategies			
Highly Effective	Effective	Needs Improvement	Ineffective
• All syllabi reflect a rich blend of research- and practice-based content that addresses the essential problems of schooling, leadership, and administration. Readings and learning activities in each course almost always promote a better understanding of the existing research on course content. Students are engaged in critically assessing implications for practice.	• Course sequence, teaching strategies, learning activities, and assessments are described in materials. Student outcomes are clearly stated, and program design is aligned with these outcomes. Program faculty have developed a crosswalk of course content, learning activities, and assessments that are aligned with ISLLC or other leadership standards and elements.	• Course sequence, teaching strategies, learning activities, and assessments are described in materials. Student outcomes are described. Program faculty have developed a crosswalk of course content that is aligned with ISLLC or other leadership standards and elements.	• No evidence in this area.

The implementation of supervised clinical practice with authentic opportunities to apply and practice leadership competencies			
Highly Effective	Effective	Needs Improvement	Ineffective
• Has formally established collaborative relationships with one or more local districts, professional associations, or other agencies to develop and support sites for clinical study and residency. Candidates are provided a sustained school internship with substantial and regular field experiences over an extended period of time (6–12 months). • Includes planned, purposeful, developmentally sequenced, standards-based supervision of students in clinical settings. Field experiences and clinical internship offer a range of opportunities for candidate responsibility in leading, facilitating, and making decisions typical of those made by educational leaders. Candidates are provided with opportunities in two or more types of school settings and a variety of community organizations. Provides candidates with multiple opportunities to work with students and teachers from diverse groups. Supervised and coached by both program and field-based supervisors.	• Has a collaborative relationship with one or more local districts, professional associations, or other agencies to develop sites for clinical study and residency. Candidates are provided a sustained school internship with substantial and regular field experiences over at least one entire semester. • Includes planned, developmentally sequenced, standards-based supervision of students in clinical settings. Field experiences and clinical internship offer a few opportunities for candidate responsibility in leading, facilitating, and making decisions typical of those made by educational leaders. Candidates are provided with opportunities to gain experiences in a school setting and community organizations. Provides candidates with occasional opportunities to work with students and teachers from diverse groups. Supervised by program and/or field-based supervisors.	• Consults with one or more local districts, professional associations, or other agencies to develop sites for clinical study and residency. Candidates are provided a school internship with intermittent field experiences over a semester. • Lacks structured, standards-based supervision of students in clinical settings. Field experiences and clinical internship do not offer opportunities for candidate responsibility in leading, facilitating, and making decisions typical of those made by educational leaders. Candidates are not provided with opportunities to gain experiences in different types of school settings or community organizations. Provides candidates few or no opportunities to work with students and teachers from diverse groups.	• No evidence in this area.

Program Processes			
The presence of collaborative relationships with other organizations (e.g., universities, school districts) for program enrichment and development, research, and practicum experiences			
Highly Effective	Effective	Needs Improvement	Ineffective
• Has formally established collaborative relationship (through MOU or other mechanism) with one or more local districts, professional associations, or other agencies to (1) recruit program candidates, (2) select program candidates, (3) strengthen program focus and content, (4) develop and support sites for clinical study and residency, (5) develop and support sites for applied research, (6) promote diversity within the preparation program and field, and (7) provide feedback on the program and program graduates.	• Has a collaborative relationship with one or more local districts, professional associations, or other agencies to (1) recruit program candidates, (2) select program candidates, (3) develop sites for clinical study and residency, (4) develop sites for applied research; (5) promote diversity within the preparation program, and (6) provide feedback on the program and program graduates.	• Consults with one or more local districts, professional associations, or other agencies to (1) develop sites for clinical study and residency and (2) provide feedback on the program and program graduates.	• No evidence in this area.
The ongoing use of evaluation practices to support improvement			
Highly Effective	Effective	Needs Improvement	Ineffective
• Competency-based formative data are used multiple times during the program to give students feedback about their leadership skills, practices and knowledge in individual courses and overall. Standards-based summative assessments of student performance are used in courses and the program as a whole.	• Competency-based formative data are used at least once in the program to give students feedback about their leadership skills, practices and knowledge in individual courses and overall. Standards-based summative assessments of student performance are used in courses.	• Competency-based formative data are used to give students feedback about their leadership skills, practices and knowledge in some courses. Standards-based summative assessments of student performance are used in some courses.	• No evidence in this area.

Program Outputs			
Assessments of graduate knowledge and skills developed through participation in the program			
Highly Effective	Effective	Needs Improvement	Ineffective
• Program bases assessments of candidate and graduate knowledge on an articulated theory of action for the program and national (or local or state) leadership standards. Student outcomes are clearly stated, and assessments are aligned with these outcomes.	• Program's assessment of candidate and graduate knowledge is aligned with national (or local or state) leadership standards. Student outcomes are articulated. Student assessments are aligned with outcomes and standards.	• Program's assessment of candidate and graduate knowledge is aligned with national (or local or state) leadership standards. Student assessments are aligned with these standards.	• No evidence in this area.
Graduates' perceptions of readiness for leadership roles based on participation in the program			
Highly Effective	Effective	Needs Improvement	Ineffective
• Program makes use of a graduate exit survey to learn about graduates' perceptions of their readiness. At least 90%[2] of program graduates report high levels of readiness. The graduate perception data is disaggregated across each of the following five areas to inform program improvement: (1) rigor and quality of recruitment and selection; (2) experience and expertise of instructors; (3) use of a defined competency framework that is aligned with leadership standards; (4) research-based and conceptually coherent content, curriculum, and instructional and assessment practices; and (5) supervised clinical practice.	• Program makes use of a graduate exit survey to learn about graduates' perceptions of their readiness. At least 70% of program graduates report high levels of readiness. The graduate perception data is disaggregated across at least two of the following five areas to inform program improvement: (1) rigor and quality of recruitment and selection; (2) experience and expertise of instructors; (3) use of a defined competency framework that is aligned with leadership standards; (4) research-based and conceptually coherent content, curriculum, and instructional and assessment practices; and (5) supervised clinical practice.	• Program makes use of a graduate exit survey to learn about graduates' perceptions of their readiness. Less than 70% of program graduates report high levels of readiness. Data is examined as a part of program improvement planning.	• No evidence of attempts to either collect or disaggregate graduate perception data.

---

2. The percentages included in this section are based on the Initiative for Systemic Program Improvement through Research in Educational Leadership (INSPIRE) Program Graduate Survey, which is part of the INSPIRE program evaluation suite. We strongly recommend that programs use the INSPIRE survey tools; however, if a different set of survey tools is used, then these percentages would need to be adjusted to reflect their norms. It would be important to include response rates for reported perception data measures.

NewLeaders

Program Graduate Outcomes			
360° evaluations of program graduates, including assessments from teachers, supervisors and program graduates			
Highly Effective	Effective	Needs Improvement	Ineffective
• Program makes use of a 360° survey to learn about perceptions of program graduates' effectiveness once those graduates have taken on leadership positions. Based on these surveys, >70% of principals are rated effective, OR >50% of assistant principals are rated effective on a principal rubric (or >70% if on an assistant principal rubric).[3] The 360° data is disaggregated across each of the following five areas to inform program improvement: (1) rigor and quality of recruitment and selection; (2) experience and expertise of instructors; (3) use of a defined competency framework that is aligned with leadership standards; (4) research-based and conceptually coherent content, curriculum, and instructional and assessment practices; and (5) supervised clinical practice.	• Program makes use of a 360° survey to learn about perceptions of program graduates' effectiveness once those graduates have taken on leadership positions. <70% of principals are rated effective AND <50% of assistant principals are rated effective on a principal rubric (or <70% if on an assistant principal rubric). The 360° data is disaggregated across at least two of the following five areas to inform program improvement: (1) rigor and quality of recruitment and selection; (2) experience and expertise of instructors; (3) use of a defined competency framework that is aligned with leadership standards; (4) research-based and conceptually coherent content, curriculum, and instructional and assessment practices; and (5) supervised clinical practice.	• Program makes use of a 360° survey to learn about perceptions of program graduates' effectiveness once those graduates have taken on leadership positions. <70% of principals are rated effective AND <50% of assistant principals are rated effective on a principal rubric (or <70% if on an assistant principal rubric). Data is examined as a part of program improvement planning.	• No evidence of attempts to collect, disaggregate or use 360° data.
Student outcome growth in schools led by program graduates[4]			
Highly Effective	Effective	Needs Improvement	Ineffective
• Program analyzes data on student outcome growth and use this analysis to inform program improvement.	• Program analyzes data on student outcome growth and uses this analysis to inform program improvement.	• Program examines data on student outcome growth as part of program improvement planning.	• No evidence of attempts to examine data on student outcome growth.
Improvement in measures of academic and nonacademic outcomes and school effects in schools led by program graduates (see note 4)			
Highly Effective	Effective	Needs Improvement	Ineffective
• Program analyzes measures of academic and nonacademic outcomes in schools led by graduates. Program analyzes measures of school effects in schools led by graduates. The results of these analyses are used to inform program planning and improvement.	• Program examines measures of academic and nonacademic outcomes in schools led by graduates, as well as measures of school effects in schools led by graduates, and considers ways the data might inform program changes.	• Program examines measures of academic and nonacademic outcomes in schools led by graduates, as well as measures of school effects in schools led by graduates.	• No evidence of attempts to examine such data.

