

Principal Talent Management According to the Evidence: **A Review of the Literature**

OCTOBER 2016



GEORGE W. BUSH
INSTITUTE



GEORGE W. BUSH INSTITUTE

Housed within the George W. Bush Presidential Center, the George W. Bush Institute is an action-oriented, nonpartisan policy organization with the mission of cultivating leaders, advancing policy, and taking action to solve today's most pressing challenges. The work is achieved through three Impact Centers – Domestic Excellence, Global Leadership, and our Engagement Agenda – by which the Bush Institute engages the greater community with its important work.

Rooted in President and Mrs. Bush's belief that "excellent schools must first have excellent leaders," the George W. Bush Institute developed the Alliance to Reform Education Leadership (AREL) to transform districts' talent management of school principals and to provide school districts with knowledge and tools to attract and retain effective principals. AREL's mission is to help ensure there is an effective principal – able to significantly advance student achievement – at the helm of every school.

To learn more about our work, visit www.bushcenter.org.

We are grateful for the continued generous support of AT&T Foundation, Bass Foundation, CME Group Foundation, Rainwater Charitable Foundation, Sid W. Richardson Foundation, The Morris Foundation, and The Prudential Foundation.

American Institutes for Research was contracted to support the George W. Bush Institute's research on education leadership.



George W. Bush Institute Project Team:

Eva Chiang, Deputy Director of Education Reform and Evaluation and Research; Dr. Catherine Jaynes, Director of Evaluation and Research; Anne Wicks Humphrey, Director of Education Reform; Abby Hoak-Morton, Associate of Education Reform.

American Institutes for Research Project Team:

Dr. Matthew Clifford, Principal Investigator; Dana Chambers, Project Director; Mariann Lemke, Senior Advisor; Dr. Kwang Suk Yoon, Lead WWC Reviewer; Dr. Yinmei Wan, WWC Reviewer

We thank our partners at Collaborative Communications who helped us improve this report. While their work greatly shaped this final document, we take full responsibility for all errors and omissions.

Contents

2	Executive Summary
4	Introduction
7	Approach to the Review
10	Summary of Findings
12	Detailed Findings of WWC Literature Review
12	Principal Working Environment
14	Principal Preparation
18	Principal Professional Learning
21	Principal Compensation and Incentives
24	Conclusion and Recommendations
26	Endnotes
27	Appendix A. Proposed Search Terms
28	Appendix B. Summary of Studies That Were Eligible for Review but Did Not Meet Standards (With or Without Reservations)
30	Appendix C. Citations Marked Ineligible for Review Based on Initial Screening
33	Appendix D. Summary of Results
35	Appendix E. WWC Review Protocol
47	Appendix F. Detailed Findings of WWC Review
59	Appendix G. Comparison of how Every Student Success Act (ESSA) defines “evidence-based” activity, strategy, or intervention; Supplementary explanation of how What Works Clearinghouse (WWC) characterizes “credibility of evidence.”
60	Complete List of All Studies Reviewed for the Initial Eligibility Screening

Executive Summary

In recognition of the importance of effective school leadership, many states and districts have implemented policies and programs that aim to improve their capacity to attract and retain great principals. Too often, though, these policies and practices do not combine to form a coherent system that ensures effective leadership development.

To be successful, these efforts require a comprehensive, systematic approach to principal talent management (PTM) that encompasses the entire continuum of a principal's career: **preparation, recruitment and selection, professional learning, performance evaluation, and compensation and incentives**. A systematic approach to PTM must also account for the powerful influence of the **principal working environment**. The working environment impacts all of the aforementioned components of the PTM system and can affect principals across a whole range of their professional experiences—from their in-district residencies to curricular and budgetary decision-making processes.

This literature review aims to provide district leaders with an understanding of the research and best evidence regarding the components of effective PTM systems. Based on the What Works Clearinghouse (WWC) standards as the criteria for identifying studies with rigorous research designs and evidence of causal relationships, our review focuses on two key outcomes of PTM systems and components: the extent to which certain policies and practices lead to improved **student achievement and principal retention**. Our review also highlights gaps in the existing research and offers recommendations for district leaders, policymakers, education-focused researchers, and funders of education leadership research, policy, and practice.

Literature Review Key Findings

- While there is clear evidence that principals play a critical role in improving student achievement, PTM research is **still emerging**. The limited amount of available evidence on the effectiveness of PTM practices reflects a relative lack of corresponding research.
 - Thus, **there is a need for more rigorous studies of PTM systems** especially given the importance of principal effectiveness to student achievement.
 - **There is also a need for more rigorous studies of the individual PTM components**—preparation, recruitment and selection, professional learning, performance evaluation, compensation and incentives, and working environment—to inform PTM systems building efforts.
- **Only four dimensions of PTM**—working environment, preparation, professional learning, and compensation and incentives—featured at least one study eligible for review.
- **Six studies of PTM components** (i.e., two for principal preparation, three for professional learning, and one for compensation) **met WWC criteria with or without reservations**.
- Of the six studies that met WWC criteria, **two studies** (one in the area of professional learning and one in the area of compensation) **had a positive impact on student achievement**.
 - In the area of professional learning, the National Institute for School Leadership's Executive Development program **had a positive impact on both reading/ELA and math achievement**.
 - In the area of principal compensation, the **Teacher Incentive Fund's pay-for-performance bonuses had potentially positive effects on students' reading/ELA achievement outcomes**.

Introduction

Great principals are essential to school and student success. Research clearly shows that principals have a significant impact on student achievement. Of all school-based factors, only classroom teachers have a greater impact on student achievement than principals.¹

Consequently, the ability to attract, support, and retain effective school leaders represents one of the most important organizational assets for any school district.

A number of related factors can influence principals' effectiveness and longevity, such as the quality of their initial training, the types of supports they receive as new principals, the culture and climate of their schools and districts, and the opportunities available for ongoing professional growth. Many school districts and states have increasingly focused on developing principal talent management (PTM) systems as a means to address these factors.¹

Effective PTM systems support principals' careers from start to finish, spanning the preparation and licensure processes as well as the ongoing cycle of principal evaluation and professional growth. PTM, in its ideal form, is a systematic approach to developing an effective school leadership workforce that encompasses the following dimensions:

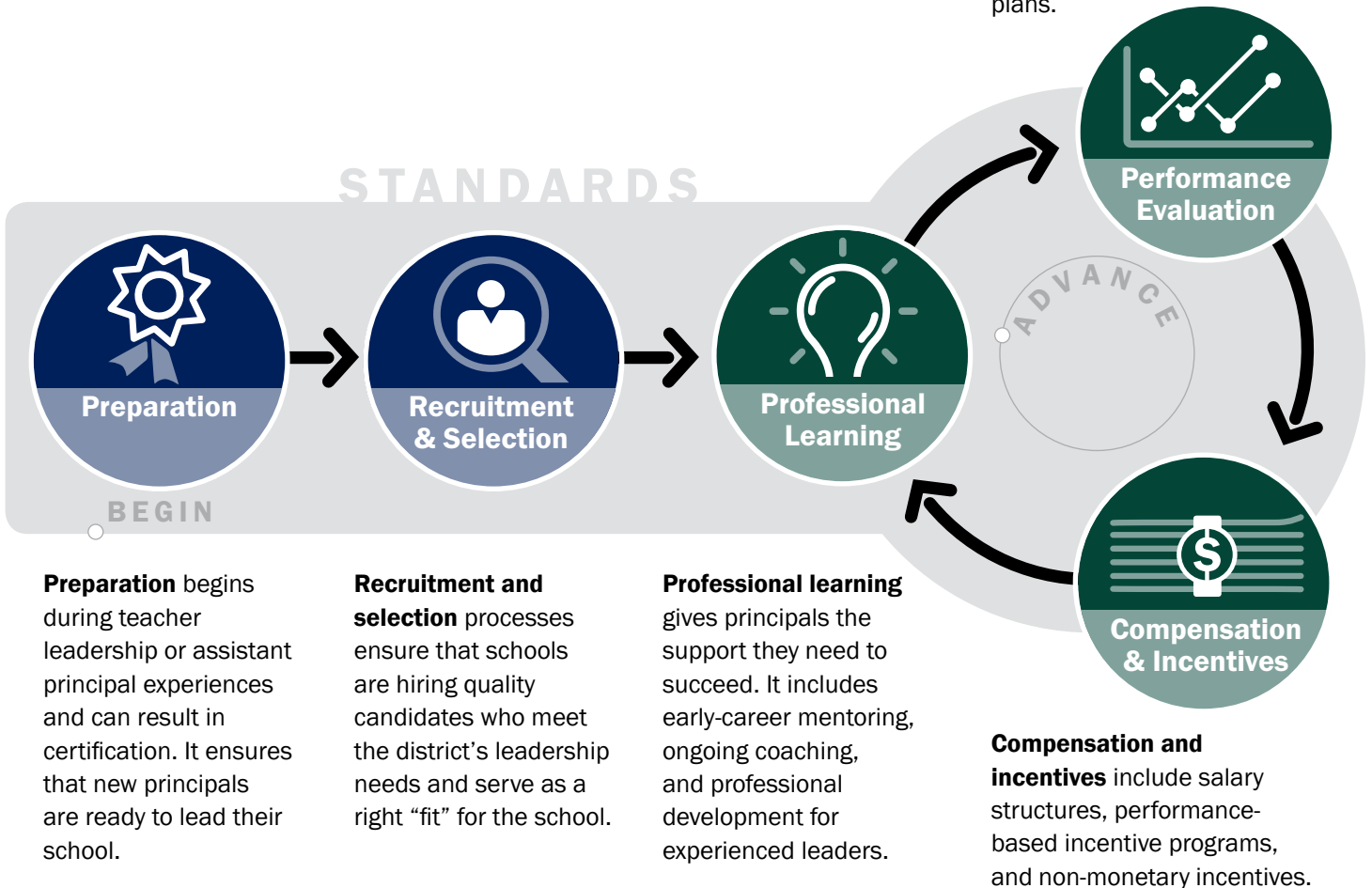
- Working Environment
- Preparation
- Recruitment and Selection
- Professional Learning
- Performance Evaluation
- Compensation and Incentives

More and more states and districts are implementing policies and programs that aim to improve some or all of these PTM components.² However, these efforts are often isolated from one another because they are provided by multiple agencies or from multiple departments within the district that do not communicate leaders' development. PTM is different in that it brings all elements together, emphasizing a common set of standards for principal performance assessment and development across the career path (e.g., Professional Standards for Educational Leadership, 2015)ⁱⁱ and closely coordinating within (e.g., between professional development offices and HR offices involved in hiring) and across organizations (e.g., between preparation programs and districts) so that leadership development is part of a coherent end-to-end system.ⁱⁱⁱ A detailed framework for PTM systems is presented in Figure 1.

FIGURE 1 A Framework for Principal Talent Management

A Framework for Principal Talent Management follows the continuum of a principal's career. It begins with the linear progression of preparation and recruitment, followed by more cyclical components once the principal is leading a school. Each component is **rooted in standards** and **measurable competencies** and influenced by a principal's **working environment**.

Performance evaluation systems that are fair and valid help inform recruitment and provide information for individual professional learning plans.



Preparation begins during teacher leadership or assistant principal experiences and can result in certification. It ensures that new principals are ready to lead their school.

Recruitment and selection processes ensure that schools are hiring quality candidates who meet the district's leadership needs and serve as a right "fit" for the school.

Professional learning gives principals the support they need to succeed. It includes early-career mentoring, ongoing coaching, and professional development for experienced leaders.

Compensation and incentives include salary structures, performance-based incentive programs, and non-monetary incentives.



The **working environment** includes district policies and practices that give a principal the right supports, balanced with the autonomy to make critical decisions. The working environment has an impact on aspiring principals' experiences during preparation programs, which may include in-district residency experiences, and may influence where an aspiring principal chooses to work. Once a principal is employed, the working environment could influence a principal's ability to make decisions and have support for curriculum, programming, teacher and staff talent management, and professional learning opportunities. The working environment also encompasses district-level policies and practices related to accountability and other demands on a principal's time and effort.

Approach to the Review

This literature review focuses on two specific outcomes of PTM policies and practices: **effects on student achievement and effects on principal retention**. This narrow focus necessarily excludes other outcomes—such as the effects of PTM on principal practice or school climate—that may be of interest in relation to student achievement. However, the variable quality of measures for these other outcomes, and the resulting variable quality of research claims, contributes to their exclusion from this document. **Two primary research questions** motivate this review:

- To what extent do specific PTM policies and practices improve student achievement?
- To what extent do specific PTM policies and practices improve principal retention?

The results of this review provide the foundation for a discussion of the current state of research on each individual component of PTM systems—as well as the state of research on PTM systems overall. To this end, our review also addresses **three secondary questions** in addition to the two main research questions listed above:

- Given the literature review methodology, what are the gaps in knowledge on PTM that might be addressed by future research?
- Are there particular challenges, methodological or other, that may have contributed to gaps in more rigorous quantitative research on the extent to which talent management policies and practices are associated with principal retention or student achievement?
- In light of the research findings, how should states and districts strengthen their approach to PTM?

The scope of this literature review is determined by a protocol that includes and excludes studies based on their use of research designs that can detect causal relationships. We adopted the What Works Clearinghouse (WWC) standards to evaluate the extent to which each reviewed study meets the highest criteria for research design and evidentiary rigor. A full complete description of the review protocol is available in the appendix. Broadly, [our review protocol](#) included the following activities:

- Our team searched multiple sources to find all published and unpublished studies of PTM and PTM components. This initial search identified 251 studies.
- These 251 studies were then screened against predetermined eligibility criteria; based on this screening, only 11 studies were eligible for inclusion in the WWC standards review. Two of these evaluated the same intervention and the more recent of the two incorporates the findings from the earlier study. Consequently, we included only the most recent of these studies in the WWC review.
- Of the ten studies reviewed by WWC-certified reviewers, six met WWC standards, either with or without reservations. The findings for these six studies are profiled in our review. The other four studies are summarized in [the appendix](#), but not profiled because they do not meet WWC standards.

This type of literature review is particularly relevant in light of the new Every Student Succeeds Act (ESSA) evidence guidelines, which encourage state and district decision-makers to select principal improvement interventions that are backed by “well-designed and well-implemented” research. This evidentiary requirement provides incentives for districts to enact interventions that have a proven impact on principal performance.

Some district leaders, as they plan for the implementation of ESSA, are still uncertain as to which programs and interventions are considered evidence-based and which criteria they must meet. In order to be considered as ESSA-eligible, backed by well-designed, well-implemented research, the evidence supporting a given program or intervention must meet the criteria for one of the following tiers:³

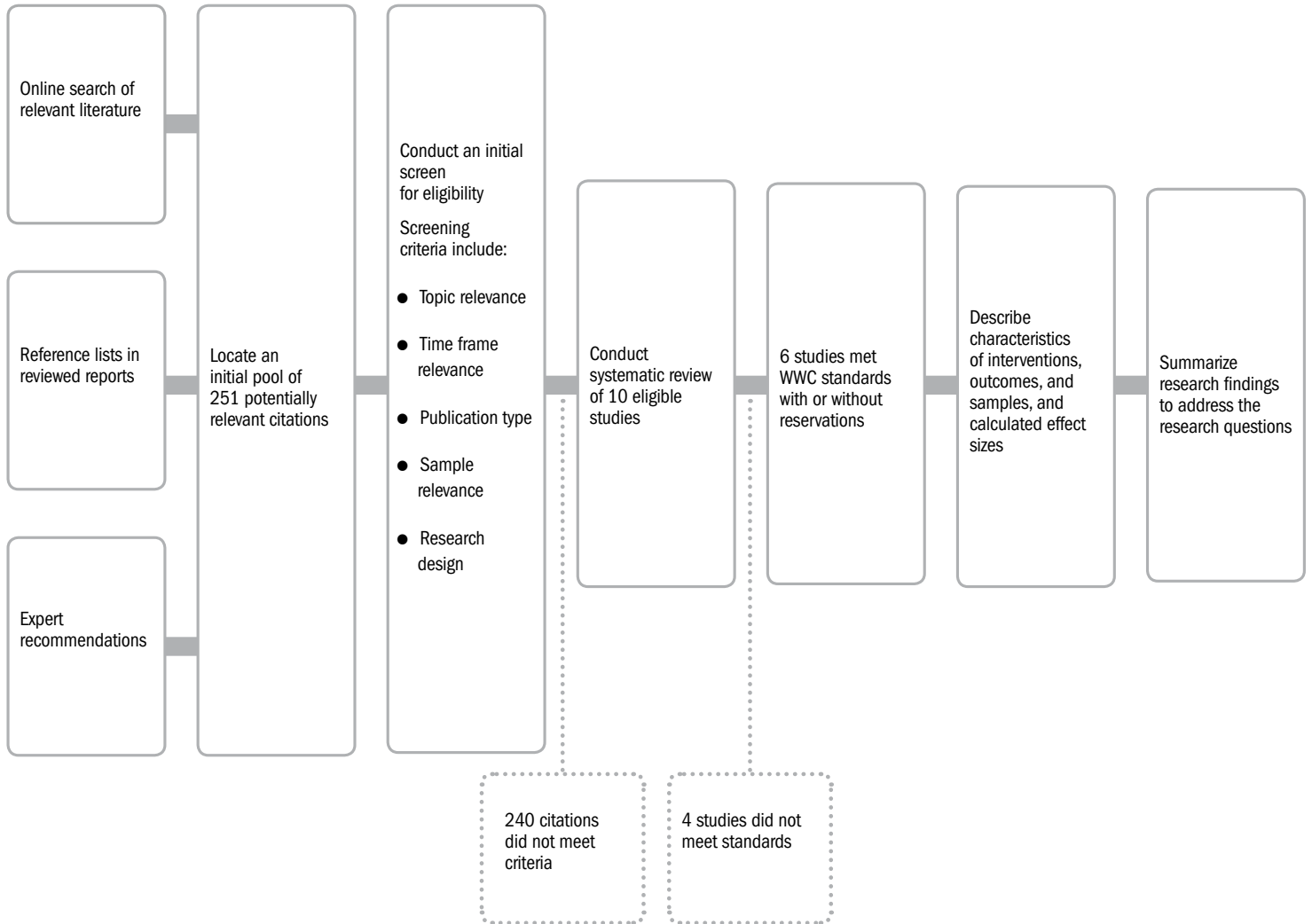
- **Tier I:** Strong evidence from at least one well-designed and well-implemented randomized control study (RCT)
- **Tier II:** Moderate evidence from at least one well-designed and well-implemented quasi-experimental study
- **Tier III:** Promising evidence from at least one well-designed and well-experimented correlational study that controls for selection bias
- **Tier IV:** For non-Title I school improvement activities, demonstrates a rationale based on research or a positive evaluation plan that results in improvement.

