



REVIEW OF *PREPARING PRINCIPALS TO RAISE STUDENT ACHIEVEMENT*

Reviewed By

Edward J. Fuller

Penn State University

June 2014

Summary of Review

RAND's recent evaluation attempts to determine the New Leaders principal preparation program's impact on student test scores, concluding that New Leader principals are slightly more effective, albeit only for certain grade levels, subject areas, and districts. But the study's overall conclusion is problematic for at least eight reasons. First, the effect sizes are quite small—less than 2 percentile points in the lower grades, where the data allow for more reliable analyses—and the study's results are more mixed than its bottom-line conclusion would suggest. Indeed, most of the results find no statistically significant impact of New Leader principals on student test scores, and there were nearly as many negative findings as positive ones. Also, the study fails to acknowledge the extant research on the limitations of the study's value-added model (VAM) with school fixed effects approach. Additionally, problems include a failure to control for the same variables in each VAM, for peer effects, or for the scores of student stayers and leavers. The study also failed to discuss the potential impact of changes in the way student achievement was measured in the districts or to explore how the distribution of school-level scores might have influenced the results. Yet in the face of these problems and limitations, this study implies—incorrectly—that it can offer a valid causal determination that the New Leaders principals had a small positive effect on student test scores.

Kevin Welner

Project Director

William Mathis

Managing Director

Erik Gunn

Managing Editor

National Education Policy Center

School of Education, University of Colorado
Boulder, CO 80309-0249
Telephone: (802) 383-0058

Email: NEPC@colorado.edu
<http://nepc.colorado.edu>

Publishing Director: Alex Molnar



This is one of a series of Think Twice think tank reviews made possible in part by funding from the Great Lakes Center for Education Research and Practice. It is also available at <http://greatlakescenter.org>.

This material is provided free of cost to NEPC's readers, who may make non-commercial use of the material as long as NEPC and its author(s) are credited as the source. For inquiries about commercial use, please contact NEPC at nepc@colorado.edu.

REVIEW OF *PREPARING PRINCIPALS TO RAISE STUDENT ACHIEVEMENT: IMPLEMENTATION AND EFFECTS OF THE NEW LEADERS PROGRAM IN TEN DISTRICTS*

Edward J. Fuller, Penn State University

I. Introduction

Over the past decade, researchers and policymaker have increasingly focused on the importance of principals in improving student outcomes, particularly as measured by changes in student test scores. A small but growing body of research, in fact, concludes that principals have an important impact on student test scores independent of the other factors affecting achievement.¹

This increased interest in school leader effectiveness has sparked attentiveness to the quality of the programs that prepare school leaders.² For example, organizations like the National Conference of State Legislatures³ and the Center for American Progress⁴ have called for the creation and adoption of principal preparation program (PPP) accountability systems to mirror those being created for teacher preparation programs. While such calls do not focus on test-based accountability exclusively, the impact of PPP graduates on K-12 students' test scores is a primary component of many proposed systems.

Issues associated with a lack of data and inadequate statistical methods, however, have made the evaluation of PPPs exponentially more difficult than evaluations of teacher preparation programs.⁵ Undeterred by both the paucity of extant research on how to best evaluate PPPs and the methodological weaknesses of existing studies, several states (including Florida, Ohio, Massachusetts, and New York) have developed PPP evaluation systems that use metrics such as graduates' effectiveness in improving student achievement, principal placement rates, and the retention rates of graduates in leadership positions as indicators of program efficacy.

A recently released RAND report, *Preparing Principals to Raise Student Achievement: Implementation and Effects of the New Leaders Program in Ten Districts*,⁶ takes on the monumental task of evaluating the effectiveness of the New Leaders (NL) principal preparation program. The New Leaders program, a non-profit organization founded in 2000 by a team of social entrepreneurs and based in New York City, says its mission is to improve student outcomes by preparing effective leaders and improving the working conditions of principals.⁷

The RAND evaluation relied on the work of a large team of researchers, including Susan Gates, Laura Hamilton, Paco Martorell, Susan Burkhauser, Paul Heaton, Ashley Pierson, Matthew Baird, Mirka Vuollo, Jennifer J. Li, Diana Lavery, Melody Harvey, and Kun Gu. The evaluation was based primarily on a statistical analysis of student achievement data. In addition, the evaluators collected qualitative data through surveys of principals and case studies of beginning principals.