Scale: Very Effective—Thorough and exceeds standard; Effective—Meets basic standard; Needs Improvement—Does not yet meet standard; Ineffective—No evidence of meeting the standard.

3. These percentages are based on the INSPIRE survey suite. We strongly suggest that programs use the INSPIRE survey, but if they do not, their norms would need to be adjusted based on the norms of the surveys they use.
4. This section of the rubric would only be used if the state education agency is able to provide programs with data needed to conduct the suggested analyses.

New Leaders

SEP³ Toolkit | STATE EVALUATION OF PRINCIPAL PREPARATION PROGRAMS

Tool D:
Handbook for Targeted Review

I. INTRODUCTION

This handbook outlines the process for targeted review of programs. As described in the guide, targeted review occurs in any year following the release of annually reported data (as opposed to in-depth review, which occurs on a multiyear cycle). Every program—on its own—may want to undertake the kind of examination described below in order to support ongoing improvement efforts. But it is not realistic for states to review every program every year, given their limited capacity. Thus, we recommend that states prioritize efforts to improve programs whose annual report data show the greatest cause for concern. In these cases, targeted review can be a useful strategy for confirming or disconfirming concerns raised in the annual report.

The targeted review model we describe in this handbook is informed by research, evaluation and improvement efforts within the field. As with the other tools, we offer this model with the expectation that states will adapt it to fit local needs.

The handbook is divided into five sections. Section II describes the purpose of targeted reviews, as well as timing considerations. Section III delineates the measures used to assess program quality. Section IV outlines the processes employed during program review. Section V discusses the necessary credentials and training for reviewers.

II. PURPOSE AND TIMING

Targeted review is designed to respond to data patterns identified in the annual report, enabling state personnel (or their designees) to diagnose and select programs with a pattern of concerning results for further review. The targeted review enables states to explore underlying causes for the concerning results. For example, a program with a 100% admissions rate might appear to be unselective, but in-depth review might reveal that potential applicants are prescreened by partner districts, leading to a high-quality pool of applicants—which would allay concerns about the high admissions rate. On the other hand, in-depth review might reveal that the high admissions rate is caused by pressures to maximize profit despite the quality of candidates, which would cause the state to identify admissions as an area in need of improvement for continued state approval in its conclusions for the in-depth review process.

Reviews allow state personnel and other team members to suggest improvements for the program's design and execution and, in some cases, to learn in detail about effective practices that can be shared with other programs. Further, targeted review can form the basis for state decisions about program status, with findings contributing to a state's determination as to whether a program can continue to operate and whether any specific improvements are required in the short-term as a condition for continued operation.

Although it might be ideal for each program to have an in-depth review to foster continuous improvement, states have a limited capacity to conduct such reviews. Since states have a responsibility to hold programs accountable, we recommend that they use annual reports to identify a small number of programs with the most concerning results for targeted review. The actual number of programs targeted depends on both a review of annual reports and the state's capacity to conduct the more intensive targeted reviews. In the meantime, programs not targeted for state review might choose to use their annual report results (especially in comparison to other programs' results) to identify areas that they want to investigate for themselves in support of their own continuous improvement. And, as described in Section III of the guide and Tool C, we recommend that all programs be reviewed in depth on a periodic basis in conjunction with accreditation timelines or other review processes.

Once states have established normative targets or performance targets for annual reports (see Tool B), they will need to establish decision rules for triggering a targeted, in-depth review. For example, a state may decide to prioritize reviews for programs that have a large number of indicators below expectations. For any program meeting the threshold for review, we recommend that the review occur as soon as possible after concerns are identified.

III. REVIEW MEASURES

Targeted review examines the same measures as those in the annual report—measures related to program inputs, program processes, program outputs, and program graduate outcomes. But a targeted review also explores program functions and features underlying the indicators in the annual report, asking:

- *To what extent are the data in the annual report accurate reflections of program quality?* Reviewers have the opportunity to examine additional data, disaggregate it (for example, by geographic area or school district), and analyze trends over time.

- *Why is the program meeting or not meeting expectations on a given measure?* Reviewers can engage program leaders in an analysis of the factors contributing to a program's relative strengths and weaknesses.

- *What should be done to improve program performance?* Reviewers and program leaders can analyze gaps between current and desired outcomes and identify strategies to close those gaps.

It is crucial to note that reviewers might conclude that the concerns triggering a review do not reflect problems of program quality. The implications of such a conclusion are described below.

IV. THE REVIEW PROCESS

The process of review depends on the scope of concerns emerging from the annual report. If the state considers the concerns to be relatively minor, program leaders could invite program leaders to submit a short memorandum providing evidence of program quality and indicating any improvement steps they are already taking. State leaders can review these documents, amend the annual report as necessary, and consider the review complete; or, state leaders might engage a team to conduct a site visit to gather more information.

For more significant reviews, the process proceeds as follows:

1. State leaders identify the particular measure(s) of concern needing deeper investigation and assign a review team.

See above for details about the identification of areas of concern. See the next section for details on the review team.

2. Program leaders gather additional data relevant to the measure(s) of concern and assemble knowledgeable program staff and stakeholders to contribute to the review.

Since the review is targeted and focused on developing an understanding of the measure(s) of concern, we recommend that the data collected be similarly focused and specific. For example, if the annual report suggests that a program is not adequately selective in admissions, program leaders might assemble: recruitment materials (e.g., brochures, work plans for recruitment), admissions materials (e.g., selection instruments, a cross-section of essays with names redacted), and admissions data disaggregated by important characteristics (e.g., sending district, race/ethnicity, gender, prior teaching experience).

The program team members need relevant knowledge and expertise in order to assess evidence sources. Continuing the example above, a deeper investigation of admissions would benefit from interviews with staff who have responsibility for recruitment and admissions, as well as staff from districts targeted for recruitment.

3. The review team analyzes the available data, develops conclusions, and identifies recommended action steps for program improvement.

Critical to the review is the use of a rigorous process for analyzing data on measures of concern and developing sound conclusions and action steps. We recommend two approaches to analysis: root cause analysis and gap analysis. These may be used in combination.

Root Cause Analysis

Root cause analysis is a well-established methodology used across a range of industry sectors to trace problems back to their source. Beginning with an identified problem (e.g., low satisfaction rates from program graduates), it asks program leaders and reviewers to consider why the problem exists. This inquiry continues until participants agree that they have arrived at the root cause of the problem. Because it is designed to surface and address the substantive issues that are leading to underperformance, it lends well to devising action plans to remedy those issues. For further information and tools for conducting root cause analysis, see http://asq.org/learn-about-quality/root-cause-analysis/overview/overview.html.

Consider this example: A program's annual report indicates that only 50% of graduates think the program content was rigorous and relevant to them in their current school leadership roles. Using root cause analysis, a review team looks more closely at the data and determines that most of the graduates concerned about program content now work in one school district. By asking why these graduates might be dissatisfied, it becomes clear (based on input from knowledgeable field supervisors from the program) that the district has seen a steady increase in the percentage of students who are non-native English speakers and has made the implementation of differentiated instructional strategies for these students a top priority for principals at

NewLeaders

University Council for Educational Administration
Quality Leadership Matters   since 1954

all levels. Further inquiry surfaces two facts: (1) for the past three years, the program has used a series of adjunct faculty with relatively little experience with English language learners to teach the only courses where this content is addressed; and (2) neither the syllabi for those courses nor the leadership practicum include opportunities to learn or practice up-to-date strategies for supporting English language learners. The review team concludes that these are root problems and need to be addressed in an action plan for improvement.

Gap Analysis

Gap analysis is a strategy used in program evaluation by organizations in both education and business to determine gaps between a desired state and a current one. Data are collected on key indicators that may indicate the status of a program based on relevant variables. Stakeholders then critically analyze the data for areas in which there are significant gaps and determine strategies to achieve the desired state using an improvement plan. The plan is then monitored using the same key indicators that were used for the original gap analysis. For further information and tools for conducting gap analysis, see http://www.ahrq.gov/professionals/systems/hospital/qitoolkit/d5-gapanalysis.pdf.

One of the benefits of the analysis methodologies described here is that they can help determine whether the data of concern reflect program quality or are attributable to other factors. Consider this example: A program is flagged for review because, for two years in a row, its rate of placement into principal and assistant principal roles has been low. Further review shows that the district where the program has historically placed candidates has just completed contract negotiations with substantial changes in compensation: Pay differentials for teachers assuming teacher leadership roles have been increased, and salaries for administrators have gone up. In part because of these changes, fewer principals and assistant principals are leaving, and candidates are being encouraged to take on teacher leadership roles. In this case, the program's data about placement may not be cause for concern.

4. State-designated reviewers complete a draft summary report identifying action steps for the program and the state.

At the end of the site visit, the review team meets with program-level leaders to discuss key findings and action steps based on the analysis. If the review disconfirms the original concerns, there may not be any action steps. If the original concerns are verified, state-designated reviewers develop a summary report that includes specific action steps for both the program (as needed) and the state. These action steps are designed to support continuous program improvement.

The draft report is shared with both the appropriate state authority and the institution undergoing review. The institution subsequently has an opportunity to respond to any and all descriptions, recommendations and action steps before the report is finalized. Based on the institution's feedback and in light of any additional data or information provided, a final version of the report is developed, again identifying specific action steps for both the program and the state. The state authority monitors the program based on the report's action steps and timeline. A determination is made as to whether further targeted review or a full in-depth review is warranted. In the unlikely event that program leaders believe their program has been misrepresented, they have avenues to appeal this determination.