The original intent of this review was not to identify studies that meet the ESSA evidence tiers. However, the studies cited here could be considered for this purpose. In addition, we have created the Principal Talent Management Framework (Bush Institute 2016), which identifies additional policies and practices that may line up with the range of ESSA evidence tiers. As a part of this review, we used WWC review criteria as the basis for identifying “well-designed and well-implemented” studies. The WWC standards, at a minimum, require evidence from a quasi-experimental study. Furthermore, WWC maintains that substantively important and statistically significant findings serve as a minimum base of evidence. This means that the studies contained in this literature review would hypothetically meet Tier I or Tier II of the ESSA requirements.

Other literature reviews—such as Herman, et al. (2016)—have adhered to more expansive standards that allow for the inclusion of a greater number of PTM-focused studies. This review, given the stringency of the WWC standards, excludes some of the studies included in other reviews such as studies with promising evidence bases or theoretical linkages, that, while compelling, do not meet the WWC standards. For a detailed explanation of the differences between ESSA evidence tiers and WWC standards, see [Appendix G](#).

Figure 2 explains the systematic review process we used to identify studies eligible for inclusion and evaluate and review them against the WWC standards. [Appendix E](#) provides a complete account of our review protocol.

FIGURE 2 Summary of Literature Review Process



Summary of Findings

Research clearly shows that effective principals have a positive impact on student achievement.⁴ Yet, many schools do not have a cohesive system in place that provides continual resources, training, and supports for principals. As a result, they struggle to recruit, develop, and retain great school leaders. With a systematic approach to PTM that spans the continuum of a principal's career, schools can implement the policies that lay the groundwork for effective leadership—and higher student achievement.

Unlike the substantial evidence on the effect of great principals, the research on PTM systems—and the specific components of those systems—is still in a nascent state. There is not yet a large body of empirical evidence that attests to the ability of specific PTM policies and practices to increase student achievement or principal retention. However, the limitations of the research are not an indication that PTM systems and their components are ineffective. Rather, these limitations are a reflection of the research itself and of the rigorous standards applied in our literature review.

Moreover, the limitations of the empirical research underscore the need for more rigorous studies of PTM systems and of individual components. This is especially true in light of the strong evidence that demonstrates the important effect principals have on student achievement.

Notably, our literature review did not identify any studies of PTM systems, and their effects on student achievement or principal retention, that meet the established review criteria. The majority of published studies examine a specific component of PTM; most of these highlight single case studies, correlational studies, quasi-experimental studies, studies without a comparison group, or simply a set of policy recommendations.

Our screening of 251 study reports identified 10 studies on PTM component areas that were eligible for WWC review. The remaining 240 studies were deemed ineligible for review and inclusion based on topic, timeline, sample relevance, publication type, or research design.

- **Six studies of PTM components met WWC criteria with or without reservations.** Of 10 studies reviewed, we identified six studies that met WWC criteria. *Two principal preparation studies, three principal professional learning studies, and one principal compensation study* met the criteria with or without reservations. Most eligible studies did not meet the WWC standards due to a failure to demonstrate comparability between intervention and comparison groups prior to the new policy or practice.
- **Of the six studies that met WWC criteria for methodological rigor, two studies showed positive effects on student achievement.** Two rigorous studies (one in the area of professional learning and one in the area of compensation and incentives) displayed statistically significant, positive effects on student achievement. None of the studies displayed a statistically significant effect on principal retention.

The two studies that demonstrated the positive impact of PTM components on student achievement focused on the areas of principal learning and principal compensation. Nunnery et al. (2011) found that National Institute for School Leadership's Executive Development program positively impacted reading/ELA and math achievement in schools in Massachusetts and Pennsylvania. Chiang et al. (2015) found that the Teacher Incentive Fund's pay-for-performance model had a positive impact on reading/ELA achievement. Encouragingly, these two studies point to the promise of PTM and its potential to lead to improved student achievement.

Figure 3 summarizes the 10 studies eligible for WWC review by PTM component and details the strength of the evidence each study provides regarding the impact of particular programs or practices on student achievement or principal retention.

FIGURE 3 Summary of WWC Effectiveness Ratings, by Area of Talent Management

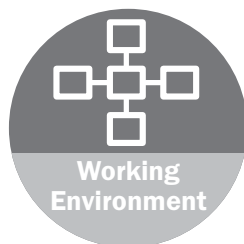
System or Component	Results of Eligibility Screening	Results of Review of Study Methods ^a			Results of Review of Study Evidence		
	Eligible	Meet WWC standards without reservations	Meet WWC standards with reservations	Do not meet WWC standards	Outcome	Effectiveness rating ^b	Extent of evidence ^c
PTM systems	0				(No studies)		
Preparation	4		2	2	ELA/Reading achievement	No discernible effects	Medium to large
					Math achievement	No discernible effects	Medium to large
Recruitment and selection	0				(No studies)		
Professional learning	3	1	2		ELA/Reading achievement	Potentially positive effects	Medium to large
					Math achievement	Potentially positive effects	Medium to large
Performance evaluation	0				(No studies)		
Compensation and incentives	2	1		1	ELA/Reading achievement	Potentially positive effects	Small
					Math achievement	No discernible effects	Small
					Principal retention at the school	No discernible effects	Small
Working conditions	1			1	(No studies)		
Total	10	2	4	4			

Notes:

- a. See [Appendix E](#) on how the ratings for individual studies and the overall ratings for studies within each PTM area were determined.
- b. Effectiveness ratings are based on all studies that meet WWC standards with or without reservations within each area of principal talent management. Outcomes were averaged across time points, studies, and interventions to obtain a summative effectiveness rating for each outcome domain within each PTM area. See figures 4, 6 and 8 for effectiveness ratings for outcomes at each time point and for each intervention; see Appendix F for detailed findings for individual studies. “No discernible effects” indicates that none of the studies show statistically significant or substantively important effects, either positive or negative. “Potentially positive effects” indicates that there is some evidence of a positive effect with no overriding contrary evidence. More specifically, at least one study shows statistically significant or substantively important positive effects; AND, fewer or the same number of studies show indeterminate effects than show statistically significant or substantively important positive effects; AND, no studies show statistically significant or substantively important negative effects.
- c. A “small” amount of evidence denotes the outcome domain includes only one study, or one setting, or is based on a total sample size of fewer than 14 schools or principals. “Medium to large” amount of evidence denotes that the domain includes more than one study, AND more than one setting, AND the domain findings are based on a total sample of at least 350 students, or 14 schools across studies. See tables [E4](#) and [E5](#) in Appendix E for more details on how these ratings were determined.

Detailed Findings of WWC Literature Review

The following subsections provide detailed findings on each of the four PTM component areas where at least one study was eligible for review: *working environment*, *preparation*, *professional learning*, and *compensation and incentives*. Each profiled study met WWC standards—either with or without reservations—but did not necessarily produce statistically significant findings. Studies that were eligible for review but did not meet WWC standards were only summarized, not profiled. Figure 4 lists each of the pre-identified studies. It also provides a summary of the number of studies that met WWC criteria, met WWC criteria with reservations, or did not meet criteria. Note that all four studies that do not meet the WWC criteria are summarized in Appendix B.



Principal Working Environment

Studies Reviewed

Pulliam, C.L., LaCaria, L., Schoeneberger, J., & Algozzine, B. (2014)

Strategic Staffing Initiative (SSI)

Overall, there is no evidence on whether specific interventions in the area of the principal working environment improve student achievement or principal retention outcomes. The policies and practices that influence a principal’s administrative environment are distinct from other factors, such as student demographics or the socioeconomic factors of a school’s neighborhood, that may influence working conditions but are largely out of district and policymaker control.

The principal’s working environment is impacted by numerous factors over which the district has some control. These factors range from principal supervisor caseloads and principal autonomy over budgetary or professional learning decisions to the general conditions of classrooms and school facilities. Additional policies that may influence the administrative working environment include state teacher evaluation and assessment requirements, state and federal reporting requirements, and other mandated aspects of the job.

Overview

In our review of the eligible research on the principal’s working environment, we did not find any studies that met WWC standards with or without reservations. Just one study failed to meet WWC standards (see Figure 4).

Notably, the Pulliam et al. (2014) study of the Strategic Staffing Initiative—a district-level turnaround, restructuring effort that assigned school leaders and key staff members from successful settings to schools that experienced deep and chronic levels of low student performance—failed to meet WWC standards because the intervention and comparison groups were not equivalent prior to the intervention. A brief description of the Pulliam et al. (2014) study is presented in [Appendix B](#).

FIGURE 4 Summary of Results of Eligibility Screening and WWC Review

Study Citation	Intervention Name	Location	Outcomes of Interest		Study Design	Meeting WWC's Group Design Standards	Reason for not meeting standards
			Principal Retention	Student Achievement			
PREPARATION							
George W. Bush Institute (2016)	Four principal preparation programs selected by Bush Institute	Four urban districts in US	No	Yes	QED	Meets standards with reservations	
Corcoran, S. P., Schwartz, A. E., & Weinstein, M. (2012)	Aspiring Principals Program	New York City	Yes	Yes	QED	Meets standards with reservations	
Gates, S. M., Hamilton, L. S., Martorell, P., Burkhauser, S., Heaton, P., Pierson, A., & Gu, K. (2014)	The New Leaders Principal-Preparation Program	Various school districts nationwide	Yes	Yes	QED	Does not meet standards	Failed to establish baseline equivalence
Vanderhaar, J. E., Munoz, M. A., & Rodosky, R. J. (2006)	District-driven preparation programs and university-based preparation program	A large, Midwestern, urban school district	No	Yes	QED	Does not meet standards	Failed to establish baseline equivalence
PROFESSIONAL LEARNING							
Jacob, R., Goddard, R., Kim, M., Miller, R., & Goddard, Y. (2014)	McREL's Balanced Leadership® Professional Development (BLPD) Program	Michigan	Yes	Yes	RCT	Meets standards without reservations	
Nunnery, J. A., Ross, S. M., & Yen, C. J. (2010)	NISL's Executive Development Program	Pennsylvania	No	Yes	QED	Meets standards with reservations	
Nunnery, J.A., Ross, S.M., Chappell, S., Pribesh, S., Hoag-Carhart, E. (2011)	National Institute of School Leadership's (NISL's) Executive Development Program	Massachusetts	No	Yes	QED	Meets standards with reservations	
COMPENSATION AND INCENTIVES							
Chiang, H, Wellington, A., Hallgren, K., Speroni, C., Herrmann, M., Glazerman, S., & Constantine, J. (2015)	Pay-for-performance bonuses	Across US	Yes	Yes	RCT	Meets standards without reservations	
Vanlaar, A. (2011)	Denver Public Schools' (DPS) Professional Compensation System for Administrators ("Principal ProComp")	Denver, CO	Yes	No	RD	Does not meet standards	Failed all four pilot regression discontinuity design standards
WORKING ENVIRONMENT							
Pulliam, C.L., LaCaria, L., Schoeneberger, J., & Algozzine, B. (2014)	Strategic staffing initiative (SSI)	A large district in Southeastern district in US	No	Yes	QED	Does not meet standards	Failed to establish baseline equivalence

Key Meets standards without reservations Meets standards with reservations Does not meet standards



Principal Preparation

Studies Reviewed

George W. Bush Institute. (2016).	Four principal preparation programs selected by Bush Institute
Corcoran, S. P., Schwartz, A. E., & Weinstein, M. (2012).	NYC Leadership Academy's Aspiring Principals Program
Gates, S. M., Hamilton, L. S., Martorell, P., Burkhauser, S., Heaton, P., Pierson, A., & Gu, K. (2014).	The New Leaders Principal-Preparation Program
Vanderhaar, J. E., Munoz, M. A., & Rodosky, R. J. (2006).	District-driven preparation programs and university-based preparation program

Principal preparation encompasses policies, programs, and practices that typically target candidates prior to when they receive certification as an elementary or secondary school principal. To determine whether preparation affects student achievement or principal retention, this set of studies examined impacts once principals earned their certification and obtained positions within schools.

The review identified four eligible research studies on principal preparation, **two of which met WWC standards with reservations.**

- The Corcoran et al. (2012) study and the Bush Institute principal preparation study (2016) **met WWC standards with reservations.** While both studies used a quasi-experimental design, they both established baseline equivalency according to WWC-accepted procedures.
- The Gates et al. (2014) study of the New Leaders principal preparation program and the Vanderhaar et al. (2016) study of a Midwestern principal preparation program **did not meet WWC standards.** Both studies used quasi-experimental designs and comparison groups that were not equivalent to the treatment groups.

Below, we provide an overview of the Corcoran et al. (2012) and the Bush Institute principal preparation studies (2016). The findings of these two studies are documented in more detail in [Appendix F. Detailed Findings of WWC Review](#). The Gates et al. (2014) and Venderhaar (2016) studies that do not meet WWC criteria are described in [Appendix B](#).

Corcoran et al. (2012)—The NYC Leadership Academy’s Aspiring Principals Program

Overview

NYC Leadership Academy (NYCLA) and the New York City Department of Education (NYCDOE) are partners in this program, which had been operating for a decade prior to the time of the study. The program prepares principals to lead high-need New York City (NYC) schools and implement practices that improve outcomes for all students. While NYCLA is considered a principal preparation program, it is unique in that the majority of applicants to the NYCLA Aspiring Principals Program (APP) obtain their New York state principal licensure (School Building Leader license) from another institution beforehand. The 14-month program has a rapid preparation timeline, which includes a six-week “Summer Intensive” with a standards-based curriculum that simulates on-the-job challenges. Following the Summer Intensive, APP candidates participate in a school-based six-month residency under the mentorship of a principal. They then use the following “planning summer” to transition into a principal position at a high-need school.

Effectiveness of the Intervention

According to the WWC Procedures and Standards Handbook, there is a **small degree of supporting evidence for this intervention** because the outcome domain includes one eligible study. This study compares student achievement and principal retention in schools led by APP graduates to student achievement and principal retention in schools led by graduates of other preparation programs. To measure student achievement in the ELA/Reading domain, the study examined four separate intervention effects based on the duration of program implementation. Our review found that:

- NYCLA APP **did not have a discernible impact** on student ELA/Reading achievement after the second year of implementation, after which there was not a clear pattern of effects (i.e., after Year 2, 3, 4, and 5).
- The **NYCLA APP did not have discernible impacts on student math achievement**—regardless of different program durations.

The quasi-experimental study also compared school-wide student learning growth in ELA/Reading and mathematics and school principal retention. However, the analysis of principal retention **did not meet WWC standards** because the study did not demonstrate baseline equivalence for the intervention and comparison groups; therefore, it is not discussed in this review.