II. Findings and Conclusions of the Report

The report's primary conclusion is that "students who attended schools led by New Leaders principals experienced slightly larger achievement gains on average than similar students in schools led by non-New Leaders principals" (p. xviii). This over-arching conclusion is divided into findings for the lower grades (grades 3 through 8) and high schools. With respect to the lower grades, the study asserts,

spending three or more years in a school with a New Leaders principal was associated with achievement gains that translate into a change of 0.7 to 1.3 percentile points for a typical student in mathematics and reading (p. xviii).

At the high school level, the study claims,

students in schools where the New Leaders principal had three or more years of tenure experienced gains in reading achievement of about 3 percentile points in reading but no significant difference in mathematics (p. xviii).

The report does list some limitations of the effort. In particular, it cautions about drawing conclusions about the efficacy of the program because New Leaders principals were more likely to remain in their positions for three or more years and research suggests principal tenure is positively associated with student outcomes.⁸

III. The Report's Rationale for Its Findings and Conclusions

The conclusions about the effectiveness of New Leaders principals are based on a relatively sophisticated value-added model (VAM) analysis of student test scores in 10 districts employing at least five New Leaders principals. Specifically, based on the VAM results pooled across grade levels and tests within each site and then pooled again across sites, the study claims that the small and positive effect sizes indicate that principals prepared by New Leaders are slightly more effective than other principals in these districts.

Yet, not only were the effect sizes quite small, there were some statistically significant negative findings; moreover, there were as many districts with negative NL impacts as there were with positive NL impacts. The results, then, were not very consistent across years, grade levels, and sites, and the positive results were very small.

Further, the results are based entirely on the unexamined assumption that the statistical approach employed in the study yields accurate estimates of principal effectiveness. As will be discussed below, extant research⁹ concludes that no known statistical approach can accurately identify the independent effect of principals on test scores apart from the myriad number of other factors that influence test scores.

IV. The Report's Use of Research Literature

Despite the six pages of references on whether principals influence student outcomes and the avenues through which principals influence student outcomes, this study ignores the available research on the characteristics of effective principal preparation programs, evaluations of efforts to estimate principal effectiveness, and studies related to the evaluation of PPPs. These are critical omissions—particularly with respect to the evaluation of principal effectiveness—because the literature in these areas goes to the very heart of the study. Indeed, while the research base on evaluating efforts to estimate principal effectiveness on test scores is not particularly voluminous, all of the existing studies¹⁰ convincingly argue that current statistical approaches cannot disentangle the impact of principals on changes in student test scores from other factors. Thus, such estimates are not particularly accurate.

V. Review of the Report's Methods

At the lower grade level, the statistical methods employed in this study are superior to most evaluations of principal or PPP effectiveness. At the high school level, the cross-sectional nature of the analysis makes the approach particularly problematic since longitudinal data from the same cohort of students is necessary to accurately estimate changes in student outcomes.

Both sets of analyses employed a number of commonly used control variables at both the student- and school-levels. Importantly, both sets of analyses controlled for the unobserved characteristics of students and schools through student and school fixed effects, respectively. There is widespread consensus that estimates of principal effectiveness should employ a school fixed effects approach since important unobserved school characteristics likely influence student outcomes.¹¹

Essentially, 10 different student-level VAMs were employed at the lower grade level, and the results were aggregated to the school level within each site. In general, these VAMs controlled for student characteristics such as participation in the federal free and reduced-price lunch program (FRPL), English as a Second Language (ESL) or Limited English Proficiency (LEP) status, participation in special education, and being over-aged relative to one's peers. School characteristics generally included the average percentages of FRPL, ESL, LEP, special education, White, African American, Hispanic, and Asian students as

well as the number of students in the school, indicators of special types of schools (charter or magnet), and the start-up year of the school. Finally, all 10 analyses of lower grades included grade effects, school-year effects, and school-level effects.

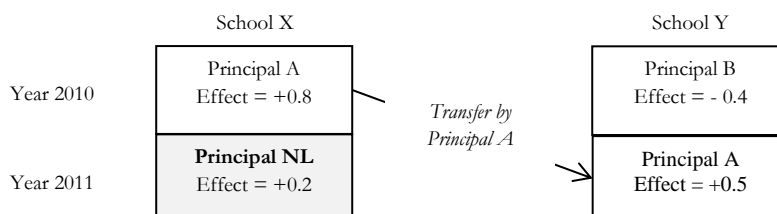
Only five sites had enough New Leaders principals and available data to be included in the high school analyses. As with the lower grade analyses, the high school analyses employed separate VAMs for each site. These VAMs generally included the same student and school control variables as the VAMs at the lower grade level.