V. THE REVIEW TEAM

We recommend the review team consist of at least three professionals, including one representative from the state department of education, one faculty member from outside the institution, and one school or district administrator. One of these individuals, preferably the individual with the most experience conducting reviews, would be designated the review team leader and would be responsible for submitting the final draft of the review report.

Reviewers possess relevant expertise for making professional judgments, including (as relevant to the measures driving the review):

- content expertise in leadership;

- understanding of adult learning theory;

- knowledge of current research on effective leadership preparation;

- ability to analyze programs for curricular alignment, evidence-based instructional strategies, and effective performance-based assessment;

- strong facilitation skills and expertise around leading root cause and gap analysis protocols; and

- understanding of program evaluation data and how to interpret it appropriately.

This list of reviewer expertise is similar to the list for in-depth reviews. We would particularly note the need for at least one reviewer with strong facilitation skills who can lead the program team through sophisticated analysis processes.

We recommend that potential reviewers fill out an application to participate as a reviewer in which they: (1) provide information on their experience with program review, (2) share their knowledge and experience with the standards and review process used in the state, and (3) explain why they are interested in becoming reviewers. We also recommended that reviewers submit at least one letter of recommendation that speaks to their ability to make program judgments. If faculty members are from the state in which the review is taking place, they should work in state-approved programs that have been identified as either effective or highly effective.

In addition to having reviewers with the above areas of expertise, states should provide training to reviewers to ensure that review team members are prepared to conduct an in-depth review. At a minimum, training familiarizes reviewers with state and national preparation program standards, how to read an annual report, how to conduct a review, how to collect and analyze data, how to facilitate analysis protocols, and ultimately how to present the data in a fair and meaningful way. Training includes information and guidance around these practices, as well as the opportunity for simulated practice (e.g., study a sample program report, participate in a mock review exercise). Training is provided on a regular basis to ensure sufficient capacity within the state for the review process.

Resource A:
Overview of How Preparation Programs
Are Currently Reviewed by States

I. STATE AUTHORITY FOR THE EVALUATION OF EDUCATIONAL LEADERSHIP PREPARATION PROGRAMS

There are two broad policy levers for improving school leadership: preparation program evaluation and individual candidate licensure. At the federal level, there is interest in fostering greater preparation program accountability as indicated by the U.S. Department of Education's proposed changes to the teacher preparation accountability requirements found in Title II of the Higher Education Act. However, thus far, there is no comparable guidance regarding leadership preparation programs, leaving oversight of an estimated 978 principal preparation programs to individual states.[1]

II. OVERSIGHT OF PREPARATION PROGRAMS

States have the authority to regulate both preparation programs for school leaders and licensing requirements for individuals seeking to become school leaders. Given these two policy levers, there has been more activity by states in adopting policies regarding individual candidate certification/licensure than in adopting policies related to the evaluation and oversight of preparation programs.[2] For example, only 23 states require preparation program reviews at specified intervals, but 40 states have regulations on the prior teaching experience of administrative candidates.

When states do enact policies focused on preparation programs, the emphasis is on program standards and program oversight, not program content or outcomes. All 50 states and the District of Columbia use nationally recognized leadership standards, and roughly half of states have some type of program oversight, as detailed in Table 1. However, few states regulate selection processes, such as the use of performance-based assessments (five states), or program elements, such as parameters for clinically rich internships. Only 15 states require 300+ hours of field-based experience, despite research evidence suggesting that the quality of clinical experience is probably the most important determinant of an effective preparation program.[3] While the number of hours does not necessarily indicate program quality, more hours do create an opportunity for a more meaningful field experience.

Table 1 presents findings from Anderson and Reynolds' review of the regulatory language regarding the approval and ongoing oversight of principal preparation programs.[4] For each element, an "X" indicates the element is required.

1. Briggs, K., Cheney, G. R., Davis, J., & Moll, K. (2013). *Operating in the dark: What outdated state policies and data gaps mean for effective school leadership.* Dallas, TX: The George W. Bush Institute. Briggs, Cheney, Davis, and Moll's findings are based on self-reporting by state education agencies in 2012.
2. Anderson, E., & Reynolds, A. (accepted). State of state policies for principal preparation program approval and candidate licensure. *Journal of Research in Educational Administration*.
3. Orr, M. T., & Barber, M. E. (2009). Program evaluation in leadership preparation and related fields. In M. D. Young, G. M. Crow, J. Murphy, & R. T. Ogawa (Eds.), *Handbook of research on the education of school leaders* (pp. 457–498). New York, NY: Routledge.
4. Anderson & Reynolds (accepted).

Table 1: State-by-State Results: Preparation Program Approval and Oversight

STATE	Requires state review at specified intervals	Includes documentation and site visits in plan for initial program oversight	Requires oversight team to have relevant experience and training	Includes feedback mechanism to improve practice
AL				
AK				
AZ		X	X	
AR	X	X	X	X
CA		X	X	X
CO		X		
CT	X	X	X	X
DE	X			X
DC		X	X	
FL	X	X		X
GA	X	X	X	X
HI				
ID	X	X	X	
IL	X	X	X	X
IN				X
IA	X	X	X	X
KS	X	X	X	X
KY	X	X	X	X
LA	X	X	X	X
ME	X	X	X	X
MD	X	X	X	X
MA		X	X	
MI	X			
MN	X	X	X	X
MS				
MO				
MT				
NE				
NV	X			
NH		X		X
NJ			X	
NM				
NY			X	
NC				
ND	X	X	X	X
OH				
OK				
OR				
PA		X		
RI	X	X	X	X
SC				
SD		X	X	
TN		X	X	X
TX	X			X
UT				
VT			X	X
VA	X	X	X	X
WA	X	X	X	X
WV	X	X	X	X
WI	X	X	X	X
WY				
Total	**23**	**28**	**27**	**25**

III. DATA COLLECTED FROM PREPARATION PROGRAMS

There is great variability in the data collected by states on their principal preparation programs. Most states collect basic workforce information on the number of preparation program graduates (31 states) and the number of licenses issued (43 states), but fewer collect data on the performance of those graduates or the programs from which they graduate. A few key findings from a study by Briggs, Cheney, Davis, and Moll (2013)[5] are summarized below. They are clustered by the same categories of indicators that we use in our proposed annual report: program processes, program outputs and program graduate outcomes.

Program Processes

- Only 27 states report including in their standards five key practices that current research shows are important to principal effectiveness: recruiting and selecting teachers, developing and supporting teachers, assessing and rewarding teachers, implementing data-driven instruction, and developing a positive school culture.

- Only five states report requiring principal preparation programs to include all key programmatic components that research shows are critical for effective programs: clear program purpose, competency framework, recruitment, candidate selection, coursework, clinical leadership experience, and program completion requirements.

Program Outputs

- Seventeen states collect data on principal job placement rates by principal preparation programs.

- Eleven states collect data on principal job retention rates among graduates of principal preparation programs.

Program Graduate Outcomes

- Fourteen states collect data on principal program graduates' job effectiveness as measured by on-the-job evaluations.

- Thirteen states collect data on principal program graduates' job effectiveness as measured by their impact on student achievement.

A. Data Required for Preparation Program Approval

For the explicit purposes of program approval, states require a range of documentation, from written applications (46 states and DC) to evidence of program graduates raising student achievement (eight states). Table 2 presents survey responses collected by the Bush Institute (Briggs et al., 2013) from state departments of education regarding the required data for program approval. Seven elements of the program approval process are noted with the timetable for approval.

---

5. Briggs et al. (2013).

Table 2: Data Required for Program Approval[6]

STATE	Written application	Site visit(s)	Evidence of graduates earning licensure	Evidence of graduates securing admin jobs	Evidence of graduates retaining jobs	Evidence of graduates' job effectiveness	Evidence of graduates raising student achievement	Programs required to reapply for approval every ___ years
AL	X	X	X	X	X	X	X	3 TO 7
AK		X						7
AZ	X	X						2 TO 5
AR	X	X	X			X	X	7
CA	X	X				X		7
CO	X	X						5
CT	X	X	X					5
DE	X							3
DC	X	X	X					3 OR 7
FL	X	X						6
GA	X	X	X					7
HI	X	X	X	X	X	X		1
ID	X	X	X					3
IL	X		X					1
IN	X							7
IA	X	X	X					7
KS	X	X	X					7
KY	X	X	X		X	X	X	7
LA	X	X	X			X	X	7
ME	X	X						5
MD	X	X						7
MA	X	X						5
MI	X							5
MN	X	X	X					5
MS	X	X				X		1
MO	X	X	X					1
MT	X	X	X					7
NE	X	X				X		7
NV	X	X						7
NH	X	X	X					5
NJ	X	X						5
NM								7
NY	X							7
NC						X	X	5
ND	X		X					7
OH	X	X	X	X	X	X	X	5 TO 7
OK	X							7
OR	X	X						7
PA	X		X					7
RI	X	X						1 TO 5
SC	X	X	X	X	X	X	X	7
SD	X	X	X	X				7
TN	X	X	X			X	X	7
TX	X	X						5
UT	X							5
VT		X						7
VA	X	X				X		2
WA	X	X	X	X				5
WV	X							7
WI	X	X	X	X				5
WY	X							7

6. Ibid.

IV. COMPARISON OF CURRENT DATA COLLECTION ON LEADERSHIP PREPARATION PROGRAMS AND ANNUAL REPORT INDICATORS[7]

Table 3 compares the proposed annual report indicators described in the guide with available research on what data are currently collected by state departments of education. The study by Briggs et al. (2013) made a distinction between data collected regarding ongoing program oversight and data collected explicitly for program approval. For example, Briggs and colleagues argued that states should monitor the number of graduates for program oversight because they need to know whether their preparation programs are producing enough new principals to meet the anticipated number of school leader vacancies each year. This distinction is indicated below. Currently, relatively few states collect the types of data this guide recommends.