WWC Improvement Index

The WWC reports the magnitude of study findings in two ways: (a) effect sizes (i.e., standardized mean differences) and (b) a WWC-calculated “improvement index.” In order to help readers judge the practical importance of an intervention’s effect, the WWC translates effect sizes into “improvement index” values. The improvement index for an individual study finding represents the difference between the percentile rank corresponding to the mean value of the outcome for the intervention group and the percentile rank corresponding to the mean value of the outcome for the comparison group distribution. The improvement index can be interpreted as the expected change in percentile rank for an average comparison group student if the student had received the intervention.

Source: WWC, 2015.

George W. Bush Institute (2016)—AREL Innovative Principal Preparation Programs

Overview

The George W. Bush Institute (2016) evaluated four selected principal preparation programs that meet a general set of criteria that experts and qualitative research associate with promising practices in principal preparation. Those criteria include:

- Program alignment to research-based competencies
- Evidence of a significant experiential learning component
- Evidence of a rigorous recruitment and selection process
- On-the-job support throughout the early years of a principal's career
- Evidence of a collaborative partnership between the program and the district served
- A demonstrated commitment to collecting data that can inform continual improvements to a program.

Although the programs vary in terms of their specific situation (i.e., university program or non-traditional program), approaches, duration, and the curricula used to prepare leaders, each strongly aligns with the above selection criteria.

Effectiveness of the Intervention

The Bush Institute principal preparation study (2016) evaluates the relative effectiveness of inexperienced principals from the four programs of interest against inexperienced principals who graduated from other programs within their district.^{iv} Our review of the study shows that the four selected principal preparation programs **did not have a discernible impact on student achievement in ELA/reading or math domains** in the four partnering districts.

The study also examines the effectiveness of all principals (regardless of experience level) prepared by the four programs of interest against principals trained by other programs in each of the four districts. However, that specific analysis does not appear in this review because it does not use an eligible design.

FIGURE 5 Summary of Effectiveness Rating Across Studies that Focused on Principal Preparation: By Outcome Domain

Outcome domain (Duration of program implementation)	Intervention (Study)	Rating of effectiveness	Domain Average Improvement index (percentile points)	Number of principals	Extent of evidence
ELA/READING ACHIEVEMENT DOMAIN					
ELA/Reading achievement (after Year 2 of intervention)	NYCLA APP (Corcoran et al., 2012)	No discernible effects	0	301	Small
ELA/Reading achievement (after Year 3 of intervention)	NYCLA APP (Corcoran et al., 2012)	No discernible effects	0	275	Small
ELA/Reading achievement (after Year 4 of intervention)	NYCLA APP (Corcoran et al., 2012)	No discernible effects	1	168	Small
ELA/Reading achievement (after Year 5 of intervention)	NYCLA APP (Corcoran et al., 2012)	No discernible effects	-3	60	Small
ELA/Reading achievement (after 1-4 years of intervention)	Bush Institute principal preparation study	No discernible effects	-1	155	Small
MATH ACHIEVEMENT DOMAIN					
Math achievement (after Year 2 of intervention)	NYCLA APP (Corcoran et al., 2012)	No discernible effects	-1	301	Small
Math achievement (after Year 3 of intervention)	NYCLA APP (Corcoran et al., 2012)	No discernible effects	-1	275	Small
Math achievement (after Year 4 of intervention)	NYCLA APP (Corcoran et al., 2012)	No discernible effects	1	168	Small
Math achievement (after Year 5 of intervention)	NYCLA APP (Corcoran et al., 2012)	No discernible effects	-1	60	Small
Math achievement (after 1-4 years of intervention)	Bush Institute principal preparation study	No discernible effects	-1	155	Small

Notes: Detailed findings are documented in Tables F1 through F10. In the table, “no discernible effects” signifies that none of the studies show statistically significant or substantively important effects—either positive or negative. In the table, a “small” amount of evidence signifies that the outcome domain includes only one study, one setting, or is based on a total sample size of fewer than 14 schools or principals. See tables E4 and E5 in Appendix E for more details on the process for determining these ratings.



Principal Professional Learning

Studies Reviewed

Jacob, R., Goddard, R., Kim, M., Miller, R., & Goddard, Y. (2014)	McREL's Balanced Leadership® Professional Development (BLPD) Program
Nunnery, J. A., Ross, S. M., & Yen, C. J. (2010)	NISL's Executive Development Program
Nunnery, J.A., Ross, S.M., Chappell, S., Pribesh, S., & Hoag-Carhart, E. (2011)	National Institute of School Leadership's (NISL's) Executive Development Program

Principal professional learning encompasses policies, programs, and practices that aim to improve current school principals' leadership and managerial skills. Professional learning (also referred to as professional development) programs typically include activities such as principal coaching and formal coursework. This category does not entail professional learning that empowers principals to improve teachers' abilities to instruct students. However, principals may also attend these types of trainings with their teachers.

Our review of the eligible research on principal professional learning found that **one study met WWC standards without reservations** and **two studies met WWC standards with reservations**:

- The Jacob et al (2014) study is a well-designed and well implemented RCT **that met WWC group design standards without reservations.**
- The Nunnery et al. (2010) study in Pennsylvania schools used a quasi-experimental design but featured groups that were equivalent at the baseline. Therefore, the study met **WWC standards with reservations.**
- The Nunnery et al. (2011) study evaluated the same intervention—i.e., NISL's EDP—but assessed its implementation in Massachusetts schools. This study also used a quasi-experimental design but demonstrated that groups of schools were equivalent at baseline. Therefore, the study **met WWC standards with reservations.**

Below, we summarize the findings from our review of these three studies. See Figure 6 and [Appendix F. Detailed Findings of WWC Review](#) for more detail.

Jacob et al. (2014)—The McREL Balanced Leadership Professional Development Program

Intervention Overview

The BLPD program aims to enhance principals' effectiveness and improve student outcomes through a focus on 21 “key leadership responsibilities” (e.g., culture, order, communication, etc.).^v Waters et al. (2003) finds that these responsibilities are associated with increased student achievement. Over a cohort-based set of ten two-day workshops, BLPD trains principals to use the set of practices required to carry out the 21 leadership responsibilities. The trainings also seek to empower principals with an understanding of the change process and establish a “purposeful community to focus organizational

resources on agreed upon goals” (Jacob et al., 2014, p. 2). Notably, the BLPD approach also aligns with other major school leadership frameworks such as the Wallace Foundation Principal Pipeline Initiative’s five practices of effective school leaders (Wallace Foundation, 2013) and the Interstate School Leaders Licensure Standards (ISLLC, 2008).^{vi}

Effectiveness of the Intervention

Our review of the Jacob et al. (2014) study shows that **McRel’s PBDP program has no discernible effects on student achievement in ELA/reading or math domains after the third year of implementation.**

Jacob et al. (2014) also examined the McREL BLPD program’s impact on principal retention. Even though the analysis met WWC standards, the report did not provide enough information to allow for the calculation of effect size (see Appendix F). Therefore, the analysis of the outcome is not documented in Figure 6.

School Leadership’s Executive Development Program

Overview

The NISL EPD aims to “train school leaders to drive their schools to high performance” (Nunnery et al, 2010, p. 6), featuring a focus on “the role of the principal as a strategic thinker, instructional leader, and creator of a just, fair, and caring culture in which all students meet high standards” (p. 6). The program design launched with a multi-million-dollar investment in research and pilots across a five-year period. The structure of the program includes not only in-person and virtual coursework, but an emphasis on interactive learning, with “simulations, case studies, school evaluations, and online activities” (p. 6).^{vii}

Effectiveness of the Intervention

We reviewed two studies that examined the NISL EPD’s effectiveness. Nunnery et al (2010) evaluated the effectiveness of the intervention on student achievement in Pennsylvania schools, while Nunnery et al. (2011) did so in Massachusetts schools. As Figure 6 shows, there is some evidence **that NISL’s EDP has a positive impact on student achievement in both ELA/reading and math domains:**

- Nunnery et al. (2010) produced *an average improvement index of three for ELA/reading achievement* after Year 3 of the program in Pennsylvania schools, a potentially positive effect size.
- Detailed findings in [Appendix F](#) indicate that this domain average improvement index of three stems from four separate ELA/reading outcome measures (i.e., proficiency on Pennsylvania state reading assessment) across four different samples (i.e., within-district matching elementary school sample, out-of-district matching elementary school sample, middle school sample, and high school sample).
- In addition, Nunnery et al. (2010) produced a *domain average improvement index of five for math achievement* after Year 3 of the program in Pennsylvania schools, a potentially positive effect size.
- Nunnery et al (2011) concluded that the same NISL EDP intervention did not have a *discernible impact in Massachusetts schools after Year 4 of the intervention*. These results are based on Massachusetts Comprehensive Assessment Program (MCAP) scores.
- The program produced an improvement index rating of four in the ELA/reading domain and produced an index rating of six in the math domain. Even though these are relatively large improvement index values, *they are not statistically significant*.

FIGURE 6 Summary of Effectiveness Rating Across Studies that Focused on Principal Professional Learning: By Outcome Domain

Outcome domain (Duration of program implementation)	Intervention (Study)	Rating of effectiveness	Domain Average Improvement index (percentile points)	Number of principals	Extent of evidence
ELA/READING ACHIEVEMENT DOMAIN					
ELA/Reading achievement (after Year 3 of intervention)	McREL's BLPD (Jacob et al., 2014)	No discernible effects	0	119	Small
ELA/Reading achievement (after Year 3 of intervention)	NISL's EDP (Nunnery et al., 2010)	Potentially positive effects	3	202	Small
ELA/Reading achievement (after Year 4 of intervention)	NISL's EDP (Nunnery et al., 2011)	No discernible effects	4	1,015	Small
MATH ACHIEVEMENT DOMAIN					
Math achievement (after Year 3 of intervention)	McREL's BLPD (Jacob et al., 2014)	No discernible effects	1	119	Small
Math achievement (after Year 3 of intervention)	NISL's EDP (Nunnery et al., 2010)	Potentially positive effects	5	202	Small
Math achievement (after Year 4 of intervention)	NISL's EDP (Nunnery et al., 2011)	No discernible effects	6	1,015	Small

Notes: Detailed findings are documented in Tables [F11](#) through [F15](#).

"No discernible effects" connotes that none of the studies show statistically significant or substantively important effects—either positive or negative. "Potentially positive effects" indicate that there is some evidence of a positive effect without overriding contrary evidence. More specifically, at least one study shows statistically significant or substantively important positive effects; fewer or the same number of studies show indeterminate effects as opposed to statistically significant or substantively important positive effects; and none of the studies show statistically significant or substantively important negative effects. A "small" amount of evidence connotes that the outcome domain includes only one study, one setting, or is based on a total sample size of fewer than 14 schools or principals. See tables [E4](#) and [E5](#) in Appendix E for more details on the process and methodology used to determine these ratings.



Principal Compensation and Incentives

Studies Reviewed

Chiang, H, Wellington, A., Hallgren, K., Speroni, C., Herrmann, M., Glazerman, S., & Constantine, J. (2015).	Pay-for-performance bonuses
Vanlwaarden, A. (2011).	Denver Public Schools' (DPS) Professional Compensation System for Administrators ("Principal ProComp")

The principal compensation and incentives component includes initiatives that incentivize principal performance, or in some cases, willingness to work in “hard to staff” or “high needs” schools and districts. Examples of compensation programs could include higher salaries or signing bonuses for principals that commit to serve in challenging schools for a given period of time. Examples of incentives might include bonuses for achieving certain student growth benchmarks or subsidies for higher education coursework.

Our review of the eligible research on principal compensation and incentives found **one study that met WWC standards without reservations and one study that did not meet WWC standards** (see Figure 7).

- The Chiang et al. (2015) study is a randomized control trial (RCT) that tested the effectiveness of the Teacher Incentive Fund’s (TIF) pay for performance program. As a well-designed and well-implemented RCT, it met WWC standards without reservations. The findings of this study are documented in more detail in Figure 7 and in *Appendix F. Detailed Findings of WWC Review*.
- The Vanlwaarden (2011) study of Denver Public Schools’ Professional Compensation System for Administrators (Principal ProComp) is based on a unique research approach called regression discontinuity design (RDD).⁵ To review the Vanlwaarden (2011) study, we used WWC’s pilot RDD standards. Our review shows that the study failed to meet relevant standards for RDD.^{viii}

Below, we describe findings from our review of the Chiang et al. (2015) intervention and its effectiveness. The Vanlwaarden (2011) study is described in [Appendix B](#).

Chiang et al. (2015)—Teacher Incentive Fund Pay-for-Performance Bonuses

Overview

The Teacher Incentive Fund (TIF) is a federal education grant that supports districts in the design and implementation of performance-based compensation programs for both teachers and leaders in underserved schools.

Requirements of the TIF grant compensation and incentives programs stipulate that:

1. Teacher and leader evaluations must include measures of student achievement growth and a minimum of two observations as part of the summative effectiveness rating;
2. Grantees must include pay for performance bonuses—that are based on educator effectiveness ratings—under the new evaluation systems. While TIF grants do not prescribe the specific type or amount of bonuses that should be given under the compensation program, they provide general guidance on “how to structure pay for performance bonuses to be substantial, differentiated, and challenging to earn” (Chiang et al., 2015, p. xix);
3. Beyond the pay for performance bonuses, teachers and leaders must also have opportunities to receive additional financial compensation in exchange for taking on “additional roles and responsibilities” (p. xx).

Effectiveness of Intervention

The Chiang et al. (2015) study is a RCT designed to test the effects of TIF’s pay for performance bonuses on student achievement and principal retention for a subset of 2010 TIF grantees in ten school districts. As Figure 7 shows, **TIF’s pay for performance bonuses had potentially positive effects on students’ ELA/reading achievement outcome—whether the program was implemented for one or two years.** Even though the improvement index for respective duration of implementation was low (i.e., one), it was statistically significant. On the other hand, the same intervention did not produce potentially positive effects on students’ math achievement outcomes. Lastly, TIF’s pay-for-performance bonuses did not yield any discernible effects on principal retention—regardless of the duration of program implementation (i.e., one or two years).

The Chiang et al. (2015) study also examined the impacts of pay-for-performance on schools’ student achievement growth ratings (based on student performance but measured on a scale of one to four). We considered this an eligible outcome for the purpose of this review and noted the analysis met WWC standards without reservations. However, the report does not provide enough information to allow for calculation of effect sizes for this outcome. Hence, the outcome is not documented in Figure 7.

The study also examined other principal outcomes that were not eligible for review such as: principals’ effectiveness as measured by principals’ observation ratings, principals’ attitudes toward TIF program as well as their satisfaction with professional opportunities, evaluation system, and school environment (measured by survey responses).

FIGURE 7 Summary of Effectiveness Rating Across Studies that Focused on Principal Compensation and Incentives: By Outcome Domain

Outcome domain (Duration of program implementation)	Intervention (Study)	Rating of effectiveness	Domain Average Improvement index (percentile points)	Number of principals	Extent of evidence
ELA/READING ACHIEVEMENT DOMAIN					
ELA/Reading achievement (after Year 1 of intervention)	TIF's Pay-for-Performance Bonuses (Chiang et al, 2015)	Potentially positive effects	1	134	Small
ELA/Reading achievement (after Year 2 of intervention)	TIF's Pay-for-Performance Bonuses (Chiang et al, 2015)	Potentially positive effects	1	134	Small
MATH ACHIEVEMENT DOMAIN					
Math achievement (after Year 1 of intervention)	TIF's Pay-for-Performance Bonuses (Chiang et al, 2015)	No discernible effects	1	134	Small
Math achievement (after Year 2 of intervention)	TIF's Pay-for-Performance Bonuses (Chiang et al, 2015)	Potentially positive effects	1	134	Small
PRINCIPAL RETENTION DOMAIN					
Principal retention (after Year 1 of intervention)	TIF's Pay-for-Performance Bonuses (Chiang et al, 2015)	No discernible effects	9	134	Small
Principal retention (after Year 2 of intervention)	TIF's Pay-for-Performance Bonuses (Chiang et al, 2015)	No discernible effects	9	134	Small

Notes: Detailed findings are documented in Tables [F16](#) through [F21](#).

“No discernible effects” connotes that none of the studies show statistically significant or substantively important effects—either positive or negative. “Potentially positive effects” indicate that there is some evidence of a positive effect without overriding contrary evidence. More specifically, at least one study shows statistically significant or substantively important positive effects; fewer or the same number of studies show indeterminate effects as opposed to statistically significant or substantively important positive effects; and none of the studies show statistically significant or substantively important negative effects. A “small” amount of evidence connotes that the outcome domain includes only one study, one setting, or is based on a total sample size of fewer than 14 schools or principals. See tables [E4](#) and [E5](#) in Appendix E for more details on the process and methodology used to determine these ratings.