While this study arguably uses the most appropriate statistical approach to estimating a principal's effect on test scores, evaluation of such efforts indicate that the final comparison of NL principals with other principals rests upon the idiosyncratic nature of the small number of relationships between principals at a given point in time and the principals employed in the same school or in a school within the principal's connected network of school at another point in time.¹²

First, each principal is compared with only a very small set of other principals. Small comparison groups result in comparisons that do not yield accurate estimates of principal effectiveness.

Specifically, as the study notes, the use of school fixed effects forces the estimates of the effectiveness of principals employed at a given point in time to be compared with the estimates of effectiveness of two sets of principals: first, the principals immediately preceding or following the principal in the same school and, second, the set of principals connected to these preceding and following principals employed in other schools.

Under this approach, the comparison group for a principal would necessarily be a small number of other individuals—often only one or two other principals.¹³ For example, as



shown in Figure 1, let us assume Principal A worked in School X and then transferred from School X to School Y to take the place of Principal B while Principal NL—a newly hired New Leaders graduate—replaced Principal A in School X.

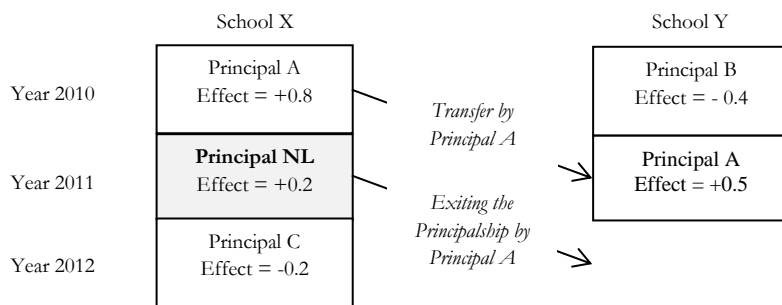
Figure 1. Connected Network of Comparison Principals in Two Schools

In this scenario, the estimates of the effectiveness of Principal NL could be compared with the estimates of Principals A and B, but not with any other principals. The effectiveness of Principal B, thus, would be largely determined by whether Principals A and B were effective or ineffective. For example, Principal NL, with an estimated effectiveness of +0.2,

might be considered to have average effectiveness because Principal A was more effective (+0.8 and +0.5) while Principal B was less effective (-0.4). But, if Principal A had negative estimates of effectiveness in both schools, then Principal NL might be considered an effective principal. *Thus, even if Principal NL's true effectiveness was the same in both scenarios, the estimates of Principal NL's effectiveness would be determined in large part by the effectiveness of Principal A.* The point here is that estimates of principal effectiveness would be based on arbitrarily connected relationships among principals and schools rather than a systemic estimate across all principals and schools.

In districts with a longer time frame, the connected network of schools and principals might be larger, thus increasing the comparison group of principals. For example, instead of only being compared with Principals A and B as in Figure 1, Principal NL would now be compared with Principals A, B, and C.

If we assume that the average principal remained at her or his school for three years, then the greatest number of principals employed at the same school as the New Leaders



principal would be three. Thus, the estimates of the New Leaders principal would be largely determined by the effectiveness of these other three principals. Even if all three of these other principals transferred from their original schools to different schools within the same district, the number of comparison principals might expand to 6 or even 8 other

Figure 2. Connected Network of Comparison Principals in Two Schools with a Longer Time Frame

principals. But, as Chiang, *et al.* found in Pennsylvania, most comparisons would include only 1 or 2 other principals.¹⁴ The RAND study did not document the number of principals involved in the comparisons for each district or for the pooled results; nor did the study even discuss this issue.

This approach to estimating principal effectiveness is problematic for several reasons. First, each principal is compared with only a very small set of other principals. Thus, the effectiveness of the prior and subsequent principals greatly influences the estimate of effectiveness of the newly hired NL principals and newly hired non-NL principals. Such small comparison groups result in comparisons that do not yield accurate estimates of principal effectiveness.¹⁵

Second, and related to the first issue, the length of tenure of the principal immediately preceding a newly hired principal could influence the estimate of the effectiveness of the principal under study.¹⁶ For example, let us assume the tenure of a principal preceding a

NL principal was greater than the tenure of the principal immediately preceding a non-NL principal. Given that the effectiveness of principals generally increases with tenure at a school¹⁷ and principals following principals with extended lengths of tenure often benefit from the policies, procedures, and systems implemented by the prior principal,¹⁸ the NL principal could appear more effective than the non-NL principal simply because the NL principal landed in a more advantageous position than the non-NL.¹⁹ The study did not control for the different lengths of tenure of the principals preceding the newly hired principals. Even if the researchers had wanted to control for this factor, the report's appendix suggests such data was simply not always available.