Table 3: Extent to Which Data for Annual Report Indicators Are Currently Being Collected

Recommended Annual Report Indicators	Data Collected for Ongoing Program Oversight	Data Collected for Initial or Periodic Program Approval
Program Inputs		
Number of program starters	No	No
Admissions rate	No	No
Percentage of admitted candidates who have strong instructional expertise	No	No
Percentage of candidates who demonstrate leadership potential through experiences leading adults	No	No
Program Processes		
Use of performance-based assessments	No	No
Number of required residency hours	15 states require 300+ hours of field-based experience[8]	
11 states require the reporting of a clinical leadership experience by programs	No	
Program Outputs		
Percentage of candidates who complete the program	31 states were able to report the number of graduates from state-approved principal preparation programs[9]	No
Licensure exam results	30 states collect this data[10]	No evidence on this indicator
Number and percentage of program completers who become licensed by the state	43 states were able to report the number of principal licenses granted	24 states require evidence of licensure obtainment by graduates
Percentage of graduates placed as school leaders in the state within 3 years	13 states collect data on job placement rates	7 states require evidence of graduates securing administrative jobs
Retention rate	8 states collect data on retention	5 states require evidence of graduates retaining jobs

---

7. Except for evidence footnoted separately, all of the data were drawn from Briggs et al. (2013).
8. Anderson & Reynolds (accepted).
9. Briggs et al. (2013) did not address whether states collect data regarding the number of program starters, which would be needed to calculate the percentage who graduate.
10. Anderson & Reynolds (accepted).

New Leaders

uc ea   University Council for Educational Administration
Quality Leadership Matters   since 1954

Program Graduate Outcomes		
Percentage of graduates rated effective or above based on teacher and/or leadership effectiveness ratings	15 states collect data on principal performance based on principal evaluations	13 states require evidence of job effectiveness by graduates
Percentage of principals who graduated from the program and improved student outcomes after 3 years of leading a school based on consistent and methodologically sound measures of aggregated individual student growth	11 states collect data on principal performance based on student achievement impact (not clear on time frame for tenure of graduate)	8 states require evidence of graduates raising student achievement (unspecified methodology and time frame)
Other data points: measures of leadership practice; measures of school climate; measures of noncognitive student outcomes	No evidence on this indicator	No evidence on this indicator

New Leaders

Resource B:
Key Program Review Processes Impacting
Higher Education Programs

There are four types of reviews that higher education leadership preparation programs are likely to experience. They include accreditation reviews, institutional reviews, state reviews, and professional association reviews.

In addition, the United States Department of Education has recently called on states to become more involved, proposing changes in the teacher preparation accountability requirements in Title II of the Higher Education Act as a means of monitoring preparation programs in teacher education. The proposed rules—which may be extended to leadership preparation programs— would require states to do the following:[1]

- Use federally established criteria to report results at the individual-program level, including information from surveys of graduates and employers, candidate placement and retention rates, and growth in student achievement.

- Rate programs as "low performing," "at risk," "effective" and "exceptional."

- Report results on a state report card. Institutions would be required to post report cards on their websites.

In the following sections, we describe what each type of review entails and how often it typically occurs. The final section, "Factors Preparation Providers Consider to Be Beneficial Sources of Change," discusses selected findings from a survey conducted by the University Council for Educational Administration (UCEA) for the Wallace Foundation during the spring of 2013 that focused on the use of standards in program improvement efforts.

I. ACCREDITATION REVIEWS

Accreditation is a review process based on a set of expectations and standards. It is meant to ensure competency and credibility.

There are several levels of accreditation in higher education. There is the accreditation of the institution (i.e., the university), the school or college, and the individual program. The accreditation of an institution (referred to as "institutional accreditation") is typically conducted by a national or regional organization.[2] In the United States, colleges and universities are accredited by one of approximately 60 different accrediting bodies. These associations focus on certain types of colleges—for instance, trade and technical institutions, or religious colleges, such as seminaries and Bible colleges. Regional accreditation agencies are recognized by the U.S. Department of Education to accredit degree-granting colleges and universities (see http://www2.ed.gov/admins/finaid/accred/accreditation_pg6.html for a list of those that are recognized by the Secretary of Education). There are six regions of the United States in which regional agencies operate. The regional accreditation agencies have similar standards for accrediting colleges and universities.

---

1. By all accounts, the processes used and types of data collected for the accreditation of teacher education programs are still in a state of flux, and as Pavlakis and Kelley (2013) pointed out, this instability and energy may offer new opportunities. Pavlakis, A., & Kelley, C. (2013). *An analysis of accreditation in medicine, psychology, teacher education, engineering and law: Challenges, strengths, opportunities, and new directions*. Unpublished report written for the UCEA-Wallace Foundation Project on Leveraging Program Change.
2. The Council of Higher Education Accreditation has an extensive list of regional accrediting organizations (http://www.chea.org/Directories/regional.asp)

Accreditors are private, nongovernmental organizations created for the specific purpose of reviewing higher education institutions and programs. The purpose of accreditation is to provide an assessment of institutional quality for students, families, governmental officials, and the public at large.

A. Background on Accreditation

Accreditation in the field of teacher education was formalized in 1954 through the founding of the National Council for Accreditation of Teacher Education (NCATE), which was developed to professionalize teaching. It is still relatively new in comparison to accreditation in other professions but is experiencing many of the same growing pains as other fields have experienced.

National accreditation in education is a voluntary, peer-reviewed process that includes an evaluation of the professional education unit (the school, college, department, or body that is in charge of training teachers or other school personnel) and is based on a set of standards. The Council for the Accreditation of Educator Preparation (CAEP)[3] accredits institutions that train over 70% of America's teachers. In a sizeable number of states, accreditation through CAEP is mandated for all or some categories of institutions.

The perception that the CAEP accreditation process may not be worth the effort required has long plagued accreditation in education.[4] Half of teacher preparation programs have not sought national accreditation, and Vergari and Hess referred to the link between accreditation and educator quality as "a matter of faith."[5]

The efforts leading to the development of CAEP reignited discussions and debates "over the form and function of professional standards for educators," reflecting the "continuing lack of consensus about what makes a great teacher or school leader."[6] Teacher education standards and the accreditation process are viewed by some as too prescriptive and politically charged, and debates about inputs versus outputs—as well as lack of evidence on outcomes—have opened the doors for external groups to intervene.

A full accreditation cycle typically takes seven years (though some organizations have cycles of six, nine or 10), with reports occurring in years one, three and six and a site visit occurring in year seven. During this time, the organization under review assesses its programs and its impact on student learning and demonstrates the capacity, commitment, and competence to support high-quality student learning and ongoing school improvement.[7]

3. CAEP was formed in 2010 when NCATE merged with the Teacher Education Accreditation Council (TEAC) following competing efforts between the organziations. In 1997, TEAC created an alternative accreditation system that focused on evidence-based outcomes related to learning, validity of assessments of learning, and continuous improvement and quality (Murray 2010). Whereas NCATE's standards were externally developed, TEAC required institutions to select research-based principles that guide preparation programming and curriculum. Murray, F. (2010) Lessons from Ten Years of TEAC's Accrediting Activity. *Issues in Teacher Accreditation*. 19(1). 8-19.
4. Murray, F. (2005). On building a unified system of accreditation in teacher education. *Journal of Teacher Education*, 56(4), 30-31; Vergari. S. and F. Hess. 2002. The accreditation game. *Education Next*, 2(3): 48-57.
5. Vergari & Hess (2002, p. 57).
6. Pavlakis & Kelley (2013, p. 14).
7. An accreditation timeline and process overview can be found on page 9 of the "Guide to CAEP Accreditation" (http://www.ctc.ca.gov/educator-prep/accred-files/guide-caep-accred.pdf).

B. Key Challenges for Accreditation

In an effort to advance understandings of accreditation and its connection to program quality, Pavlakis and Kelley[8] conducted an analysis of accreditation standards and processes across five professional fields: medicine, psychology, teacher education, engineering and law. The lack of conclusive data linking accreditation to program quality or to the success of graduates is one of the key challenges facing accreditation in each of the five fields.[9] Because the evidence is often lacking, mixed, or inconclusive, in many cases it is challenging to say with certainty whether or not accreditation supports program improvement. This research gap can reduce confidence in the value of accreditation and hinder compliance. Yet Volkwein, Lattuca, Harper, and Domingo's study[10] on the impact of Engineering Criteria 2000 (the premier accrediting organization in engineering) on engineering education may be a helpful model for those interested in evaluating accreditation in other professional fields. Volkwein et al. examined the impact of the change in accreditation on a representative national sample of 203 engineering programs at 40 institutions and found that Engineering Criteria 2000 was succeeding in its quality assurance goals. For example, Volkwein et al. found that half to two-thirds of faculty reported increasing their use of more active learning methods in a course they teach regularly, that graduates had different educational experiences (such as more collaborative engagement in their learning and more interaction with faculty) than graduates a decade earlier, and that graduates reported statistically significant gains in nine areas related to accreditation goals (e.g., using modern tools, working well in teams, and applying experimental skills in analysis and interpretation).