Conclusion and Recommendations

Overall, this review points to a limited body of rigorous research that examines the impacts of PTM systems and system components on principal retention and student learning. Some will be surprised by these findings, especially given the number of leadership-related interventions and the extent of articles and papers that address the topic each year. **Many of these studies, however, either did not address student achievement or principal retention—or were designed in a manner inconsistent with WWC’s rigorous methodological criteria.** Experimental and quasi-experimental studies of leadership interventions employing comparison groups are challenging to design and implement, but nonetheless offer the most compelling evidence that they are effective across diverse contexts.

Based on the findings from this review, we offer the following recommendations:

For researchers and funders of research, *PTM According to the Evidence* raises, again, the need for well-designed and well-implemented educational leadership research studies, particularly those related to principal talent management. In studying the impact of interventions aimed at improving school leadership, there are clear methodological challenges that are similar to and also different from studies of the impact of teacher-related interventions. For example, it may take a significant amount of time for efforts to demonstrate that school leadership impacts student achievement—especially given that principals’ relationship to student learning is not direct. It is often difficult to identify the required, large sample of principals experiencing a given program for a sufficient period of time. Addressing these challenges may mean reconsidering which programs or practices to study—or the timeframe for studying impact.

We recommend researchers and funders of research develop multi-year, methodologically diverse research agendas on PTM and its components, which can result in compelling evidence and justify investment in well-designed experimental and quasi-experimental studies. For some PTM components, less rigorous studies that provide “promising evidence” of positive effects, in ESSA terms, may provide more useful evidence than a single study that provides moderate or strong evidence. For example, a particular program or policy may be most relevant to a particular geographic area or type of principal (e.g. new principals). It may be more useful to focus on this group rather than trying to establish a large enough or exact enough sample to do an RCT. In pursuing a well-formed research agenda, we can identify where and how we can establish causal links through research—and where we really cannot—which would be an important step moving forward.

In addition, this review focuses on student achievement and principal retention. Because of a principal’s role in a school, some might argue that efforts to improve school leadership might more easily or quickly be observed through changes in a principal’s practice, in school climate, or in other outcomes. To address this challenge, researchers and funders of research need to invest in work that produces high-quality, reliable, and low-burden measures of practice, school climate, and other relevant outcomes.

For states and district policymakers, the gaps in empirical evidence present a difficult but familiar position of trying to select and implement strategies that will lead to principal and ultimately student success—without strong information on how to do that successfully. Even without specific supporting evidence, detailed examples from states and districts that have implemented efforts in the area of PTM still allow policymakers to collectively build knowledge and develop strategies that may be ripe for in-depth study after they’ve had a chance to mature.

In light of statistics that illustrate declining retention in the principal workforce and poor professional supports, state and district decision-makers may need to adopt and implement strategies with less rigorous evidence (as defined by the WWC criteria). When doing so, states and districts will need to present compelling arguments based upon available research or expert opinion.

We recommend that state and district decision-makers engage with researchers to rigorously study PTM and its components through federal or foundation funding, thus contributing to the evidence base. At a minimum, states and districts should engage in internal evaluation that examines programmatic impact and continuously improves services to principals.^{ix} Studies that do not meet WWC criteria can still yield useful information for improving local PTM systems and can lead to new and better programs to improve school leadership.

States and districts should also recognize their ability to influence the amount and quality of research available in the field. They should continue to press service providers, researchers, and others for compelling evidence on PTM and its components by asking tough questions about programs, policies, and the methodology supporting claims of effectiveness.





With the evidence and recommendations presented here, district leaders and state policymakers have the tools to begin to implement PTM and greatly expand the supports available for principals across their careers. Researchers, meanwhile, can build on a cohort of studies that point to the positive impact of an effective PTM framework. Significant work still needs to be done—particularly in raising awareness about the importance of school leadership and encouraging the development of studies that assess these policies and practices. If that work is done rigorously, states and districts can improve their abilities to recruit and retain great principals—and improve student achievement in the process.




- BACK TO TABLE OF CONTENTS
- i. The U.S. Department of Education's Teacher Incentive Fund and School Leadership Program provided support for development of innovative principal talent management systems. The Wallace Foundation has provided funding and support for principal pipeline development in six large, urban school districts, and the Gates Foundation has provided funding to support teacher evaluation development and district partnership development, both of which address the need for talented principals.
 - ii. National Policy Board for Educational Administration (2015). *Professional Standards for Educational Leaders 2015*. Reston, VA. <http://www.ccsso.org/Documents/2015/ProfessionalStandardsforEducationalLeaders2015forNPBEAFINAL.pdf>
 - iii. In *A Framework for Principal Talent Management*, the George W. Bush Institute provides an in-depth overview of PTM, which include examples in practice.
 - iv. The study also examines the effectiveness of all principals (regardless of experience) prepared through the four programs of interest as compared to all principals trained by other programs in each of the four districts. However, that analysis is ineligible for this review because it does not use an eligible design.
 - v. For more information on the BLPD Framework and key leadership responsibilities, see: <http://files.eric.ed.gov/fulltext/ED544245.pdf>
 - vi. The ISLLC 2008 standards were recently "refreshed" as the Professional Standards for Education Leaders (PSEL, 2015)
 - vii. For more information on the NISL EPD, see: <http://www.nisl.net/executive-development-program/>
 - viii. Four relevant standards for RDD include: (1) The average treatment effect for an outcome must be estimated using a statistical model that controls for the forcing variable; (2) A graphical analysis displaying the relationship between the outcome and forcing variable—including a scatter plot and a fitted curve—must be included in the report; (3) The study must provide evidence that an appropriate parametric, semi-parametric, or nonparametric model was fit to the data; and (4) Any constraints on the relationship between the outcome and the forcing variable (e.g., constraining the slope of the relationships to be the same on both sides of the cutoff) need to be supported by either a statistical test or graphical evidence.
 - ix. Bryk et al., 2015.
 - x. This list is representative and may not be exhaustive of all the search terms used. AIR searched for each talent management component in conjunction with each outcome.
 - xi. This review protocol was adapted from the Review Protocol for Teacher Training, Evaluation, and Compensation Version 3.1 (May 2015), which guides the review of research that informs the What Works Clearinghouse (WWC) intervention reports in the Teacher Training, Evaluation, and Compensation topic area. The protocol is used in conjunction with the WWC Procedures and Standards Handbook (version 3.0).
 - xii. While this review protocol is guided by these two specific research questions, the broader purpose of the literature review publication addresses two additional research questions: (1) What gaps are apparent in knowledge on leadership talent management that might be addressed by future research? And (2) Are there particular methodological or other types of challenges or limitations that may have contributed to gaps in more rigorous, quantitative research on the extent to which talent management policies and practices are associated with principal retention or student achievement?
 - xiii. While there is a great deal of research documenting the extent to which working conditions that are outside of the district's control (e.g., student demographics, parental involvement, and other contextual factors), are associated with the outcomes of interest, working conditions associated with these external factors are not the focus of this literature review.
 - xiv. This document borrows heavily from the WWC Procedures and Standards Handbook (version 3.0).
 - xv. Note that our review does not include author queries, which provide study reviewers the opportunity to follow up directly with study authors to clarify methodological or other questions related to the review.
 - xvi. Depending on the number and nature of studies that meet the WWC standards, the review team may decide to combine findings from studies that address similar policies or practices under the same area of principal talent management (that is, preparation, recruitment and selection, induction and mentoring, evaluation and professional growth, compensation and incentives, and working conditions).
 - xvii. U.S. Department of Education. (2016). *Non-regulatory guidance: Using evidence to strengthen education investments*. Washington, DC: Author. Retrieved from <http://www2.ed.gov/policy/elsec/leg/essa/guidanceusesinvestment.pdf>
1. Davis, Darling-Hammond, LaPointe, & Meyerson, 2005; Darling-Hammond, Meyerson, LaPointe, & Off, 2010; Levine, 2008; Wallace Foundation, 2015.
 2. For example, the Wallace Foundation invested \$75 million to enhance principal pipelines in six urban school districts.
 3. Herman, Gates, Chavez-Herrerias, & Mark Harris, 2016
 4. Davis, Darling-Hammond, LaPointe, & Meyerson, 2005; Darling-Hammond, Meyerson, LaPointe, & Off, 2010; Levine, 2008; Wallace Foundation, 2015.
 5. Under an RDD, the effect of an intervention is estimated as the difference in mean outcomes between treatment and comparison group units at the cutoff, adjusting statistically for the relationship between the outcomes and the variable used to assign units to the intervention.
 6. The study failed all the criteria under each of the four pilot standards: integrity of the forcing variable, sample attrition, continuity of the outcome-forcing variable relationship, functional form and bandwidth.

Principal effectiveness, principal talent management, principal professional development, human capital, talent management, human capital talent management, principal effectiveness and student achievement, school district principal effectiveness strategies, principal retention, principal efficacy, principal satisfaction, principal attitudes and principal retention, principal compensation and principal retention, principal improvement strategies, principal selection, principal selection best practice, principal recruitment, principal development and teacher retention, educator environment and principal retention, school climate and principal effectiveness, school climate and principal retention, principal preparation, principal induction, mentoring and principal retention, high quality principal mentoring, principal mentoring and principal effectiveness, principal evaluation, principal compensation, educator environment and principal retention, educator environment and principal attitudes, educator environment and high quality principal practice, educator environment and principal talent pool, principal working conditions, principal working conditions and principal retention, principal working conditions and principal attraction, principal working conditions and principal satisfaction, principal working conditions and student achievement, principal autonomy, principal autonomy and principal retention, principal autonomy and principal attraction, principal autonomy and principal satisfaction, principal autonomy and student achievement, principal pipeline development, leadership pipeline development

Note: This list is representative and may not be exhaustive of all the search terms used. AIR searched for each talent management component in conjunction with each outcome.

APPENDIX B.1 Summary of Studies That Were Eligible for Review but Did Not Meet Standards (With or Without Reservations)

Study Citation	Intervention Name	Location of Study	Principal Retention Outcomes?	Student Achievement Outcomes?	Study Design	WWC Rating
Gates, S. M., Hamilton, L. S., Martorell, P., Burkhauser, S., Heaton, P., Pierson, A., ... & Gu, K. (2014). <i>Preparing principals to raise student achievement: Implementation and effects of the New Leaders program in ten districts.</i>	The New Leaders Principal-Preparation Program	Various school districts nationwide	Yes	Yes	QED	 Does not meet standards
Vanderhaar, J. E., Munoz, M. A., & Rodosky, R. J. (2006). <i>Leadership as accountability for learning: The effects of school poverty, teacher experience, previous achievement, and principal preparation programs on student achievement.</i>	District-driven preparation programs and university-based preparation program	A large, Midwestern, urban school district	No	Yes	QED	 Does not meet standards
Vanlwaarden, A. (2011). <i>Principal ProComp and retention: 2011 report brief.</i>	Denver Public Schools' (DPS) Professional Compensation System for Administrators ("Principal ProComp")	Denver, CO	Yes	No	RDD	 Does not meet standards
Pulliam, C.L., LaCaria, L., Schoeneberger, J., & Algozzine, B. (2014). A preliminary analysis of a strategic staffing initiative. <i>Educational Policy, 28</i> (4), pp. 578-603.	Strategic staffing initiative	A large Southeastern district in U.S.	No	Yes	QED	 Does not meet standards

Key  Meets standards without reservations  Meets standards with reservations  Does not meet standards

The Gates et al. (2014) study examines the implementation of the New Leaders program and its impact on student achievement. The New Leaders programs involves both principal preparation and partnering with districts to improve working conditions for principals. The study uses a quasi-experimental design that compares the achievement of students who attended schools led by New Leaders principals with the achievement of similar students who attended schools led by non-New Leaders principals within the same district. The study found statistically significant differences in achievement gains between students in schools led by New Leader principals and students in schools led by non-New Leaders principals in reading and mathematics at the lower grade levels—and in reading at the high school level. Our review concludes that the study does not meet WWC standards because the report does not provide sufficient data to establish the equivalence between the intervention and comparison groups prior to the intervention.

The Vanderhaar et al. (2006) study examines the relationship between school leadership preparation programs and student achievement in a large, Midwestern, urban school district. Our review focuses on the analyses that compare the impact of two kinds of principal preparation on student achievement: (1) participating in the district-driven principal preparation program (versus not participating); and (2) participating in the preparation program from the “primary” university (one metropolitan university where 56% of the district’s principals were from) versus from other universities. The study found no statistically significant main effect for either kind of principal preparation. Our review shows that the study does not meet WWC standards because the equivalence of the intervention and comparison groups prior to the intervention cannot be established.

The Vanlwaarden (2011) study examines the effect of Denver Public Schools’ Professional Compensation System for Administrators (Principal ProComp) on the retention of principals and assistant principals. This program aims to attract, retain, and train high-quality principals to work in Denver Public Schools by providing them with various individual incentives. The study uses a regression discontinuity design (RDD) to examine how three incentives—the Hard-to-Serve (HTS), Top Performing (TPS), and High Growth Schools (HGS) incentives—may be influencing retention of administrators that work in schools that are near the cut-off points used to make award decisions for each incentive. The analyses suggest a positive effect of the three incentives on administrator retention, particularly for administrators near the cut-off points. Our review of this study using WWC’s pilot RDD standards shows that the study failed to meet all relevant standards for RDD.⁶

The Pulliam et al. (2014) study examines the impact of the Strategic Staffing Initiative (SSI) on student outcomes. The SSI reassigned principals and key staff members from settings in which they were successful to chronically low-performing schools and allowed principals increased autonomy related to curriculum and instruction. The study uses a quasi-experimental design that compares the achievement of students in each of the SSI elementary schools with students in a closely matched school within the district. The study found that the student achievement gains and growth in reading, mathematics, and science in SSI schools were similar to those in comparison schools after one year of implementation. The study does not meet WWC standards due to nonequivalence of the intervention and comparison groups prior to the intervention.

This complete citation list includes all studies, articles, reports, and other documentation that researchers screened and determined were ineligible for WWC review after not meeting eligibility criteria. Studies were deemed ineligible for review because they did not meet one or more of the following eligibility criteria:

- The study was not a primary analysis of the effect of an intervention.
- The study did not have an eligible design.
- The study did not use a sample aligned with the protocol.
- The study did not include an outcome within a domain specified in the protocol.
- The study was not published in the relevant time frame.

- ACTION UNITED Education Fund. (2012). *Revolving doors: Findings from Philadelphia's highest-poverty schools*. Philadelphia, PA: Author.
- Augustine, C. H., Gonzalez, G., Ikemoto, G. S., Russel, J., Zellman, G. L., Constant, L., & Dembosky, J. W. (2009). *Improving school leadership: the promise of cohesive leadership systems*. Santa Monica, CA: RAND Corporation.
- Baker, B.D., Punswick, E., & Belt, C. (2010). *School leadership stability, principal moves, and departures: Evidence from Missouri*. *Educational Administration Quarterly*, 46(4).
- Barnes, C.A., Camburn, E., Sanders, B.R., & Sebastian, J. (2010). *Developing instructional leaders: Using mixed methods to explore the black box of planned change in principals' professional practice*. *Educational Administration Quarterly*, 46(2), 241–279.
- Béteille, T., Kalogrides, D., & Loeb, S. (2012). Stepping stones: Principal career paths and school outcomes. *Social Science Research*, 41(4), 904–919.
- Braun, D., Gable, R., & Kite, S. (2011). Relationship among essential leadership preparation practices and leader, school, and student outcomes in K-8 schools. *International Journal of Educational Leadership Preparation*, 6(3), 1–21.
- Briggs, K., Rhines Cheney, G., Davis, J., & Moll, K. (2012). *Operating in the dark: what outdated state policies and data gaps mean for effective school leadership*. Houston, TX: George W. Bush Institute.
- Bryk, A. S., Gomez, L. M., & Grunow, A. (2015). *Learning to improve: How America's schools can get better at getting better*. United States: Harvard Educational Publishing Group.
- 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 Corporation.
- Camburn, E. M., Goldring, E., May, H., Supovitz, J., Barnes, C., & Spillane, J. P. (2007, April). *Lessons learned from an experimental evaluation of a principal professional development program*. Paper presented at the American Educational Research Association, Washington, DC.
- Clark, D., Martorell, P., & Rockoff, J. (2009). *School Principals and School Performance* (Working Paper 38). Washington, D.C: Center for Analysis of Longitudinal Data in Education Research.
- Creighton, T.B., & Jones, G.B. (2001, August) *Selection or self-selection? How rigorous are our selection criteria for education administration preparation programs?* Paper presented at the 2001 conference of the National Council of Professors of Educational Administration, Houston, TX.