The RAND study simply fails to explain these details of its analytic approach, thus leaving the reader in the dark about the small and uncontrolled nature of the comparisons.

VI. Review of the Validity of the Findings and Conclusions

While the statistical analyses in this study are similar to the best available analyses of principal effectiveness and are superior to most efforts to evaluate PPP effectiveness, there are a number of serious issues that further render the study highly problematic.

Its most glaring and important flaws are the failure to adequately explain the details of the analyses and the failure to include a review of the literature on estimating principal effectiveness. If the study had faithfully reviewed the available literature, the study would have recognized the severe limitations of the approach employed; it also would acknowledge that the estimates used were inaccurate indicators of a principal's influence on test scores.

For example, Grissom, Kalgoides, and Loeb (p. 22) conclude their evaluation of statistical efforts to estimate principal effectiveness by stating,

it is important to think carefully about what the measures are revealing about each principal's contribution and to use the measures for what they are, which is *not* as a clear indicator of principals' specific impact on student test score growth.²⁰

Branch, Hanushek, and Rivken as well as Chiang, Lipscomb, and Gill arrive at a similar conclusion—that the estimates of effectiveness are not particularly accurate. Indeed, Chiang, *et al.* (p. 26) contend that if principal evaluations include principals with less than three years of experience (as this study does), then “using measures of school effectiveness as the basis for [principal] evaluations is an invalid method for gauging the principals' true effectiveness.”²¹

As noted, the study found very small effect sizes and inconsistent results across grade levels and districts, yet concluded that New Leaders principals are more effective than other principals. This conclusion is not justified. We simply don't know whether New Leaders principals are more effective; only a study that included a wide variety of

additional measures, such as surveys of teachers, supervisors, and students associated with the principals and interviews/surveys of graduates of the PPPs, would provide enough evidence to make such a claim. Even then, the claim would need to be made cautiously as researchers have simply been unable to agree on a strategy to accurately estimate principal effectiveness.²²

In addition, there are at least 6 other problems with this report. These include:

- 1) The study does not convincingly rule out district effects.
- 2) The study relies on multiple VAMs that employ different variables from one site to another.
- 3) The study fails to account for peer effects and the academic profiles of entrants and leavers.
- 4) The study does not examine the potential impact that changes in the pre- and post-achievement tests could have had on the results.
- 5) The study does not explore how the distribution of scores could have affected the results.
- 6) The executive summary presents an overly positive and unbalanced picture of the study's findings.

1) The study does not convincingly rule out district effects.

While the study notes that the effects of New Leaders principals varies across districts, the study does not convincingly rule out that district factors heavily influenced the results. At the lower grade level, the study notes that four districts had positive results, four districts had negative results, and two districts had insignificant results. It is unclear how the study derived these characterizations, particularly with respect to New York, one of the two districts described as having small and insignificant results. In New York, 17 of the possible 24 results (2 school levels x 3 years of principal tenure x 2 subject areas x 2 approaches) were statistically significant and negative. Thus, one could argue 5 of the 10 districts had negative results.

What we don't know is how these districts differed and how these differences might be driving the results found in the study. This possibility is left largely unexplored, thus leaving the reader wondering if the positive results say more about the districts than about New Leaders.

2) The study relies on multiple VAMs that employ different variables and, thus, different ways of measuring effectiveness

The study relies on 10 different VAMs at the lower grade level and 5 different VAMs at the high school level. Not all of the VAMs include the same set of variables. Some of the VAMs

exclude key variables, such as participation in the FRPL program, the percentage of students receiving free lunch, and the percentage of male students. Other VAMs include only one year of principal tenure data—another key variable in the analyses. The VAMs also rely on a wide array of different tests in different subject areas across the sites, and these tests most certainly measure different constructs within a subject area. Further, as mentioned below, the tests—and likely the constructs measured—changed over time. Because of all of these issues, principal effectiveness was measured in different ways across time and across districts. In fact, because different tests were used by different districts, even the construct of “principal effectiveness” as measured by test scores is likely different across districts. The study had to tackle these problems because of the very nature of the placement of New Leaders principals across many districts. The point here is not that the study did a poor job in addressing these issues, but rather that the study did not adequately acknowledge these issues as potentially problematic when pooling results across many principals in different years and in different districts.