Within the field of education, empirical research linking accreditation reviews to program improvement is scant. However, one project focused on the Educational Leadership Constituent Council's (ELCC) accreditation review process is worth mentioning. Based on a survey of educational leadership faculty, Machado and Cline found strong alignment between the content of educational leadership preparation programs and either the Interstate School Leaders Licensure Consortium (ISLLC) or ELCC standards.[11] Of the 222 survey respondents, 80% asserted that there was "moderate to substantial observable evidence of program-standards alignment."[12] The alignment process was described by many respondents as ongoing. Furthermore, 10% reported aligning their programs to leadership standards starting in 1996, when the ISLLC standards were first released; 75% reported engaging in program-standards alignment by 2003; and the remaining 15% indicated that alignment work began after 2004. Some states use the ELCC national recognition process, while others use state program review processes (that may utilize ELCC standards). CAEP is introducing a third option for states known as "program review with feedback."

II. INSTITUTIONAL REVIEWS

Internal program reviews provide a periodic and comprehensive evaluation of a university or college's academic and/or research programs and student support services to ensure continuous improvement and to aid in institutional planning and budget allocation.

The purpose of an internal program review is to direct decision making regarding the development, approval, and management of programs and services and to ensure alignment with the college's mission, core themes and identified strategic objectives. To achieve these goals, internal program review procedures encourage self-study and purposeful planning within programs and services. In addition, an essential element of the internal program review is the identification and evaluation of student learning outcomes as key indicators of effectiveness.

---

8. Pavlakis & Kelley (2013).
9. Ibid.
10. Volkwein, J. F., Lattuca, L. R., Harper, B. J., & Domingo, R. J. (2007). Measuring the Impact of Professional Accreditation on Student Experiences and Learning Outcomes. *Research in Higher Education*, 48 (2), 251-282.
11. Machado, C., & Cline, D. (2010). Faculty perceptions: Where do educational administration programs stand with the ISLLC/ELCC standards? *NCPEA Educational Leadership Review*, 11(1), 10–17.
12. Ibid, p. 12.

Many institutions (though not all) conduct internal reviews in conjunction with professional disciplinary or specialized accreditation (or reaccreditation) review cycles. University administrators keep lists of the different accrediting organizations that review programs and colleges within their institutions and generally organize the institutional review to coincide with the accreditation cycle. In some states (e.g., Rhode Island), the state office of higher education maintains a similar listing and requires that institutions forward copies of program self-studies and accrediting team reports according to the accreditation schedule.

Those programs that are not reviewed by national accreditation agencies undergo a program review initiated by the institution. Within a cyclical time frame of three to seven years, each institution determines a specific schedule for program reviews. Institutional reviews can also be triggered early, such as when the number of students enrolled in a given program falls below a certain number for several years.

The self-study developed for professional or specialized accreditation reviews normally provides the essential requirements of internal program review; however, if there is a difference in the requirements for the internal reviews, the entity under review currently must meet the requirements of each reviewer.

The basic components of a typical internal program review process include the following:

1. a self-study, recommendations, and preliminary implementation plan completed by the faculty or staff associated with the program;

2. review and recommendations by the peer program review committee (depending on the type of institution, peers may be from the same college, from a different college within the university, or from a peer institution);

3. revision of the self-study, recommendations, and preliminary implementation plan in response to the peer program review committee;

4. presentation of the final draft of the internal program review to the college council;

5. final approval by the appropriate senior administrator (chief academic officer, chief student affairs officer, chief fiscal officer) of all elements of the internal program review documents; and

6. implementation of actions to improve program effectiveness and quality as needed.

III. STATE REVIEWS

There are several entities at the state level that review colleges and schools of education. The state office of higher education generally has review requirements that may or may not align with those of an external accrediting organization. Generally, states have a set of specific data points that they collect from universities that focus on the students enrolled, financial aid provided, resources available (e.g., libraries), class sizes, faculty characteristics, teaching assignments, incentives for promoting quality, research dollars, service, graduation rates, retention rates, graduate job placement, and alumni satisfaction. Collecting this information from all colleges at the same time enables state personnel to develop comparison reports.

State departments of PK-12 education are also involved in reviewing colleges and schools of education, as well as other organizations and entities that seek to provide certification programs for educational professionals. With regard to educational leadership, the introduction of the ISLLC standards in 1996 prompted some states to adopt program review and redesign initiatives that involved alignment to state or national leadership standards and the adoption of specific program features, such as expanded field experience requirements.[13]

---

13. Baker, B. D., Orr, M. T., & Young, M. D. (2007). Academic drift, institutional production, and professional distribution of graduate degrees in educational leadership. *Educational Administration Quarterly*, 43(3), 279–318. doi:10.1177/0013161X07303320

Some states have developed relationships with CAEP. These relationships can take several forms, such as delegating the review of programs to CAEP and accepting CAEP reviews in place of state reviews. In some cases, a state will partner with CAEP to conduct the reviews. In such cases, state department personnel will identify a faculty member and practitioner from the state to collaborate with CAEP volunteers in the review of materials and on the site visit. However, some states require that a college be accredited and will reserve the right to conduct program reviews.

In some states (e.g., North Carolina, Iowa, and Kentucky), low-quality programs that were unable to meet these new requirements were eliminated as the result of a state review. These reviews usually involve some kind of expert or critical friend review process (see description of critical friend reviews in section IV). The existing research on the effects of such processes on program change indicates that state policies and strategies intended to promote redesign of principal preparation programs have produced episodic change in a few institutions but have fallen short of expectations.[14] Furthermore, research concerning the impact of such processes on program graduates is inconclusive.[15]

A. Key Challenges for State Reviews

More research is needed to gain a robust understanding of the processes used in state reviews and their impact on principal preparation programs and candidates in those programs. Two issues in particular may impact the effectiveness of such processes. First, it is questionable whether state departments of education have the capacity to translate and implement policy in addition to supporting preparation program redesign, particularly in times of financial cutbacks.[16] The second issue concerns the process itself. A one-size-fits-all redesign process that does not take into account different institutional types, missions, and capacities is unlikely to yield desired results.[17]

A 2006 report from the Southern Regional Education Board[18] concluded that state policies and strategies intended to promote the redesign of principal preparation programs have improved some programs but have not led to the sort of deep change needed to ensure all candidates master the knowledge and skills needed to be effective school leaders. Murphy et al. had similar findings, stating, "The results of reform are uneven and fall short of the mark."[19]

State reform initiatives often utilize a one-size-fits-all approach that does not take into account the context or capacity of the university preparation program.[20] Research universities and regional colleges, regardless of size or mission, are generally required to use the same framework for reform; yet faculty roles, resources available, and audiences served at these institutions are often quite different. It is important to note that context does matter when implementing change.

Further, if all university-based preparation programs in a state are required to participate in reforms, questions arise about the intentions and capacity of the state department of education as well as the intended impact of those reforms. Particularly in times of financial cutbacks, many raise the question: Do state departments of education have the capacity to translate and implement policy and support preparation program redesign?[21] Murphy et al. cautioned that without adequate attention to both the technical aspects of reform (e.g., staffing) and the adaptive aspects of reform (e.g., changes in core values and beliefs), change will be superficial. The magnitude of the resources (e.g., faculty time and funding) and institutional changes (e.g., changes in admissions requirements and processes) needed to support this type of work are significant. Without adequate institutional and state-level capacity to support redesign, implementation, and monitoring of program quality after implementation, most reform efforts will fall short of expectations.

---

14. Murphy, J., Moorman, H. N., & McCarthy, M. (2008). A framework for rebuilding initial certification and preparation programs in educational leadership: Lessons from whole-state reform initiatives. *Teachers College Record*, 110(10), 2172–2203; Southern Regional Education Board. (2006). *Challenge to lead: The momentum continues.* Atlanta, GA: Author.

15. McCarthy, M. M., & Forsyth, P. B. (2009). An historical review of research and development activities pertaining to the preparation of school leaders. In M. D. Young, G. M. Crow, J. Murphy, & R. T. Ogawa (Eds.), *Handbook of research on the education of school leaders* (pp. 86–128). New York, NY: Routledge.

16. Young, M. D. (2013). Is state-mandated redesign an effective and sustainable solution? *Journal of Research on Leadership Education*, 8(2), 247–254.

17. Ibid.

18. SREB (2006).

19. Murphy et al. (2008, p. 2186).

20. Phillips (2013); Young (2013); Young, M. D., & Brewer, C. (2008). Fear and the preparation of school leaders: The role of ambiguity, anxiety and power in meaning making. *Education Policy*, 22(1), 106–129.