- Donmoyer, R., Yennire-Donmoyer, J., & Galloway, F. (2012). *The search for connections across principal preparation, principal performance, and student achievement in an exemplary principal preparation program*. *Journal of Research in Leadership Education*, 7(1), 5–43.
- Friedman, B. A., Friedman, M. A., & Markow, D. (2008). Predictors of principals' satisfaction with their schools. *Journal of Educational Administration*, 46(5), 598–612.
- Fuller, E., Young, M., & Baker, B. D. (2011). Do principal preparation programs influence student achievement through the building of teacher-team qualifications by the principal? An exploratory analysis. *Educational Administration Quarterly*, 47(1), 173–216.
- Fuller, M. & Young, E. (2009). *Tenure and retention of newly hired principals in Texas*. Paper presented at the University Council for Educational Administration, Austin, TX.
- Fuller, M. & Young, E. (2009, April). *Tenure and retention of newly hired principals in Texas*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.
- Furgeson, J., Knechtel, V., Sullivan, M., Tuttle, C.C., Akers, L., Anderson, M.A., and Nichols-Barrer, I. (2014). *KIPP leadership practices through 2010-2011: Technical report*. Mathematica Policy Research: Cambridge, MA.
- Gates, S., et al. (2003). *Who is Leading Our Schools? An Overview of School Administrators and Their Careers*. Santa Monica, CA: RAND Corporation.
- Hamilton, L.S., Engberg, J., Steiner, E.D., Nelson, C.A., & Yuan, K. (2012). *Improving school leadership through support, evaluation, and incentives: The Pittsburgh Principal Incentive Program*. Santa Monica, CA: RAND Corporation.
- Hemphill, K., & Nauer, C. (2010). *Managing by the numbers: Empowerment and accountability in New York City's schools*. New York, NY: Center for New York City Affairs, The New School for Management and Urban Policy.
- Kaplan, L. S., Owings, W. A., & Nunnery, J. (2005). Principal quality: A Virginia study connecting interstate school leaders' licensure consortium standards with student achievement. *NASSP Bulletin*, 89(643), 28–44.
- Knoepfel, R. C., & Rinehart, J. S. (2008). Student achievement and principal quality: Explaining the relationship. *Journal of School Leadership*, 18(5), 501–527.
- Kwan, Y.K.P., & Walker, A.D. (2009) Are we looking through the same lens? Principal recruitment and selection. *International Journal of Educational Research*, 48(1), 51–61.
- Leithwood, K. (2011). *Characteristics of high performing school systems in Ontario: Part I*. Ontario, Canada: Institute for Educational Leadership.
- May, H., Huff, J., & Goldring, E. (2012). A longitudinal study of principals' activities and student performance. *School Effectiveness and School Improvement*, 23(4), 417–439.
- Miller, A. (2013). Principal turnover & student achievement. *Economics of Education Review*, 36, 60–72.
- Newton, R., Giesen, J., Freeman, J., Bishop, H., & Zeitoun, P. (2003). Assessing the reactions of males and females to attributes of the principalship. *Educational Administration Quarterly*, 39, 504–532.
- Orphanos, S., & Orr, M. T. (2014). Learning leadership matters: The influence of innovative school leadership preparation on teachers' experiences and outcomes. *Educational Management Administration & Leadership*, 42(5), 680–700.
- Orr, M. T. & Orphanos, S. (November 2011). How graduate-level preparation influences the effectiveness of school leaders: A comparison of the outcomes of exemplary and conventional leadership preparation programs for principals. *Educational Administration Quarterly*, 47, 18–70.

- Owings, W. A., Kaplan, L. S., & Nunnery, J. (2005). Principal quality, ISLLC standards, and student achievement: A Virginia study. *Journal of School Leadership, 15*(1), 99–119.
- Papa, F. C., Jr., Lankford, H., & Wyckoff, J. (2002). *The attributes and career paths of principals: Implications for improving policy*. Albany, NY: Teacher Policy Research.
- Papa, F. Jr., (2007). Why do principals change schools? A multivariate analysis of principal retention. *Leadership and Policy in Schools, 6*(3), 267–290.
- Pijanowski, J.C., & Brady, K.P. (2009). The influence of salary in attracting and retaining school leaders. *Education and Urban Society, 42*(1), 25–41.
- Porter, A.C., Polikoff, M.S., Goldring, E. B., Murphy, J., Elliot, S. N., & May, H. (2010, December). Investigating the validity and reliability of the Vanderbilt Assessment of Leadership in Education. *The Elementary School Journal, 111*(2), 282–313.
- Pounder, D. G., & Merrill, R. J. (2001). Job desirability of the high school principalship: A job choice theory perspective. *Educational Administration Quarterly, 37*(1), 27–57.
- Rammer, R. A. (2007). Call to action for superintendents: Change the way you hire principals. *Journal of Educational Research, 101*(2), 67–76
- Replogle Sheppard, R. (2010, May). *Determining factors that influence high school principal turnover over a five-year period* (Unpublished doctoral dissertation). University of North Texas, Denton, TX.
- Ross, C., Herrmann, M., & Angus, M.H. (2015). *Measuring principals' effectiveness: Results from New Jersey's principal evaluation pilot*. Washington, D.C.: Institute of Education Sciences.
- Silva, J. P., White, G. P., & Yoshida, R. K. (2011). The direct effects of principal-student discussions on eighth grade students' gains in reading achievement: An experimental study. *Educational Administration Quarterly, 47*(5), 772–793.
- Stoelinga, S. R., Hart, H., & Schalliol, D. (2008). *The work of Chicago Public Schools' principals: Leading in a complex context with high stakes (Research Report)*. Chicago, IL: Consortium on Chicago School Research.
- Teh, B., Chiang, H., Lipscomb, S., Gill, B. (2014). *Measuring school leaders' effectiveness: An interim report from a multiyear pilot of Pennsylvania's Framework for Leadership*. Washington, D.C.: Institute of Education Sciences.
- Weinstein, M., Schwartz, A. E., Jacobowitz, R., Ely, T., & Landon, K. (2009). *New schools, new leaders: A study of principal turnover and academic achievement at new high schools in New York City*. New York, NY: NYU Wagner School of Public Service.


APPENDIX D1 Summary of WWC Effectiveness Ratings, by Area of Talent Management




System or component	Outcome domain	Rating of effectiveness ^a	Improvement index (percentile points)		Number of studies	Number of principals	Number of students	Extent of evidence ^b
			Average	Range				
Preparation	ELA/Reading achievement	No discernible effects	0	-3 to 1	2	456	°	Medium to large
	Math achievement	No discernible effects	-1	-1 to -1	2	456	°	Medium to large
Recruitment and selection	(No studies)							
Professional learning	ELA/Reading achievement	Potentially positive effects	2	0 to 4	3	1,336	°	Medium to large
	Math achievement	Potentially positive effects	4	1 to 6	3	1,336	°	Medium to large
Performance evaluation	(No studies)							
Compensation and incentives	ELA/Reading achievement	Potentially positive effects	1	1 to 1	1	134	40,576	Small
	Math achievement	No discernible effects	1	1 to 1	1	134	40,852	Small
	Principal retention at the school	No discernible effects	9	9 to 9	1	134	na	Small
Working conditions	(No studies)							

Notes:

- Effectiveness ratings are based on all studies that meet WWC standards with or without reservations within each area of talent management. Outcomes measured at multiple time points are averaged. See figures 4, 6 and 8 for effectiveness ratings for outcomes at each time point and for each intervention; see Appendix F for detailed findings for individual studies. In the figure, none of the studies show statistically significant or substantively important effects—either positive or negative. Potentially positive effects indicate that there is some evidence of a positive effect with no overriding contrary evidence. More specifically, at least one study shows statistically significant or substantively important positive effects; AND, Fewer or the same number of studies show indeterminate effects than show statistically significant or substantively important positive effects; AND, No studies show statistically significant or substantively important negative effects.
- A “small” amount of evidence denotes the outcome domain includes only one study, or one setting, or is based on a total sample size of fewer than 14 schools or principals. “Medium to large” amount of evidence denotes that the domain includes more than one study, AND more than one setting, AND the domain findings are based on a total sample of at least 350 students, or 14 schools across studies. See tables E4 and E5 in Appendix E for more details on how these ratings were determined.
- One or more studies does not provide number of students.

APPENDIX D2 Summary of Results of Eligibility Screening and WWC Review

Study Citation	Intervention Name	Location	Outcomes of Interest		Study Design	Meeting WWC's Group Design Standards	Reason for not meeting standards
			Principal Retention	Student Achievement			
PREPARATION							
George W. Bush Institute (2016)	Four principal preparation programs selected by Bush Institute	Four urban districts in US	No	Yes	QED	 Meets standards with reservations	
Corcoran, S. P., Schwartz, A. E., & Weinstein, M. (2012)	Aspiring Principals Program	New York City	Yes	Yes	QED	 Meets standards with reservations	
Gates, S. M., Hamilton, L. S., Martorell, P., Burkhauser, S., Heaton, P., Pierson, A., & Gu, K. (2014)	The New Leaders Principal-Preparation Program	Various school districts nationwide	Yes	Yes	QED	 Does not meet standards	Failed to establish baseline equivalence
Vanderhaar, J. E., Munoz, M. A., & Rodosky, R. J. (2006)	District-driven preparation programs and university-based preparation program	A large, Midwestern, urban school district	No	Yes	QED	 Does not meet standards	Failed to establish baseline equivalence
PROFESSIONAL LEARNING							
Jacob, R., Goddard, R., Kim, M., Miller, R., & Goddard, Y. (2014)	McREL's Balanced Leadership® Professional Development (BLPD) Program	Michigan	Yes	Yes	RCT	 Meets standards without reservations	
Nunnery, J. A., Ross, S. M., & Yen, C. J. (2010)	NISL's Executive Development Program	Pennsylvania	No	Yes	QED	 Meets standards with reservations	
Nunnery, J.A., Ross, S.M., Chappell, S., Pribesh, S., & Hoag-Carhart, E. (2011)	National Institute of School Leadership's (NISL's) Executive Development Program	Massachusetts	No	Yes	QED	 Meets standards with reservations	
COMPENSATION AND INCENTIVES							
Chiang, H, Wellington, A., Hallgren, K., Speroni, C., Herrmann, M., Glazerman, S., & Constantine, J. (2015)	Pay-for-performance bonuses	across US	Yes	Yes	RCT	 Meets standards without reservations	
Vanlaar, A. (2011)	Denver Public Schools' (DPS) Professional Compensation System for Administrators ("Principal ProComp")	Denver, CO	Yes	No	RD	 Does not meet standards	Failed all four pilot regression discontinuity design standards
WORKING CONDITIONS							
Pulliam, C.L., LaCaria, L., Schoeneberger, J., & Algozzine, B. (2014)	Strategic staffing initiative (SSI)	A large Southeastern district in the U.S.	No	Yes	QED	 Does not meet standards	Failed to establish baseline equivalence

Key  Meets standards without reservations  Meets standards with reservations  Does not meet standards

Review Protocol for the Bush Institute’s Principal Talent Management Literature Review^{xi}

Purpose Statement

This review focuses on principal talent management policies and practices that aim to increase retention and make them more effective at improving the academic achievement of students in grades PK-12. We define principal talent management as district-level policies and practices that attract, prepare, support, and retain principals who can effectively improve school and student outcomes. For the purpose of this literature review, principal talent management policies and practices are organized into the following six areas: *preparation, recruitment and selection, evaluation, professional learning, compensation and incentives, and working conditions*. The primary outcomes of interest are principal retention and student achievement.

The following research questions guide this review^{xiii}:

- To what extent do specific principal talent management policies and practices improve principal retention?
- To what extent do specific principal talent management policies and practices improve student academic achievement?

Key Definitions

Principal. A state-certified educational administrator who holds the title of “school principal,” “chief executive officer,” or “head of school” or is otherwise administratively responsible for K-12 public, public charter, or private schools serving children and youth. For the purpose of this review, assistant principals or other school leaders are not included in the definition of “principal.” However, if a study sample includes both principals and assistant principals and the study does not present separate findings for principals and assistant principals, the study is also eligible for review.

Categories of relevant research. The review team identified and defined seven areas of principal talent management that encompassed policies and practices aimed at improving retention and increasing capacity to improve student achievement. Those categories are:

- **Preparation:** Studies that examine the effects of principal preparation program design.
- **Recruitment and selection:** Studies that examine the effects of processes used to identify quality candidates as well as select and match new school leaders who meet the leadership needs of schools and districts.
- **Performance evaluation:** Studies that examine the effects of principal performance evaluation. Such evaluations provide information to guide planning of principal professional development, meet workforce needs to inform recruitment and selection decisions and support the basis for performance-based compensation.

- **Professional learning:** Studies that examine the effects of principal professional learning programs and policies, which could provide ongoing development and support. Professional learning may include induction or onboarding support for newly placed leaders, and from there, continual coaching or mentoring.
- **Compensation and incentives:** Studies that examine the effects of financial and nonfinancial rewards for principals based on their performance or other conditions (e.g., commitments to working in underserved schools or districts).
- **Working environments:** Studies that examine the effects of district policies and practices that impact principal working conditions and support their professional practice and growth. Working environments may include the district policies and practices that can influence a principal’s administrative capacity; for example, consider policies related to principal autonomy, accountability pressures, supervisory support, and the physical working conditions of the school building.^{xiii}
- **Principal talent management systems:** Studies that examines the effects of cohesive systems containing one or more elements of support for the principal career continuum that are aligned to a set of core competencies (e.g., principal preparation, recruitment, professional learning, performance evaluation, compensation and incentives, or policies related to principals’ working conditions).

Eligibility Criteria

To be eligible for inclusion in this literature review, studies must meet certain criteria. These criteria involve the types of student populations within principals’ schools, the types of interventions used, and the time frame and location in which a study was conducted. Detailed criteria are listed in the following sections.

Eligible Populations

In this review, the following populations are of interest:

- **Grade.** Principals must be the lead of a school that serves students in grades PK–12.
- **Location.** Principals must be employed by schools located within the United States, its territories, or tribal entities.

Potential subgroups of interest for this review (only if data is available) include:

- Characteristics of principals:
 - Experience
 - Demographics (e.g., gender, age, terminal degree)
- Characteristics of students:
 - Special education status
 - English learner status
 - Economically disadvantaged (e.g., free or reduced-price lunch status)
 - Grade
- Characteristics of school settings:
 - Location of school (e.g., urban, suburban, rural)
 - School level (e.g., elementary, middle, high)
 - School governance (e.g., traditional public, charter, private)
 - School size
 - Economically disadvantaged (e.g., Title I status, percentage of students eligible for free or reduced-price lunch)

Eligible Interventions

Only replicable interventions (i.e., those that can be reproduced) are eligible for review. An intervention must include the following characteristics in order to reliably re-occur with different participants, in other settings, and at other times:

- Intervention description: skills or practices, an approach to enhancing skills or practices (e.g., strategies, activities, and materials), and targeted population
- Intervention duration and intensity
- Description of individuals delivering or administering the intervention

In this review, interventions may include district principal talent management policies, practices, or programs that attract, prepare, support—and are effective in improving school and student outcome. The programs, practices, or policies must be clearly described and commonly understood in the field and literature. Examples of principal talent management interventions might include:

- Using a set of recommended principles within preparation programs, such as increased selectivity for admissions, a clinical residency, and a standards-based curriculum (*preparation*)
- Creating partnerships between school districts and principal preparation programs. (*preparation*)
- Requiring principals to participate in a year-long clinical residency program as part of a principal preparation program (*preparation*)
- Providing targeted, ongoing coaching support for principals (*professional learning*)
- Mentoring support for new principals (*professional learning*)
- Connecting principal evaluation results to professional learning opportunities. (*evaluation*)
- Offering financial incentives for effective principal leadership (*compensation and incentives*)
- Offering financial incentives for principals that commit to working in high need school districts or schools for a specific period of time (*compensation and incentives*)
- Providing principals with autonomy over budgetary and hiring decisions in exchange for increased accountability (*working conditions*)
- Reducing the administrative caseload for principal supervisors so they can provide additional support for principals (*working conditions*)

Eligible Research

In this review, the following additional parameters define the scope of research studies to be included:

- **Topic.** The study must examine the effect of principal talent management practices, policies, and programs on principal retention and student achievement. The practices, policies, or programs can focus on any part of the principal career path, including preparation, recruitment and selection, professional learning, performance evaluation, compensation and incentives, working environments, or a system that includes multiple talent management processes.

- **Time frame.** Studies must have been released or published after 1995.
- **Sample.** The study sample must meet the requirements described in the “Eligible Populations” section above.
- **Language.** The study must be available in English to be included in the review. Studies of interventions delivered in languages other than English will not be included in the review.
- **Location.** The study must include principals working in the United States, its territories, or tribal entities.

Eligible Outcomes

This review includes outcomes in the following domains:

Student-level Outcome Domains

English language arts achievement includes outcomes in the following areas: foundational reading (word reading, fluency and/or accuracy in reading connected text, vocabulary, reading comprehension), general reading, measures of English language conventions (e.g., grammar), writing, and general English language arts achievement (i.e., on a standardized test covering an array of language arts topics).

Mathematics achievement includes outcomes in the following areas: understanding of different subjects within mathematics, including algebra, arithmetic, calculus, geometry, probability, statistics, and trigonometry; understanding of concepts and procedures; understanding of word problems and applications; and general math achievement (i.e., on a standardized test covering an array of mathematics topics).

Science achievement includes outcomes in any of the physical or life science disciplines, such as biology, chemistry, earth science, general science, and physics.

Social studies achievement includes outcomes in social studies sub-disciplines, such as civics, economics, geography, history, and world cultures.

General achievement includes a general measure of student academic achievement and is used if study authors did not distinguish students’ achievement in specific areas (e.g., math, reading). Examples include composite scores from state assessments that represent a combination of reading and math scores.

Assessments unrelated to academic achievement are not eligible outcome measures.

Student progression includes measures of students’ progression in school. Constructs include:

- **Student promotion** (e.g., students’ advancement to next grade level)
- **Student graduation** (e.g., students’ completion of the PK–12 education system)

The review is limited to student achievement indicators. We recognize principals may influence other student outcomes such as student behavior, health, and nutrition. However, these indicators are not included in the scope of this review.

Principal-level Outcome Domains

Student growth scores include measures of student achievement growth attached to a specific school. Examples of eligible measures include school scores from a value-added model or the Colorado Growth Model.

Principal retention focuses on outcomes that assess whether or not principals return to their school, their school district, or the profession from year to year. More detailed mobility outcomes generally will not be reviewed because they either are captured by the key, commonly measured outcomes of interest (e.g., an indicator for moving to another high school in the school district would be captured by a broader outcome that measures whether a principal returned to lead in the same school district) or may not be defined consistently across studies (e.g., an indicator for moving from an intervention school to a comparison school). Furthermore, the focus for the principal retention outcome must be principals' actual movement from a school, not expected movement. For example, principal ratings on whether they expect to return to their positions are not eligible for review.

Principal retention at the school includes outcomes that measure the percentage of principals who return to their positions in the same school from year to year.

Principal retention in the school district includes outcomes that measure the percentage of principals who return to their positions in the same school district from year to year.

Principal retention in the profession includes outcomes that measure the percentage of principals who return to the principalship from year to year, regardless of school location.

Evidence Standards

Eligible studies are assessed against WWC evidence standards, as described in the *WWC Procedures and Standards Handbook* Section III: Screening and Reviewing Studies (pp. 8–21).

Sample Attrition

The *WWC Procedures and Standards Handbook* discusses the sample attrition standards used by the WWC in Section III: Subsection B.2 Sample Attrition: *Is the combination of overall and differential attrition high?* (pp. 11–15).

This review uses the **liberal** boundary for attrition. The *WWC Procedures and Standards Handbook* contains a figure illustrating the attrition boundary and an associated table with attrition levels that define high and low attrition. Based on the choice of the boundary, the study review guide calculates whether attrition is high or low.

Baseline Equivalence

If the study design is a randomized controlled trial or regression discontinuity design with high levels of attrition or a quasi-experimental design, the study must demonstrate baseline equivalence of the intervention and comparison groups for the analytic sample. The onus for demonstrating equivalence in these studies rests with the authors. The *WWC Procedures and Standards Handbook* discusses how authors must demonstrate baseline equivalence in Section III: Subsection B.3 Baseline Equivalence: *Is equivalence established at baseline for the groups in the analytic sample?* (pp. 15 and 16).

Baseline equivalence must be demonstrated for the intervention and comparison groups in the analytic sample on the following pre-intervention (or baseline) characteristic:

- A pre-intervention measure of the outcome; or,
- If a pre-intervention measure is not available, an acceptable alternative pre-intervention measure, as explained below and summarized in [Table E1](#).

Acceptable measures for student achievement outcomes. For outcomes in the student achievement domains, studies must show that the groups are equivalent on an acceptable pre-intervention measure of student achievement. A pretest measure in the same subject as the outcome is preferred; however, if a same-subject pretest is not available, a pretest measure of general achievement (e.g., a combined mathematics and reading score) is acceptable. In addition, a pretest measure of mathematics achievement can establish baseline equivalence for a science achievement outcome, and a pretest measure of reading achievement can do so for a social studies achievement outcome.

Acceptable measures for student progression outcomes. For outcomes in the student progression domain, studies must show that groups are equivalent on the following set of characteristics that are correlated with student progression.

- Grade level; **AND**
- One of the following measures of student academic performance: standardized test scores, proportion of grade-age youth (could be measured by age among students in the same grade), prevalence of school behavior or discipline issues, rate of school attendance, or grade point average (GPA); **AND**
- One of the following: student race/ethnicity or a measure of degree of disadvantage (i.e., free or reduced-price lunch status, poverty status, family income, English learner status, special education status, or disability status); **AND,**
- **If the unit of assignment is the school,** a school-level measure of the student progression outcome.

Acceptable measures for student growth score outcomes. For outcomes in the student growth scores domains, studies must show that the groups are equivalent on a pre-intervention measure of the outcome. Because principal or school characteristics are not highly related to student growth scores, these are not acceptable pre-intervention measures of student growth score outcomes.

Acceptable measures for principal retention outcomes. For outcomes in the principal retention domains, studies must show that groups are equivalent on the following set of characteristics that are correlated with principal retention.

- Principal experience; **AND**
- One of the following measures of student academic performance: standardized test scores, proportion of grade-age youth (could be measured by age among students in the same grade), prevalence of school behavior or discipline issues, rate of school attendance, or GPA; **AND**
- One of the following: student race/ethnicity or a measure of degree of disadvantage (i.e., free or reduced-price lunch status, poverty status, family income, English learner status, special education status, or disability status); **AND,**
- **If the unit of assignment is the school,** a school-level measure of the outcome. For example, if the outcome is principal retention in the profession and the unit of assignment is the school, equivalence must also be demonstrated on a baseline measure of the percentage of principals who returned to the principalship.

This review requires that, in a domain with statistical adjustments, the adjustment is made only for that outcome. For example, if A, B, and C are available as pre- and post-intervention measures, and the pre-intervention difference in B requires statistical adjustment, the analysis of outcome B must adjust for B.

A review should clearly document if a study has a baseline difference in any of the following characteristics, since it could be evidence that the populations are drawn from different settings or that the intervention and comparison groups are not sufficiently comparable for the purposes of this review:

- Socioeconomic status (SES)
- Racial/ethnic breakdown
- School location (e.g., urban, rural)

The provision of such information, however, is not a requirement of the review.

Outcomes

The *WWC Procedures and Standards Handbook* discusses types of outcomes, criteria the outcomes must meet, and the methods by which outcomes are reported in Section III: Subsection B.4 (pp. 16–19). In this review, the requirements for outcome measures differs from guidance in the *WWC Procedures and Standards Handbook* in the following ways:

1. This review follows more stringent guidance with respect to reliability for student growth scores.
2. The use of a standardized test is not sufficient to establish the reliability of a student growth score; rather, authors also must show that the student growth scores meet reliability standards.
3. The onus for demonstrating that student growth scores meet reliability and validity standards rests with the authors.

For this review, measures obtained at the end of an intervention, as well as any time thereafter, are admissible. The review prioritizes immediate post-intervention findings that can inform intervention ratings and improvement indices. Measures occurring several months or years after the intervention may provide strong evidence for an intervention's effectiveness. Therefore, the review will include follow-up findings, when available and appropriate, in supplemental appendices to the report.

Statistical Adjustments

The *WWC Procedures and Standards Handbook* discusses the types of adjustments made by the WWC in Section IV: Subsection B Statistical Significance of Findings (p. 24).

Other Study Designs

Studies that use regression discontinuity or single-case designs are eligible for review using the appropriate pilot standards.

The *WWC Procedures and Standards Handbook* discusses the pilot standards for reviewing regression discontinuity design studies in Appendix D.

The *WWC Procedures and Standards Handbook* discusses the pilot standards for reviewing single-case design studies in Appendix E.

TABLE E1 Acceptable Pre-Intervention Measures by Outcome Domain

Outcome Domain	Acceptable Pre-Intervention Measures
Student achievement ^a	Achievement in the same subject as the outcome OR Achievement in the general achievement domain OR If the outcome is science, mathematics achievement OR If the outcome is social studies, reading achievement
Student progression	Grade level AND Student academic performance ^c AND One of the following: student race/ethnicity or degree of disadvantage ^c AND , if the unit of assignment is the school, a school-level measure of the student progression outcome
Student growth scores	The same measure as the outcome
Principal retention ^b	Principal experience AND Student academic performance ^c AND One of the following: student race/ethnicity or degree of disadvantage ^c AND , if the unit of assignment is the school, a school-level measure of the principal retention outcome

^aIncludes all five student achievement domains (mathematics, reading, science, social studies, and general).

^bIncludes all three principal retention domains (at the school, in the school district, and in the profession).

^cSee text for examples of acceptable measures of student academic performance and degree of disadvantage.

Process for Screening, Reviewing, and Summarizing Studies for Bush Institute Principal Talent Management Project^{xiv}

This document describes the process the review team will follow to screen, review studies, and summarize findings (see Figure 1).

Review of Individual Studies

The core element of the review process is the assessment of individual studies. The review of eligible studies against WWC standards serves as the basis for the final report that summarizes findings across studies. The review process has two steps: (a) an initial screening for eligibility and (b) a review of eligible studies using the review protocol developed specifically for this study and the *WWC Procedures and Standards Handbook* (version 3.0).

Initial Screening for Eligibility

Studies gathered during the literature search are screened against the parameters specified in the review protocol in order to identify a set of studies eligible for WWC review. A certified WWC reviewer (hereafter referred to as the first reviewer) conducts the initial screening for eligibility. According to the *WWC Procedures and Standards Handbook* (pp. 7-8), studies may be designated as *ineligible for WWC Review* for any of the following reasons:

- The study is not a primary analysis of the effect of an intervention.
- The study does not have an eligible design.
- The study does not use a sample aligned with the protocol.
- The study does not include an outcome within a domain specified in the protocol.
- The study was not published in the relevant time frame.

Review of Eligible Studies Against WWC Standards

The first reviewer will assess all studies that meet the initial screening criteria against the WWC standards. The first reviewer will then complete the Study Review Guide (a tool used by the WWC to record study characteristics and translate findings into standardized effect sizes) for each intervention. A second WWC-certified reviewer will review those SRGs and discuss any issues or problems with the first reviewer.^{xv}

The end result of reviewing a study against WWC standards is a study rating, which indicates the credibility of the supporting evidence. The three possible ratings are: Meets WWC Standards without Reservations, Meets WWC Standards with Reservations, and Does Not Meet WWC Standards.

Reporting on Findings

To the extent possible, the review team will report the magnitude and statistical significance of effect sizes for each study that meets WWC standards with or without reservations. The review team will then combine findings from individual studies into summative measures of effectiveness. These indicators will consist of the estimated degree of impacts, overall ratings of effectiveness, and a rating for the extent of evidence.

Report Findings for Individual Studies

The review will report findings for individual studies similar to those used in WWC reviews. Specifically, the review will report the magnitude of study findings in two ways—through effect sizes (i.e., standardized mean differences) and a WWC-calculated “improvement index”—as well as the statistical significance of the estimates. The *WWC Procedures and Standards Handbook* describes how these statistics are defined and calculated (pp. 22-25).

Using the estimated effect size and statistical significance level, findings from each study are classified into one of the following five categories: (a) statistically significant positive effect, (b) substantively important positive effect, (c) indeterminate effect, (d) substantively important negative (unfavorable) effect, and (e) statistically significant negative effect. The WWC criteria for selecting the categories that apply to each study based on a single or multiple outcome measures are presented in [Tables D2 and D3](#) respectively.

TABLE E2 WWC Characterization of Findings of an Effect Based on a Single Outcome Measure

Statistically significant positive effect	The estimated effect is positive and statistically significant (correcting for clustering when not properly aligned).
Substantively important positive effect	The estimated effect is positive but it is not statistically significant; yet, it is substantively important.
Indeterminate effect	The estimated effect is neither statistically significant nor substantively important.
Substantively important negative effect	The estimated effect is negative but it is not statistically significant; yet, it is substantively important.
Statistically significant negative effect	The estimated effect is negative and statistically significant (correcting for clustering when not properly aligned).

Source: The WWC Procedures and Standards Handbook, Table IV.1.

Synthesis of Findings Across Studies on the Same Intervention

If multiple studies of the same intervention (practice, policy, or program) meet WWC standards with or without reservations, the review team will combine findings across those studies, adopting a procedure similar to what the WWC uses to combine findings for intervention reports.^{xvi} Specifically, the WWC follows these steps to summarize findings across studies that examine the same intervention:

- Compute an average effect size and improvement index,
- Compute the statistical significance of the effect,
- Determine a rating for the intervention, and
- Determine the extent of evidence used to inform the intervention rating.

Each step is described in detail in the *WWC Procedures and Standards Handbook* (pp. 28-31). After computing an average effect size and the statistical significance of that effect in Steps a and b, the WWC uses a set of guidelines to determine the rating for an intervention (see [Table E4](#)). The WWC also reports on the extent of the evidence used to determine the intervention rating. The extent of evidence has two categories: (a) medium to large and (b) small, which is determined based on the number and sizes of studies (see [Table E5](#)).

Depending on the number and nature of studies that meet the WWC standards with or without reservations, the review team may decide to combine findings from studies that address similar practices, programs, or policies under the same area of principal talent management. The same procedures as those described above will be used to combine findings across policies or practices in the same area of principal talent management.

TABLE E3 WWC Characterization of Findings of an Effect Based on Multiple Outcome Measures

Statistically significant positive effect	<p>When any of the following is true:</p> <ol style="list-style-type: none"> 1. Univariate statistical tests are reported for each outcome measure and: <ul style="list-style-type: none"> ● At least half of the effects are positive and statistically significant—and no effects are negative and statistically significant—in a properly aligned analysis <p>OR</p> <ul style="list-style-type: none"> ● At least one measure is positive and statistically significant—and no effects are negative and statistically significant, accounting for multiple comparisons (and correcting for clustering when not properly aligned). <ol style="list-style-type: none"> 2. The mean effect reported for the multiple outcome measures is positive and statistically significant (correcting for clustering when not properly aligned). 3. After a multivariate statistical test in a properly aligned analysis, the omnibus effect for all combined outcome measures is reported as positive and statistically significant.
Substantively important positive effect	The reported mean effect is positive but not statistically significant; yet, it is substantively important.
Indeterminate effect	The reported mean effect is neither statistically significant nor substantively important.
Substantively important negative effect	The mean effect reported is negative but not statistically significant; yet, it is substantively important.
Statistically significant negative effect	<p>When any of the following is true:</p> <ol style="list-style-type: none"> 1. Univariate statistical tests are reported for each outcome measure and either: <ul style="list-style-type: none"> ● At least half of the effects are negative and statistically significant—and no effects are positive and statistically significant—in a properly aligned analysis, <p>OR</p> <ul style="list-style-type: none"> ● At least one measure is negative and statistically significant—and no effects are positive and statistically significant, accounting for multiple comparisons (and correcting for clustering when not properly aligned). <ol style="list-style-type: none"> 2. The reported mean effect of multiple outcome measures is negative and statistically significant (correcting for clustering when not properly aligned). 3. After a multivariate statistical test in a properly aligned analysis, the omnibus effect for all outcome measures is reported as negative and statistically significant.

Source: The WWC Procedures and Standards Handbook, Table IV.2.

TABLE E4 Criteria Used to Determine the WWC Extent of Evidence for an Intervention

Medium to large	<ul style="list-style-type: none"> ● The domain includes more than one study, AND ● The domain includes more than one setting, AND ● The domain findings draw from at least 14 schools or principals across studies.
Small	<ul style="list-style-type: none"> ● The domain includes only one study, OR ● The domain includes only one setting, OR ● The domain findings draw from fewer than 14 schools or principals across studies.