21. Young (2013).

IV. PROFESSIONAL ORGANIZATION REVIEWS

A. UCEA Membership Reviews

Reform efforts have prompted an increased focus from within the profession on how to improve the quality of leadership preparation. Organizations such as UCEA have a long history related to the development and dissemination of pertinent research and tools that have influenced leadership preparation within and beyond UCEA institutions. UCEA has long invested in the development of instructional materials, from simulations to cases to course modules, and UCEA-sponsored research projects have raised critical questions about the conditions and quality of leadership development and practice. Additionally, through its sponsorship of the National Commission on Excellence in Educational Administration, UCEA was instrumental in the creation in 1988 of the National Policy Board for Educational Administration (NPBEA), which has undertaken a series of important activities, including the development of ISLLC and the ELCC standards. NPBEA, in conjunction with the Danforth Foundation, sponsored national conferences focused on innovative preparation practices to help spread promising practices across the nation. UCEA worked with NPBEA in 2001 to sponsor the National Commission for the Advancement of Educational Leadership Preparation, which examined high-quality leadership preparation and professional development programs, their impact, and the contextual factors that influenced the success and impact of programs.

Membership in UCEA requires a rigorous, multistage review and renewal process—a process that carefully examines the quality of an institution's preparation and research programs. The process begins with the development of an application portfolio. Decisions on membership are made based on three categories of evidence: (a) eligibility, including consistency with UCEA's standards of excellence; (b) consistency with UCEA's Institution and Program Quality Standards; and (c) other supporting evidence. A set of rubrics and suggested sources of evidence are provided in the UCEA publication "UCEA Institutional and Program Quality Criteria: Guidance for Master's and Doctoral Programs in Educational Leadership."[22]

After receiving an application portfolio, UCEA Executive Committee members thoroughly review the full set of application materials using the UCEA quality criteria rubric, and they provide an overview of the applicant's strengths and weaknesses and note the absence of data necessary for rating the institution on one or more criteria. Depending on the strength of the application and the availability of data, the Executive Committee will make a recommendation to gather additional information, send a site-visit team to the institution, or let the institution know that its programs are ineligible for membership.

The site visitation is conducted by a team of two UCEA faculty with expertise in educational leadership development. The visit usually takes two full days and involves a combination of interviews with faculty, students, alumni, district partners, and institutional leadership; classroom observations; and a review of relevant program documents and evidence. Based on data from these sources, the visitation team submits a site-visitation report to the executive committee, which then makes a decision about recommending the institution to the broader membership for consideration, either as a full or provisional member. Provisional membership is recommended for those institutions that are close to meeting membership criteria but still need to improve certain aspects of their program before full membership can be offered.

Representatives of UCEA member institutions are provided access to the applying institution's membership application portfolio and site-visit report. They are provided a period of 30 days to review the applicant's materials and the site-visit report, after which time the institution's membership is put to a vote. At any point in this process, members may raise questions, request clarification, or make reasonable requests for additional evidence.

---

22. Young, M. D., Orr, M. T., & Tucker, P. D. (2012). *UCEA institutional and program quality criteria: Guidance for master's and doctoral programs in educational leadership*. Charlottesville, VA: UCEA.

Continuation of membership in UCEA involves periodic self-study. The purpose of the self-study is to provide each member institution opportunities to (a) review its commitment to improving its programs in educational administration, (b) assess its progress in the attainment of program goals, (c) exhibit its unique program qualities and strengths, and (d) describe its future program goals and opportunities. The self-study is facilitated by two UCEA resources: (a) "Developing Evaluation Evidence: A Formative and Summative Evaluation Planner for Educational Leadership Preparation Programs," and (b) the UCEA's Initiative for Systemic Program Improvement through Research in Educational Leadership (INSPIRE) survey suite. The evaluation planner is aligned to both the ELCC and UCEA standards and facilitates planning and data collection around preparation program evaluation and improvement. The planner includes a logic model, suggested sources of evidence to collect, worksheets, and key questions to drive program evaluation. The INSPIRE suite of surveys is also aligned with the ELCC and UCEA standards. It provides 360° data on the features, quality and impact of educational leadership preparation programs (including candidate, practicing leader, teacher, supervisor and program perspectives), and it reflects the key elements of the evaluation planner logic model regarding how preparation impacts leadership, school conditions, and student learning.

UCEA's program review is considered by members to be an important and impactful self-assessment process. Like the critical friends review used by Murphy et al., UCEA's program review process involves both program self-assessments and external review and feedback, followed by recommendations and technical assistance. The feedback from these reviews addresses both technical and adaptive elements influencing program quality.[23] These processes resemble in some ways the accreditation review processes used by CAEP and TEAC in that they are standards- and evidence-based, yet they take the review one step further by providing programs with actionable feedback and advice.

B. Quality Measures

The Quality Measures (QM) review and improvement process resembles UCEA's approach as well. The QM standards or "indicators of quality" were made publicly available in 2009 by the Education Development Center.[24] The indicators reflect research on the features, both in terms of content and clinical practices, associated with exemplary principal preparation programs and evolved through work with Wallace Foundation–funded districts, universities, and states to assess the quality of their principal preparation programs.[25]

This 2009 version of QM focuses on two program features in particular, Program Course Content and Pedagogy and Program Clinical Practice. The five course-content indicators focus on the content domains of the 2008 ISLLC standards. For example, it is recommended that content "be logically and sequentially organized and aligned with state professional standards and school district performance expectations."[26] Furthermore, the Education Development Center asserted that program pedagogy should include "problem-based learning strategies designed to incorporate real school contexts and make extensive use of formative and summative assessments."[27] The five clinical practice indicators characterize a high-quality experience as one that is full-time, yearlong, carefully sequenced, and organized around opportunities to practice leadership skills in a real-world setting. As Darling-Hammond and colleagues describe, "Like other program coursework, the clinical practice should provide formal formative and summative assessments and offer interns the opportunity to develop competencies in more than one context."[28]

23. Murphy et al. (2008).
24. King, C. (2014). *Partnership effectiveness continuum*. Waltham, MA: Education Development Center, Inc.
25. Darling-Hammond, L., LaPointe, M., Meyerson, D., & Orr, M. T. (2007). *Preparing school leaders for a changing world: Executive summary*. Stanford, CA: Stanford University, Stanford Educational Leadership Institute.
26. King, C. (2013). Quality Measures Principal Preparation Program Self-Assessment Toolkit: for use in developing, assessing, and improving principal preparation programs. Waltham, MA: Education Development Center, Inc.
27. Ibid, p. 2.
28. Ibid, p. 3.

The QM document *Principal Preparation Program Quality Self-Assessment Rubrics* asserts, "An effective self-assessment of principal preparation program quality requires a clear understanding by all participants in the process of what is meant by 'quality.'"[29] Thus, the QM designers worked to develop user-friendly tools as well as a process that decreased ambiguity about what program elements were to be assessed and what would be considered acceptable evidence. QM places particular emphasis on the review and consideration of "supporting evidence" in determining the degree to which programs reflect research-based indicators of quality. Specifically, tools (e.g., rubrics and assessments) were developed to facilitate program self-assessment and to build consensus around the features and attributes of high-quality programs. A hands-on program review involving program faculty and external consultants in the review of program artifacts, data, and faculty work is portrayed as invaluable and critical to the revision process. It is argued that "these tools and processes, when used together, will provide improved guidance to program self-assessment team efforts to more accurately determine the quality of their principal preparation programs."[30]

They also calibrate the indicators of quality along a developmental scale (well-developed, developed, emerging, and beginning). QM has been used primarily by Wallace-funded principal preparation programs. Programs have used the QM rubrics, along with a handbook containing guidelines for selecting and analyzing credible evidence, to structure their self-assessment of core features. Self-assessments enable program teams to determine where they fall on a developmental scale and then use the results of the analysis to plan improvements in the quality of their programs.

C. Self-Study Models

The UCEA and QM processes both involve a self-study component through which programs gather program information and candidate data to determine (a) the degree to which programs reflect the UCEA or QM criteria and (b) the impact of the program's content and experiences on candidates' growth, career outcomes, and leadership performance. As described above, UCEA and QM offer tools (e.g., rubrics) to facilitate program self-assessments and to build consensus around the features and attributes of high-quality programs. The tools reflect the current research and lessons learned about principal preparation program quality.

The new suite of instruments to evaluate preparation programs available through UCEA enables faculty to dig more deeply into questions of how preparation programs impact the knowledge and practice of their graduates. The INSPIRE survey suite includes a program features survey, a candidate survey, and a 360° survey. Together, the surveys provide data that help programs assess the quality and impact of various program features and content areas. When used in conjunction with the "Developing Evaluation Evidence" program evaluation planner, the surveys help guide program faculty through evaluation design, data collection, analysis and improvement cycles.[31] Given that most programs are guided by their own theory of action or program theory, which connects choices in program content, delivery, and design to expected outcomes, the surveys and planner support a variety of program designs and their unique features.

D. Critical Friend Reviews

Critical friend reviews reflect the processes used by UCEA and QM and the process described by Murphy et al. They generally involve program self-assessments; external review of program documents, data and artifacts; feedback to program faculty, including recommendations; and, in some cases, technical assistance. The QM handbook asserts, "These tools and processes, when used together, will provide improved guidance to program self-assessment team efforts to more accurately determine the quality of their principal preparation programs."[32] These processes resemble in some ways the accreditation review processes used by NCATE and TEAC in that they are standards- and evidence-based, but they take the review one step further by providing programs with actionable feedback and advice.

---

29. Ibid, p. 4.
30. Ibid, p. 2.
31. Orr, M. T., Young, M. D., Rorrer, A. K. (2010). *Developing evaluation evidence: A formative and summative evaluation planner for educational leadership preparation programs*. Charlottesville, VA: Center for the Evaluation of Educational Leadership Preparation and Practice.
32. King (2013).