Source: The WWC Procedures and Standards Handbook, Table IV.4.

TABLE E5 Criteria Used to Determine the WWC Rating of Effectiveness for an Intervention

<p>Positive effects: Strong evidence of a positive effect without overriding contrary evidence</p>	<ul style="list-style-type: none"> ● Two or more studies show statistically significant positive effects; at least one of those meets WWC group design standards without reservations <p>AND</p> <ul style="list-style-type: none"> ● None of the studies show statistically significant or substantively important negative effects.
<p>Potentially positive effects: Evidence of a positive effect without overriding contrary evidence</p>	<ul style="list-style-type: none"> ● At least one study shows statistically significant or substantively important positive effects, AND ● Fewer or the same number of studies show indeterminate effects than statistically significant or substantively important positive effects, AND ● None of the studies show statistically significant or substantively important negative effects.
<p>No discernible effects: No affirmative evidence of effects</p>	<ul style="list-style-type: none"> ● None of the studies show statistically significant or substantively important effects, either positive or negative.
<p>Mixed effects: Evidence of inconsistent effects</p>	<p>EITHER of the following two circumstances:</p> <ul style="list-style-type: none"> ● At least one study shows statistically significant or substantively important positive effects, AND ● At least one study shows statistically significant or substantively important negative effects, BUT no more than the number that show statistically significant or substantively important positive effects. <p>OR</p> <ul style="list-style-type: none"> ● At least one study shows statistically significant or substantively important effects, AND ● More studies show an indeterminate effect than statistically significant or substantively important effects.
<p>Potentially negative effects: Evidence of a negative effect without overriding contrary evidence</p>	<p>EITHER of the following two circumstances:</p> <ul style="list-style-type: none"> ● One study shows statistically significant or substantively important negative effects, AND ● None of the studies show statistically significant or substantively important positive effects. <p>OR</p> <ul style="list-style-type: none"> ● Two or more studies show statistically significant or substantively important negative effects; at least one of those has statistically significant or substantively important positive effects, AND ● More studies show statistically significant or substantively important negative effects than statistically significant or substantively important positive effects.
<p>Negative effects: Strong evidence of a negative effect with no overriding contrary evidence</p>	<ul style="list-style-type: none"> ● Two or more studies show statistically significant negative effects; at least one of those meets WWC group design standards without reservations, AND ● None of the studies show statistically significant or substantively important positive effects.

Source: The WWC Procedures and Standards Handbook, Table IV.3.



Principal Preparation

TABLE F1 Findings Included in the Effectiveness Rating for the ELA/Reading Achievement (After Year 2 of Intervention) Domain for the NYC Leadership Academy's (NYCLA) Aspiring Principals Program

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
CORCORAN, SCHWARTZ, & WEINSTEIN, 2012								
New York State ELA exam standardized scores	K-12 principals (propensity score matching sample; after Year 2 of intervention)	301 principals ^a	na	na	0.01	0.01	0	0.76
Domain average for ELA/reading achievement (Corcoran, Schwartz, & Weinstein, 2012)						0.01	0	Not statistically significant
Domain average for ELA/reading achievement across all studies						0.01	0	na

^aNumber of students not provided in the study report.

TABLE F2 Findings Included in the Effectiveness Rating for the ELA/Reading Achievement (After Year 3 of Intervention) Domain for the NYC Leadership Academy's (NYCLA) Aspiring Principals Program

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
CORCORAN, SCHWARTZ, & WEINSTEIN, 2012								
New York State ELA exam standardized scores	K-12 principals (propensity score matching sample; after Year 3 of intervention)	275 principals ^a	na	na	-0.01	-0.01	0	0.62
Domain average for ELA/reading achievement (Corcoran, Schwartz, & Weinstein, 2012)						-0.01	0	Not statistically significant
Domain average for ELA/reading achievement across all studies						-0.01	0	na

^aNumber of students not provided in the study report.

TABLE F3 Findings Included in the Effectiveness Rating for the ELA/Reading Achievement (After Year 4 of Intervention) Domain for the NYC Leadership Academy's (NYCLA) Aspiring Principals Program

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
CORCORAN, SCHWARTZ, & WEINSTEIN, 2012								
New York State ELA exam standardized scores	K-12 principals (propensity score matching sample; after Year 4 of intervention)	168 principals ^a	na	na	0.03	0.02	1	0.54
Domain average for ELA/reading achievement (Corcoran, Schwartz, & Weinstein, 2012)						0.02	1	Not statistically significant
Domain average for ELA/reading achievement across all studies						0.02	1	na

^aNumber of students not provided in the study report.

TABLE F4 Findings Included in the Effectiveness Rating for the ELA/Reading Achievement (After Year 5 of Intervention) Domain for the NYC Leadership Academy's (NYCLA) Aspiring Principals Program

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
CORCORAN, SCHWARTZ, & WEINSTEIN, 2012								
New York State ELA exam standardized scores	K-12 principals (propensity score matching sample; after Year 5 of intervention)	60 principals ^a	na	na	-0.07	-0.07	-3	0.27
Domain average for ELA/reading achievement (Corcoran, Schwartz, & Weinstein, 2012)						-0.07	-3	Not statistically significant
Domain average for ELA/reading achievement across all studies						-0.07	-3	na

^aNumber of students not provided in the study report.

TABLE F5 Findings Included in the Effectiveness Rating for the ELA/Reading Achievement (After 1-4 Years of Intervention) Domain for Four Selected Principal Preparation Programs (Bush Institute Principal Preparation Study, 2016)

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
BUSH INSTITUTE PRINCIPAL PREPARATION STUDY, 2016								
ELA exam standardized scores	District A, K-8 principals (1-4 years after the intervention)	17 principals (18,848 students)	na	na	na	-0.08	-3	0.04
ELA exam standardized scores	District B, K-8 principals (1-3 years after the intervention)	22 principals (12,804 students)	na	na	na	0.06	2	0.49
ELA exam standardized scores	District C, K-8 principals (1-4 years after the intervention)	41 principals (30,653 students)	na	na	na	0.03	1	0.33
ELA exam standardized scores	District D, K-8 principals (1-4 years after the intervention)	75 principals (86,682 students)	na	na	na	-0.05	-2	0.05
Domain average for ELA/reading achievement (BUSH INSTITUTE principal preparation study, 2016)						-0.03 ^a	-1	Not statistically significant
Domain average for ELA/reading achievement across all studies						-0.03	-1	na

^aThe average effect size is based on effect sizes that have been weighted by the sample size for each district.

TABLE F6 Findings Included in the Effectiveness Rating for the Mathematics Achievement (After Year 2 of Intervention) Domain for the NYC Leadership Academy's (NYCLA) Aspiring Principals Program

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
CORCORAN, SCHWARTZ, & WEINSTEIN, 2012								
New York State mathematics exam standardized scores	K-12 principals (propensity score matching sample; after Year 2 of intervention)	K-12 principals (propensity score matching sample; after Year 2 of intervention)	301 principals ^a	na	na	-0.02	-0.02	-1
Domain average for mathematics achievement (Corcoran, Schwartz, & Weinstein, 2012)						-0.02	-1	Not statistically significant
Domain average for mathematics achievement across all studies						-0.02	-1	na

^aThe average effect size is based on effect sizes that have been weighted by the sample size for each district.

TABLE F7 Findings Included in the Effectiveness Rating for the Mathematics Achievement (After Year 3 of Intervention) Domain for the NYC Leadership Academy’s (NYCLA) Aspiring Principals Program

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
CORCORAN, SCHWARTZ, & WEINSTEIN, 2012								
New York State mathematics exam standardized scores	K-12 principals (propensity score matching sample; after Year 3 of intervention)	275 principals ^a	na	na	-0.04	-0.04	-1	0.16
Domain average for mathematics achievement (Corcoran, Schwartz, & Weinstein, 2012)						-0.04	-1	Not statistically significant
Domain average for mathematics achievement across all studies						-0.04	-1	na

^aNumber of students not provided in the study report.

TABLE F8 Findings Included in the Effectiveness Rating for the Mathematics Achievement (After Year 4 of Intervention) Domain for the NYC Leadership Academy’s (NYCLA) Aspiring Principals Program

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
CORCORAN, SCHWARTZ, & WEINSTEIN, 2012								
New York State mathematics exam standardized scores	K-12 principals (propensity score matching sample; after Year 4 of intervention)	168 principals ^a	na	na	0.03	0.02	1	0.54
Domain average for mathematics achievement (Corcoran, Schwartz, & Weinstein, 2012)						0.02	1	Not statistically significant
Domain average for mathematics achievement across all studies						0.02	1	na

^aNumber of students not provided in the study report.

TABLE F9 Findings Included in the Effectiveness Rating for the Mathematics Achievement (After Year 5 of Intervention) Domain for the NYC Leadership Academy’s (NYCLA) Aspiring Principals Program

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
CORCORAN, SCHWARTZ, & WEINSTEIN, 2012								
New York State mathematics exam standardized scores	K-12 principals (propensity score matching sample; after Year 5 of intervention)	60 principals ^a	na	na	-0.03	-0.03	-1	0.63
Domain average for mathematics achievement (Corcoran, Schwartz, & Weinstein, 2012)						-0.03	-1	Not statistically significant
Domain average for mathematics achievement across all studies						-0.03	-1	na

^aNumber of students not provided in the study report.

TABLE F10 Findings Included in the Effectiveness Rating for the ELA/Mathematics Achievement (After 1-4 Years of Intervention) Domain for Four Selected Principal Preparation Programs (Bush Institute Principal Preparation Study, 2016)

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
BUSH INSTITUTE PRINCIPAL PREPARATION STUDY, 2016								
Math exam standardized scores	District A, K-8 principals (1-4 years after the intervention)	17 principals (18,904 students)	na	na	na	-0.06	-2	0.30
Math exam standardized scores	District B, K-8 principals (1-3 years after the intervention)	22 principals (18,269 students)	na	na	na	-0.11	-4	0.25
Math exam standardized scores	District C, K-8 principals (1-4 years after the intervention)	41 principals (31,156 students)	na	na	na	0.02	1	0.62
Math exam standardized scores	District D, K-8 principals (1-4 years after the intervention)	75 principals (88,708 students)	na	na	na	-0.02	-1	0.63
Domain average for math achievement (BUSH INSTITUTE principal preparation study, 2016)						-0.02 ^a	-1	Not statistically significant
Domain average for math achievement across all studies						-0.02	-1	na

^aAverage effect size is based on effect sizes that have been weighted by the sample size for each district.



Professional Learning

TABLE F11 Findings Included in the Effectiveness Rating for the ELA/Reading Achievement (After Year 3 of Intervention) Domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
JACOB ET AL., 2014								
Michigan Educational Assessment Program (MEAP) grade 3 reading score	Principals in schools serving grade 3 students (after Year 3 of intervention)	119 principals ^a	na	na	-0.45	-0.02	-1	0.56
MAEP grade 4 reading score	Principals in schools serving grade 4 students (after Year 3 of intervention)	115 principals ^a	na	na	-0.27	-0.01	0	0.79
MAEP grade 5 reading score	Principals in schools serving grade 5 students (after Year 3 of intervention)	109 principals ^a	na	na	.81	0.02	1	0.52
Domain average for ELA/reading achievement (Jacob et al., 2014)								
NUNNERY, ROSS, & YEN, 2010								
Proficient on Pennsylvania state reading assessment	Principals (elementary within-district matching sample after Year 3 of intervention)	72 principals / 10,463 students	0.74 (na)	0.74 (na)	0.00	0.01	0	0.57
Proficient on Pennsylvania state reading assessment	Principals (elementary out-of-district matching sample after Year 3 of intervention)	64 principals / 10,080 students	0.72 (na)	0.70 (na)	0.02	0.10	4	0.00
Proficient on Pennsylvania state reading assessment	Principals (middle school sample after Year 3 of intervention)	38 principals / 16,414 students	0.75 (na)	0.71 (na)	0.05	0.06	3	0.00
Proficient on Pennsylvania state reading assessment	Principals (high school sample after Year 3 of intervention)	28 principals / 5,569 students	0.60 (na)	0.62 (na)	-0.02	0.10	4	0.00
Domain average for ELA/reading achievement (Nunnery, Ross, & Yen, 2010)						0.07	3	Statistically significant
Domain average for ELA/reading achievement across all studies						0.07	3	Statistically significant

^aNumber of students not provided in the study report.

TABLE F12 Findings Included in the Effectiveness Rating for the ELA/Reading Achievement (After Year 4 of Intervention) Domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
NUNNERY ET AL., 2011								
Massachusetts Comprehensive Assessment Program ELA test score	Elementary, middle, or elementary-middle school principals (after Year 4 of intervention)	1,015 principals ^a	-0.46 (1.00)	-0.57 (1.00)	0.11	0.11	4	0.51
Domain average for ELA/reading achievement (Nunnery et al., 2011)						0.11	4	Not statistically significant
Domain average for ELA/reading achievement across all studies						0.11	4	na

^aThis study analyzes aggregated school-level outcomes.

TABLE F13 Findings Included in the Effectiveness Rating for the Mathematics Achievement (After Year 3 of Intervention) Domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
JACOB ET AL., 2014								
Michigan Educational Assessment Program (MEAP) grade 3 mathematics score	Principals in schools serving grade 3 students (after Year 3 of intervention)	119 principals ^a	na	na	0.78	0.04	2	0.34
MAEP grade 4 mathematics score	Principals in schools serving grade 4 students (after Year 3 of intervention)	115 principals ^a	na	na	-0.01	0.00	0	0.99
MAEP grade 5 mathematics score	Principals in schools serving grade 5 students (after Year 3 of intervention)	109 principals ^a	na	na	1.48	0.04	2	0.29
Domain average for mathematics achievement (Jacob et al., 2014)						0.03	1	Not statistically significant
NUNNERY, ROSS, & YEN, 2010								
Proficient on Pennsylvania state mathematics assessment	Principals (elementary within-district matching sample; after Year 3 of intervention)	72 principals / 10,463 students	0.82 (na)	0.82 (na)	0.00	0.08	3	0.00
Proficient on Pennsylvania state mathematics assessment	Principals (elementary out-of-district matching sample, after Year 3 of intervention)	64 principals / 10,080 students	0.81 (na)	0.78 (na)	0.03	0.13	5	0.00
Proficient on Pennsylvania state mathematics assessment	Principals (middle school sample, after Year 3 of intervention)	38 principals / 16,414 students	0.76 (na)	0.69 (na)	0.08	0.13	5	0.00
Proficient on Pennsylvania state mathematics assessment	Principals (high school sample, after Year 3 of intervention)	28 principals / 5,569 students	0.52 (na)	0.44 (na)	0.08	0.16	6	0.00
Domain average for mathematics achievement (Nunnery, Ross, & Yen, 2010)						0.13	5	Statistically significant
Domain average for mathematics achievement across all studies						0.08	3	na

^aNumber of students not provided in the study report.