V. FACTORS PREPARATION PROVIDERS CONSIDER TO BE BENEFICIAL SOURCES OF CHANGE

Below, we highlight several findings from a survey of directors of educational leadership preparation programs that is focused on program improvement levers. According to these program leaders, the most beneficial levers for program improvement include (in the following order of importance): state licensure and program requirements, national or regional accreditation review, and professional association projects or reviews.

Importantly, program directors were asked to identify factors that were most influential as well as those that were most beneficial. These two terms were differentiated in the following way: *Influential* was defined as having an impact, changing behavior or ensuring compliance, whereas *beneficial* was defined as supporting substantive, research-based, positive and sustainable change.

A. State Program Accreditation and Approval

When asked to identify which reform and program improvement factors were most beneficial to their program improvement efforts, program directors were most likely to point to state licensure and other state requirements, as shown in Table 1 below.[34] Directors described several ways that state requirements benefited their programs. Comments ranged from the state's influence on the focus of the program, such as, "State licensure requirements have required us to focus our efforts on what appears to be most important in our state," to comments such as, "The state requirements were well done and therefore were instructive for the efforts of our redesigned program."

Table 1: Percentage of Program Directors Who Rated Six Sources of Pressure in the Program Design Process as First- or Second-Most Influential and Beneficial

Source	First- or second-most influential	First- or second-most beneficial
National or regional accreditation review	78	36
State program requirements	52	41
State licensure requirements	41	65
Institutional requirements and demands	14	23
Professional association projects or reviews	11	33
Federal, state, or foundation funding requirements	4	4

Not all comments, however, highlighted the beneficial aspects of state requirements. Rather, some comments emphasized the ways in which state requirements were "not always viewed as productive by our program faculty." One director shared, "New state requirements were rigorous and helpful but went too far in micromanaging institutional programs." Another noted, "The state requirements are a double-edged sword in that they are required, but we do it well so students get exactly what they need. The institution sometimes creates obstacles for creative ways to implement programs." However, a few directors offered a more balanced perspective. One commented, "While at times we find state requirements onerous, we believe that in our state many of them have merit and provide us enough latitude to differentiate programming to meet our students' needs."

---

33. American Association of Colleges for Teacher Education (2013).
34. Young, M. D., Tucker, P. D., Mawhinney, H., & Reed, C. J. (2013). *Leveraging what works in preparing educational leaders*. Charlottesville, VA: University Council for Educational Administration.

New Leaders

B. National Accreditation

Program directors' comments regarding the beneficial effects of program accreditation reflected both the challenges and opportunities of such work. For example, several directors highlighted the opportunities provided by NCATE to "review our program carefully." One shared that a "recent change of accrediting body for the institution brought an opportunity to examine the program through a slightly different lens. The changes we made to meet the accreditation standards [TEAC] have been very beneficial." Another shared, "While NCATE, CAEP and ELCC has impacted us tremendously, it has also helped us grow and develop as a department to deliver the best program that we can to our students." Reflecting a more mixed perspective, one director explained, "We are required to be accredited by ELCC/NCATE. The programs and policies put in place to address those requirements dictate everything else so we were forced to be on the same page as instructors and that did help."

C. Professional Association Projects or Reviews

Interestingly, program directors identified professional association projects and reviews as more beneficial than influential, while the reverse was true for national and regional accreditation reviews. As some program directors explained, "Professional association reports and reviews are much more focused on the realities of administrator preparation," and, "Our faculty most value professional association change forces, as we believe them to be most informed by rigorous scholarship and praxis." Several mentioned UCEA specifically. Directors commented that "UCEA has set the best standard, which was largely embraced by [the state]" and that "UCEA is most beneficial in terms of determining what's most beneficial."

D. Other

Program directors tended to identify research on effective preparation as more beneficial than influential. In fact, several cited research as a driver of program improvement. One program director remarked, "Program improvement decisions are driven by research." Another stated, "My awareness about research and connection to what works helps as I think about program design so I ranked professional association connections first. The rest are more about meeting state or other requirements which we usually meet easily." Engaging in inquiry on their own program was also considered beneficial; one program director commented, "Our redesign was informed by a formal cycle of inquiry and data collection (both internal and external). This process had the most influence on our program's current design and delivery."

A final source considered beneficial for program change involved local needs and priorities. One director shared, "We need to consider what our aspiring school leaders want/need in the field in order to serve all children and families." Similarly, another explained, "The most beneficial element of the change process over the last five years has been that the requirement of partnerships has strengthened both the K-12/university partnerships and the university system partnerships." Program directors highlighted feedback from alumni as important as well. One director shared,

> The most beneficial forces have been the feedback provided by our alumni and by the students who take our program's courses. That's what matters most to us. The other factors listed actually distract and derail us from focusing on what matters most to us.

Finally, survey respondent comments revealed two factors considered detrimental to program redesign and improvement. The first factor dealt with institutional barriers. For example, one respondent noted, "The institution ranks low [as a beneficial influence] because they pose tuition barriers for effective recruitment/delivery." The second factor was increased competition. A director explained,

> We have to compete with other entities locally and across the state that certify school administrators and at the same time prepare students to pass the certification exam. We have tried to revise our program to meet the state standards and at the same time be able to stay competitive.

VI. FINAL THOUGHTS

Since the introduction of the ISLLC standards in 1996, a number of states have adopted program review and redesign initiatives that involve alignment to state or national leadership standards and the adoption of specific program features (e.g., expanded field experience requirements).[35] In some states, low-quality programs that were unable to meet these new requirements were eliminated. What research has been conducted on the effects of such processes on program change indicates that state policies and strategies intended to promote the redesign of principal preparation programs have produced episodic change in a few institutions but have generally fallen short of expectations.[36] Furthermore, research concerning the impact of such processes on program graduates is inconclusive.[37]

Although more research on such processes would be helpful to gain a robust understanding of the processes used and their impact on programs and candidates, two issues may impact the effectiveness of such processes: (a) the capacity of state departments of education to both translate and implement policy and support preparation program redesign, particularly in times of financial cutbacks; and (b) the need for clear guidance in taking up the evaluation role—in particular, the need to avoid a one-size-fits-all redesign process that does not take into account different institutional types, missions and capacities.

---

35. Baker et al. (2007).
36. Murphy et al. (2008); SREB (2006).
37. McCarthy & Forsyth (2009).

NewLeaders

University Council for Educational Administration
Quality Leadership Matters   since 1954

Resource C:
Relevant Publications

I. IN-DEPTH REVIEW PROCESS RESOURCES

Title: Principal Preparation Program Self-Assessment Toolkit: For Use in Developing, Assessing, and Improving Principal Preparation Programs (with accompanying handbook)
Source: Education Development Center
Publication Year: 2013
Link: Wallace Foundation
Description: Designed to help principal preparation programs assess the quality of the training offered, this kit includes "rubrics" (or ratings with descriptions of what they mean) on course content, candidate recruitment and other matters. A separate handbook is designed to guide people through effective use of the tool.

Title: Developing Evaluation Evidence: A Formative and Summative Evaluation Planner for Educational Leadership Preparation Programs
Source: University Council for Educational Administration (UCEA)
Publication Year: 2012
Link: UCEA
Description: To facilitate planning and data collection, this planner is organized by program inputs and outcomes. The planner includes a conceptual model of the link between leadership preparation and outcomes, a guide for identifying evaluation evidence, and an evaluation planning worksheet.

Title: Institutional and Program Quality Criteria: Guidance for Master's and Doctoral Programs in Educational Leadership
Source: UCEA
Publication Year: 2012
Link: UCEA
Description: This guidebook for masters and doctoral programs in educational leadership includes the UCEA Institutional and Program Quality criteria, rubrics that illuminate the difference between very effective, effective and developing practices concerning each criteria as well as suggestions for how to use the rubrics to facilitate conversations around program improvement.

Title: Designing Purposeful and Coherent Leadership Preparation Curriculum: A Curriculum Mapping Guide
Source: UCEA
Publication Year: 2012
Link: UCEA
Description: This guide provides a process and set of tools to help preparation program faculty articulate and align leadership expectations and their program's content. The resource includes worksheets to collect curricular information and analyze courses' standards alignment, content coherence, and relevance to program goals and priorities. The guide can be used to help develop new programs or to evaluate an existing program for renewal and revision.

Title: 2014 Interstate School Leaders Licensure Consortium (ISLLC) Standards
Source: Council of Chief State School Officers (CCSSO)
Publication Year: 2014
Link: UCEA
Description: Produced by the CCSSO and the National Policy Board for Educational Administration, these standards are model leadership protocols that outline what education leaders should know and be able to do to ensure that all students graduating from high school are prepared to enter college or the modern workforce. The standards outline foundational principles of education leadership that cut across grade levels and help improve student achievement and engagement.

Title: Council for the Accreditation of Educator Preparation 2013 Standards for Accreditation of Educator Preparation (and Evidence Guide)
Source: Council for the Accreditation of Educator Preparation (CAEP)
Publication Year: 2013
Link: CAEP
Description: CAEP adopted new accreditation standards for education preparation programs in August 2013. To support preparation programs in meeting the new standards, the council released an evidence guide that explains CAEP's perspective on building a "culture of evidence" through the use of data. The guide includes instructions for collecting and analyzing data, as well as guidance on what constitutes valid evidence and suggestions for collecting evidence on the impact of program graduates on student learning.