TABLE F14 Findings Included in the Effectiveness Rating for the Mathematics Achievement (After Year 4 of Intervention)

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
NUNNERY ET AL., 2011								
Massachusetts Comprehensive Assessment Program mathematics test score	Elementary, middle, or elementary-middle school principals (after Year 4 of intervention)	1,015 principals ^a	-0.38 (1.00)	-0.52 (1.00)	0.14	0.14	6	0.40
Domain average for mathematics achievement (Nunnery et al., 2011)						0.14	6	Not statistically significant
Domain average for mathematics achievement across all studies						0.14	6	na

TABLE F15 Findings Included in the Effectiveness Rating for the Principal Retention Domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
JACOB ET AL., 2014								
Retention at the school	Whole sample (after Year 3 of intervention)	122 principals	na	na	-0.16	na	na	0.05 < p ≤ .010
Domain average for principal retention (Jacob et al., 2014)						na	na	Not statistically significant
Domain average for principal retention across all studies						na	na	na

Note: The study report does not provide enough information to allow for calculation of the effect size



Compensation and Incentives

TABLE F16 Findings Included in the Effectiveness Rating for the ELA/Reading Achievement (After Year 1 of Intervention) Domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
CHIANG ET AL., 2015								
ELA/Reading test scores	Principals in districts that implemented the intervention for two years (after Year 1 of intervention)	134 principals / 40,576 students	-0.37 (1.00)	-0.40 (1.00)	0.03	0.02	1	0.04
Domain average for ELA/reading achievement (Chiang et al, 2015)						0.02	1	Statistically significant
Domain average for ELA/reading achievement across all studies						0.02	1	na

TABLE F17 Findings Included in the Effectiveness Rating for the ELA/Reading Achievement (After Year 2 of Intervention) Domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
CHIANG ET AL., 2015								
ELA/Reading test scores	Principals in districts that implemented the intervention for two years (after Year 2 of intervention)	134 principals / 40,391 students	-0.36 (1.00)	-0.39 (1.00)	0.03	0.02	1	0.03
Domain average for ELA/reading achievement (Chiang et al, 2015)						0.02	1	Statistically significant
Domain average for ELA/reading achievement across all studies						0.02	1	na

TABLE F18 Findings Included in the Effectiveness Rating for the Math Achievement (After Year 1 of Intervention) Domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
CHIANG ET AL., 2015								
Mathematics test scores	Principals in districts that implemented the intervention for two years (after Year 1 of intervention)	134 principals / 40,852 students	-0.43 (1.00)	-0.45 (1.00)	0.02	0.02	1	
Domain average for ELA/reading achievement (Chiang et al, 2015)						0.02	1	Not statistically significant
Domain average for math achievement across all studies						0.02	1	na

TABLE F19 Findings Included in the Effectiveness Rating for the Math Achievement (After Year 1 of Intervention) Domain for the Principal Compensation and Incentives Component





Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
CHIANG ET AL., 2015								
Mathematics test scores	Principals in districts that implemented the intervention for two years (after Year 2 of intervention)	134 principals / 40,759 students	-0.39 (1.00)	-0.43 (1.00)	0.04	0.03	1	0.07
Domain average for ELA/reading achievement (Chiang et al, 2015)						0.03	1	Not statistically significant
Domain average for math achievement across all studies						0.03	1	na




TABLE F20 Findings Included in the Effectiveness Rating for the Principal Retention (After Year 1 of Intervention) Domain for the Principal Compensation and Incentives Component

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
CHIANG ET AL., 2015								
Retention at the school	Principals in districts that implemented the intervention for two years (after Year 2 of intervention)	134 principals	0.80 (na)	0.73 (na)	0.07	0.24	9	0.27
Domain average for principal retention (Chiang et al., 2015)						0.24	9	Not statistically significant
Domain average for principal retention across all studies						0.24	9	na

TABLE F21 Findings Included in the Effectiveness Rating for the Principal Retention (After Year 2 of Intervention) Domain for the Principal Compensation and Incentives Component

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
CHIANG ET AL., 2015								
Retention at the school	Principals in districts that implemented the intervention for two years (after Year 2 of intervention)	134 principals	0.67 (na)	0.58 (na)	0.09	0.23	9	0.36
Domain average for principal retention (Chiang et al., 2015)						0.23	9	Not statistically significant
Domain average for principal retention across all studies						0.23	9	na

ESSA Definition	WWC Characterization			
	Quality of evidence for a single study		Quality of evidence for an intervention	
An “evidence-based” activity, strategy, or intervention demonstrates a statistically significant effect on student outcomes or other relevant outcomes	Ratings for a single study (Determined based on review of a single study against WWC standards)	Characterization of findings (Determined for each study that meets WWC standards with or without reservations)	Rating of effectiveness (Determined by combining findings from multiple studies that meet WWC standards with or without reservations)	Characterization of the extent of evidence (Determined based on the number of studies that meet WWC standards with or without reservations, the number of settings, and total sample sizes).
Strong evidence from at least 1 well designed and well-implemented experimental study	 Meets WWC Group Design Standards without reservations: <ul style="list-style-type: none"> Well designed and well-implemented randomized control trials (studies with low sample attrition and no confounding factors) 	<ul style="list-style-type: none"> Statistically significant positive effect Substantively important positive effect Indeterminate effect Substantively important negative effect Statistically significant negative effect 	<ul style="list-style-type: none"> Positive effects (Strong evidence of a positive effect without overriding contrary evidence) Potentially positive effects (Evidence of a positive effect without overriding contrary evidence) No discernible effects (No affirmative evidence of effects) Mixed effects (Evidence of inconsistent effects) Potentially negative effects (Evidence of a negative effect without overriding contrary evidence) Negative effects (Strong evidence of a negative effect without overriding contrary evidence) 	<ul style="list-style-type: none"> Medium to large: Includes more than one study AND more than one setting AND a total sample size of at least 350 students or 14 classrooms Small: Includes only one study OR one setting OR a total sample size that does not exceed 350 students or 14 classrooms
Moderate evidence from at least 1 well designed and well-implemented quasi-experimental study	 Meets WWC Group Design Standards with reservations: <ul style="list-style-type: none"> Randomized control trials that have high sample attrition but demonstrate baseline equivalence and have no confounding factors. Well designed and well-implemented quasi-experimental studies (studies that demonstrate baseline equivalence and have no confounding factors) 			
Promising evidence from at least 1 well designed and well-implemented correlational study with statistical controls for selection bias	 Does not meet WWC group design standards	N/A	N/A	N/A
Demonstrates a rationale, based on high quality research findings or positive evaluation, that a activity, strategy, or intervention will likely improve student outcomes or other relevant outcomes; and includes ongoing efforts to examine the effects of such activity, strategy, or intervention	 Does not meet WWC group design standards	N/A	N/A	N/A

Key  Meets standards without reservations  Meets standards with reservations  Does not meet standards

Notes: WWC has developed pilot standards for regression continuity designs (one type of quasi-experimental design) and pilot standards for single-case designs (one type of experimental design). Studies using those two designs can be rated with the following characterizations: meet pilot WWC standards without reservations, meet pilot WWC standards with reservations, and does not meet pilot WWC standards. However, those standards are applied only to evidence from individual studies. Findings from studies with those designs are not incorporated into reports that combine findings across studies.

Source: US Department of Education (2016) ^{xvii}

Complete List of All Studies Reviewed for the Initial Eligibility Screening

- ACTION UNITED Education Fund. (2012). *Revolving doors: Findings from Philadelphia's highest-poverty schools*. Philadelphia, PA: Author.
- Augustine, C. H., Gonzalez, G., Ikemoto, G. S., Russel, J., Zellman, G. L., Constant, L., & Dembosky, J. W. (2009). *Improving school leadership: the promise of cohesive leadership systems*. Santa Monica, CA: RAND Corporation.
- Baker, B.D., Punswick, E., & Belt, C. (2010). School leadership stability, principal moves, and departures: Evidence from Missouri. *Educational Administration Quarterly*, 46(4).
- Barnes, C.A., Camburn, E., Sanders, B.R., & Sebastian, J. (2010). Developing instructional leaders: Using mixed methods to explore the black box of planned change in principals' professional practice. *Educational Administration Quarterly*, 46(2), 241-279.
- Braun, D., Gable, R., & Kite, S. (2011). Relationship among essential leadership preparation practices and leader, school, and student outcomes in K-8 schools. *International Journal of Educational Leadership Preparation*, 6(3), 1-21.
- Briggs, K., Rhines Cheney, G., Davis, J., & Moll, K. (2012). *Operating in the dark: what outdated state policies and data gaps mean for effective school leadership*. Houston, TX: 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 Corporation.
- Camburn, E. M., Goldring, E., May, H., Supovitz, J., Barnes, C., & Spillane, J. P. (2007, April). *Lessons learned from an experimental evaluation of a principal professional development program*. Paper presented at the American Educational Research Association, Washington, DC.
- Clark, D., Martorell, P., & Rockoff, J. (2009). *School Principals and School Performance* (Working Paper 38). Washington, D.C: Center for Analysis of Longitudinal Data in Education Research.
- Corcoran, S. P., Schwartz, A. E., & Weinstein, M. (2009). *The New York City aspiring principals program: A school-level evaluation*. New York: Institute for Education and Social Policy.
- Corcoran, S. P., Schwartz, A. E., & Weinstein, M. (2012). Training your own: The impact of New York City's aspiring principals program on student achievement. *Educational Evaluation and Policy Analysis*, 34(2), 232-253.
- Creighton, T.B., & Jones, G.B. (2001, August) *Selection or self-selection? How rigorous are our selection criteria for education administration preparation programs?* Paper presented at the 2001 conference of the National Council of Professors of Educational Administration, Houston, TX.
- Donmoyer, R., Yennire-Donmoyer, J., & Galloway, F. (2012). The search for connections across principal preparation, principal performance, and student achievement in an exemplary principal preparation program. *Journal of Research in Leadership Education*, 7(1), 5-43.
- Friedman, B. A., Friedman, M. A., & Markow, D. (2008). Predictors of principals' satisfaction with their schools. *Journal of Educational Administration*, 46(5), 598-612.
- Fuller, E., Young, M., & Baker, B. D. (2011). Do principal preparation programs influence student achievement through the building of teacher-team qualifications by the principal? An exploratory analysis. *Educational Administration Quarterly*, 47(1), 173-216.
- Fuller, M. & Young, E. (2009). *Tenure and retention of newly hired principals in Texas*. Paper presented at the University Council for Educational Administration, Austin, TX.
- Fuller, M. & Young, E. (2009, April). *Tenure and retention of newly hired principals in Texas*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.
- Furgeson, J., Knechtel, V., Sullivan, M., Tuttle, C.C., Akers, L., Anderson, M.A., and Nichols-Barrer, I. (2014). *KIPP leadership practices through 2010-2011: Technical report*. Mathematica Policy Research: Cambridge, MA.
- Gates, S., et al. (2003). *Who is Leading Our Schools? An Overview of School Administrators and Their Careers*. Santa Monica, CA: RAND Corporation.
- Gates, S. M., Hamilton, L. S., Martorell, P., Burkhauser, S., Heaton, P., Pierson, A., & Gu, K. (2014). *Preparing principals to raise student achievement: Implementation and effects of the New Leaders program in ten districts*. Santa Monica, CA: Rand Corporation.
- George W. Bush Institute. (2016). *Following the Leaders: An Analysis of Graduate Effectiveness From Five Principal Preparation Programs*. Dallas: Bush Institute.
- Hamilton, L.S., Engberg, J., Steiner, E.D., Nelson, C.A., & Yuan, K. (2012). *Improving school leadership through support, evaluation, and incentives: The Pittsburgh Principal Incentive Program*. Santa Monica, CA: RAND Corporation.
- Hemphill, K., & Nauer, C. (2010). *Managing by the numbers: Empowerment and accountability in New York City's schools*. New York, NY: Center for New York City Affairs, The New School for Management and Urban Policy.
- Herman, R., Gates, S., Chavez-Herrerias, E. & Harris, M. (2015). *School Leadership Interventions under the Every Student Succeeds Act (ESSA): Volume I – A Review of the Evidence Base, Initial Findings*. Santa Monica, CA: RAND Corporation.
- Jacob, R., Goddard, R., Kim, M., Miller, R., & Goddard, Y. (2014) Exploring the causal impact of the McREL Balanced Leadership Program on leadership, principal efficacy, instructional climate, educator turnover, and student achievement. *Educational Evaluation and Policy Analysis*, 52, 187-220.
- Kaplan, L. S., Owings, W. A., & Nunnery, J. (2005). Principal quality: A Virginia study connecting interstate school leaders licensure consortium standards with student achievement. *NASSP Bulletin*, 89(643), 28-44.
- Kimball, S.M., Milanowski, T., & McKinney, S.A. (2007, April). *Implementation of standards-based principal evaluation in one school district: First year results from randomized trial*. Paper presented at the annual conference of the American Educational Research Association, Chicago, IL.

- Knoepfel, R. C., & Rinehart, J. S. (2008). Student achievement and principal quality: Explaining the relationship. *Journal of School Leadership, 18*(5), 501–527.
- Kwan, Y.K.P., & Walker, A.D. (2009) Are we looking through the same lens? Principal recruitment and selection. *International Journal of Educational Research, 48*(1), 51–61.
- Leithwood, K. (2011). *Characteristics of high performing school systems in Ontario: Part I*. Ontario, Canada: Institute for Educational Leadership.
- Max, J., Constantine, J., Wellington, A., Hallgren, K., Glazerman, S., Chiang, H., & Speroni, C. (2014). *Evaluation of the teacher incentive fund: Implementation and early impacts of pay-for-performance after one year*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- May, H., Huff, J., & Goldring, E. (2012). A longitudinal study of principals' activities and student performance. *School Effectiveness and School Improvement, 23*(4), 417–439.
- Newton, R., Giesen, J., Freeman, J., Bishop, H., & Zeitoun, P. (2003). Assessing the reactions of males and females to attributes of the principalship. *Educational Administration Quarterly, 39*, 504–532.
- Nunnery, J. A., Ross, S. M., & Yen, C. J. (2010). *The effect of the National Institute for School Leadership's executive development program on school performance trends in Pennsylvania*. Online Submission.
- Nunnery, J.A., Ross, S.M., Chappell, S., Pribesh, S., Hoag-Carhart, E. (2011). *The impact of the National Institute for School Leadership's Executive Development Program on school performance in Massachusetts: Cohort 2 results*. Online submission.
- Orphanos, S., & Orr, M. T. (2014). Learning leadership matters: The influence of innovative school leadership preparation on teachers' experiences and outcomes. *Educational Management Administration & Leadership, 42*(5), 680–700.
- Orr, M. T. & Orphanos, S. (November 2011). How graduate-level preparation influences the effectiveness of school leaders: A comparison of the outcomes of exemplary and conventional leadership preparation programs for principals. *Educational Administration Quarterly, 47*, 18–70.
- Owings, W. A., Kaplan, L. S., & Nunnery, J. (2005). Principal quality, ISLLC standards, and student achievement: A Virginia study. *Journal of School Leadership, 15*(1), 99–119.
- Papa, F. C., Jr., Lankford, H., & Wyckoff, J. (2002). *The attributes and career paths of principals: Implications for improving policy*. Albany, NY: Teacher Policy Research.
- Papa, F. Jr., (2007). Why do principals change schools? A multivariate analysis of principal retention. *Leadership and Policy in Schools, 6*(3), 267–290.
- Pijanowski, J.C., & Brady, K.P. (2009). The influence of salary in attracting and retaining school leaders. *Education and Urban Society, 42*(1), 25–41.
- Porter, A.C., Polikoff, M.S., Goldring, E. B., Murphy, J., Elliot, S. N., & May, H. (2010, December). Investigating the validity and reliability of the Vanderbilt Assessment of Leadership in Education. *The Elementary School Journal, 111*(2), 282–313.
- Pounder, D. G., & Merrill, R. J. (2001). Job desirability of the high school principalship: A job choice theory perspective. *Educational Administration Quarterly, 37*(1), 27–57.
- Pulliam, C.L., LaCaria, L., Schoeneberger, J., & Algozzine, B. (2014). A preliminary analysis of a strategic staffing initiative. *Educational Policy, 28*(4), 578–603.
- Rammer, R. A. (2007). Call to action for superintendents: Change the way you hire principals. *Journal of Educational Research, 101*(2), 67–76.
- Replogle Sheppard, R. (2010, May). *Determining factors that influence high school principal turnover over a five-year period* (Unpublished doctoral dissertation). University of North Texas, Denton, TX.
- Ross, C., Herrmann, M., & Angus, M.H. (2015). *Measuring principals' effectiveness: Results from New Jersey's principal evaluation pilot*. Washington, D.C.: Institute of Education Sciences.
- Silva, J. P., White, G. P., & Yoshida, R. K. (2011). The direct effects of principal-student discussions on eighth grade students' gains in reading achievement: An experimental study. *Educational Administration Quarterly, 47*(5), 772–793.
- Stoelinga, S. R., Hart, H., & Schalliol, D. (2008). *The work of Chicago Public Schools' principals: Leading in a complex context with high stakes (Research Report)*. Chicago, IL: Consortium on Chicago School Research.
- Teh, B., Chiang, H., Lipscomb, S., Gill, B. (2014). *Measuring school leaders' effectiveness: An interim report from a multiyear pilot of Pennsylvania's Framework for Leadership*. Washington, D.C.: Institute of Education Sciences.
- Vanderhaar, J. E., Munoz, M. A., & Rodosky, R. J. (2006). Leadership as accountability for learning: The effects of school poverty, teacher experience, previous achievement, and principal preparation programs on student achievement. *Journal of Personnel Evaluation in Education, 19*(1–2), 17–33.
- VanLwaarden, A. (2011). *Principal ProComp and retention: 2011 report brief*. Boulder, CO: University of Colorado at Boulder School of Education.
- Weinstein, M., Schwartz, A. E., Jacobowitz, R., Ely, T., & Landon, K. (2009, September). *New schools, new leaders: A study of principal turnover and academic achievement at new high schools in New York City*. New York, NY: NYU Wagner School of Public Service.

The George W. Bush Institute at the George W. Bush Presidential Center
2943 SMU Boulevard | Dallas, Texas 75205 | 214.200.4300
www.bushcenter.org | educationreform@bushcenter.org