Title: Evaluation of Teacher Preparation Programs: Purposes, Methods, and Policy Options
Source: National Academy of Education (NAEd)
Publication Year: 2013
Link: National Academy of Education
Description: NAEd's stated objective for this report was "to provide clearer information and direction around evaluation measures and systems in educator preparation." Many aspects of the relationship between teacher preparation and instructional quality are not fully understood, and existing approaches to teacher preparation program evaluation are complex, varied and fragmented. Designers and consumers of teacher preparation program evaluations could benefit from clear information about the purposes, effects, strengths and limitations of current evaluation approaches and from guidance for designing and using future evaluations. This report, the product of an analysis by a committee of the National Academy of Education, aims to fill that need.

II. THE STATE ROLE IN EVALUATING PRINCIPAL PREPARATION PROGRAMS

Title: Change Agents: How States Can Develop Effective School Leaders (Concept Paper and Guide)
Source: New Leaders
Publication Year: 2013
Link: New Leaders
Description: This concept paper is a primer for states that want to build a pipeline of highly effective principals by reforming the way they hold preparation programs accountable for results and improving their licensure systems. It lays out current challenges in the field, describes a case for outcomes-focused school leadership, and makes policy recommendations for achieving that vision. The concept paper is accompanied by an important resource, the "Change Agents Companion Guide," which serves as a framework state policymakers can use to explore fundamental questions of purpose and design based on a new vision for leadership and an assessment of tools at their disposal.

Title: The State of State Policies for Principal Preparation Program Approval and Candidate Licensure
Source: The Journal of Research on Leadership Education/Anderson and Reynolds
Publication Year: 2015 (forthcoming)
Link: Not yet available.
Description: Using a policy analysis framework by Roach and colleagues, this study explores state codes, administrative rules and regulations, and accompanying state board/department of education documents to describe state policies for principal preparation program approval and candidate licensure.

Title: Improving School Leader Preparation: Collaborative Models for Measuring Effectiveness
Source: American Institutes for Research
Publication Year: 2014
Link: Center on Great Teachers and Leaders at American Institutes for Research
Description: This research brief provides an overview of how states currently measure the effectiveness of school leader preparation programs and explores new collaborative models for continuous program improvement.

Title: Our Responsibility, Our Promise: Transforming Educator Preparation and Entry into the Profession
Source: CCSSO Task Force on Educator Preparation and Entry into the Profession
Publication Year: 2012
Link: CCSSO
Description: This report was written by the Task Force on Educator Preparation and Entry into the Profession, which is composed of current and former chief state school officers who are members of the CCSSO, with input from the National Association of State Boards of Education and the National Governors Association. The recommendations contained in this report focus on the levers for change that are the responsibility of state education agencies and, where applicable, their partner professional standards boards: licensure; program approval; and data collection, analysis, and reporting. Twenty-five states have agreed to advance the recommendations in the report.

III. ANNUAL REPORT MEASURES

Title: Initiative for Systemic Program Improvement through Research in Educational Leadership (INSPIRE) Surveys
Source: UCEA
Publication Year: 2013
Link: UCEA
Description: The INSPIRE Surveys include a suite of evaluation resources made available by the UCEA Center for the Evaluation of Educational Leadership Preparation and Practice for leadership preparation programs to produce evidence helpful in improving programs, meeting accreditation requirements, and making the case for support among various constituencies. The INSPIRE Leadership Suite currently includes:

- Preparation Program Edition (INSPIRE-PP)—This instrument enables the educational leadership preparation program to systematically document its core program features.

- Graduate Edition (INSPIRE-G)—The specific purpose of this survey is to elicit feedback from alumni on their leadership preparation experiences and learning and career outcomes.

- Leader in Practice Edition (INSPIRE-LP)—This instrument enables the educational leadership preparation programs to document leadership practices and school improvement and organizational indicators from the perspective of program graduates who are working as school principals.

- 360 Edition (INSPIRE-360)—This instrument enables the educational leadership preparation program to document leadership practices and school improvement and organizational indicators in the schools where program graduates work from the perspective of teachers.

Title: Comprehensive Assessment of Leadership for Learning (CALL)
Source: Wisconsin Center for Educational Research
Publication Year: 2012
Link: Comprehensive Assessment of Leadership for Learning
Description: CALL captures current leadership practices in five domains: focus on learning, monitoring teaching and learning, building nested learning communities, acquiring and allocating resources, and maintaining a safe and effective learning environment. The task-based focus of the survey emphasizes the work that needs to be done to improve student learning rather than the disposition or character of leaders. CALL gathers data from principals, school staff and teachers and is intended for middle schools and high schools.

Title: Vanderbilt Assessment of Leadership in Education
Publication Year: 2012
Link: Discovery Education
Description: The Vanderbilt Assessment of Leadership in Education is a research-based evaluation tool that measures the effectiveness of school leaders by providing a detailed assessment of a principal's behaviors. VAL-ED focuses on the skills and behaviors unique to the role and career of a principal, providing evidence that the appropriate and necessary instructional leadership behaviors are exhibited at the school.

IV. FOR FURTHER READING

American Educational Research Association & National Academy of Education. (2011). *Getting teacher evaluation right: A brief for policymakers*. Stanford, CA: Stanford Center for Opportunity Policy in Education.

Anderson, E., & Reynolds, A. (accepted). State of state policies for principal preparation program approval and candidate licensure. *Journal of Research in Educational Administration*.

Baker, B. D., Orr, M. T., & Young, M. D. (2007). Academic drift, institutional production, and professional distribution of graduate degrees in educational leadership. *Educational Administration Quarterly*, 43(3), 279–318. doi:10.1177/0013161X07303320

Branch, G. F., Hanushek, E. A., & Rivken, S. (2012). *Estimating the effect of leaders on public sector productivity: The case of school principals* (NBER Working Paper No. 17803). Cambridge, MA: National Bureau of Economic Research.

Briggs, K., Cheney, G. R., Davis, J., & Moll, K. (2013). *Operating in the dark: What outdated state policies and data gaps mean for effective school leadership*. Dallas, TX: The George W. Bush Institute.

Burkhauser, S., Gates, S. M., Hamilton, L. S., & Ikemoto, G. S. (2012). *First-year principals in urban school districts: How actions and working conditions relate to outcomes*. Santa Monica, CA: RAND.

Coelli, M., & Green, D. A. (2012). Leadership effects: School principals and student outcomes. *Economics of Education Review*, 31(1), 92–109.

Ehlert, M., Koedel, C., Parsons, E., & Podgurskey, M. (2013). *Selecting growth measures for school and teacher evaluations: Should proportionality matter?* (CALDER Working Paper No. 80). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Feuer, M. J., Floden, R. E., Chudowsky, N., & Ahn, J. (2013). *Evaluation of teacher preparation programs: Purposes, methods, and policy options*. Washington, DC: National Academy of Education.

Fuller, E. J., & Hollingworth, L. (2013). A bridge too far? Challenges in evaluating principal effectiveness. *Educational Administration Quarterly*, 50(3), 466–499.

Grissom, J. A., Kalogrides, D., & Loeb, S. (2014). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, 37(1), 3–28.

Machado, C., & Cline, D. (2010). Faculty perceptions: Where do educational administration programs stand with the ISLLC/ELCC standards? *NCPEA Educational Leadership Review*, 11(1), 10–17.

McCarthy, M. M., & Forsyth, P. B. (2009). An historical review of research and development activities pertaining to the preparation of school leaders. In M. D. Young, G. M. Crow, J. Murphy, & R. T. Ogawa (Eds.), *Handbook of research on the education of school leaders* (pp. 86–128). New York, NY: Routledge.

Murphy, J., Moorman, H. N., & McCarthy, M. (2008). A framework for rebuilding initial certification and preparation programs in educational leadership: Lessons from whole-state reform initiatives. *Teachers College Record*, 110(10), 2172–2203.

Orr, M. T., & Barber, M. E. (2009). Program evaluation in leadership preparation and related fields. In M. D. Young, G. M. Crow, J. Murphy, & R. T. Ogawa (Eds.), *Handbook of research on the education of school leaders* (pp. 457–498). New York, NY: Routledge.

New Leaders

University Council for Educational Administration
Quality Leadership Matters *since 1954*

Pavlakis, A., & Kelley, C. (2013). *An analysis of accreditation in medicine, psychology, teacher education, engineering and law: Challenges, strengths, opportunities, and new directions*. Unpublished report written for the UCEA-Wallace Foundation Project on Leveraging Program Change.

Southern Regional Education Board. (2006). *Challenge to lead: The momentum continues*. Atlanta, GA: Author.

Young, M. D. (2013). Is state-mandated redesign an effective and sustainable solution? *Journal of Research on Leadership Education*, 8(2), 247–254.

Young, M. D., & Brewer, C. (2008). Fear and the preparation of school leaders: The role of ambiguity, anxiety and power in meaning making. *Education Policy*, 22(1), 106–129.

Young, M. D., Crow, G. M., Murphy, J., & Ogawa, R. T. (Eds.). (2009). *Handbook of research on the education of school leaders*. New York, NY: Routledge.

Young, M. D., Tucker, P. D., Mawhinney, H., & Reed, C. J. (2013). *Leveraging what works in preparing educational leaders*. Charlottesville, VA: University Council for Education Administration.