3) The study fails to account for peer effects and the academic profiles of entrants and leavers

While a number of studies have found that peer effects influence student outcomes,²³ this study does not include peer effects in the analysis; nor does the study discuss the potential influence of peer effects. This is particularly important given that many of the schools led by New Leaders are charter schools or schools located in districts with school choice options. The academic characteristics of students can vary dramatically among schools, particularly when there are school choice options.

Moreover, the study fails to describe the academic profiles of students entering and leaving schools. This is important for two reasons. First, if schools led by New Leaders principals systemically lost lower-performing students and gained higher-performing students, this could alter the peer effects in such a way as to positively bias the results in favor of the New Leaders principal.²⁴

Second, since the analytic approaches employed in the study relied on a comparison of students remaining in a school with students entering or leaving a school, a comparison of the academic performance of entrants and leavers across schools would uncover any systematic patterns that might bias the results.

4) The study does not examine the potential impact changes in the pre- and post-achievement tests could have had on the results

As mentioned previously, the study relied on a number of different student achievement tests, and these tests changed over time in 5 of the 10 districts. Thus, for many of the analyses, the initial student achievement scores were associated with one test, while the final student achievement scores were assessed with a different test (the Appendix describing the methodology did not provide details on the number of comparisons that

were affected by the change in tests). Partially in response to this situation, the study appropriately converted student scale scores into normal curve equivalent (NCE) scores. If the change in tests resulted in either a change in the constructs measured or in a reallocation of the number of items associated with the constructs assessed by the tests, however, then the change in NCE scores as an indicator of growth could be misleading.²⁵ Although achievement tests would still be highly correlated under such scenarios, the unaccounted-for variance in this relationship, by itself, would simply swamp the small magnitude of the reported effect sizes. Thus, the change in the distribution of scores across tests complicates the comparison of scores—including NCE scores—across tests. Again, the study employs reasonably appropriate techniques to deal with such issues, but fails to discuss or explore them as possible limitations to the study.

5) The study does not explore how the distribution of scores could have impacted the results

The study does not report the average scale scores, NCE scores, or the distribution of either the scale scores or NCE scores by school. Knowing the distribution of the scores for each school—especially the NCE scores, since the analyses relied on the NCE scores—is important because answering one additional question correctly or incorrectly for students at the lower or upper tails of the distribution can translate into much greater changes in NCE scores than answering one additional question correctly or incorrectly for students in the middle of the distribution. If a disproportionate percentage of students start at the lower or upper tails of the distribution for particular schools, then changes over time can be misleading.

6) The executive summary presents an overly positive and unbalanced picture of the study's findings.

Rather than including a complete review of the findings, the report—particularly the executive summary—focuses on the few positive findings of the analyses. For example, with respect to the principal tenure analyses, there were 24 possible results across both school levels and only one positive statistically significant result. That one positive statistically significant result (principal tenure equals 3 years for reading at the high school level under the student/school fixed effects approach) was highlighted without any reference to the two negative, also statistically significant, results or the 21 results that were not statistically significant.

The study is on more solid footing with respect to the findings at the lower grade levels, but it glosses over 4 negative results and 16 not statistically significant results out of the 24 findings reported. A more balanced review would have mentioned all of these findings. If the study focused only on the positive findings because of some research-based reason, it was not clearly presented.

VII. Usefulness of the Report for Guidance of Policy and Practice

This evaluation is both useful and not useful, albeit for different interested parties.

With respect to researchers in this arena, the study provides a rich description of a thoughtful approach to evaluating principal effectiveness and PPP effectiveness. Those attempting to evaluate principals and especially PPPs will find this study quite informative.

More importantly, though, with respect to policymakers, this study is potentially harmful. The positive interpretation of small and sometimes negative results, and the failure to adequately explain the methods and the limitations of those methods, could misleadingly suggest that evaluation approaches such as the one contained in the study could be used in high-stakes accountability systems for principals and PPPs.

Current research is very clear about this—the estimates presented in the study do not accurately capture principal effectiveness and should not be used to make high-stakes decisions about individuals or programs.²⁶ Even if the results did accurately capture principal effectiveness, the very small effect sizes and inconsistencies in the findings related to NL principal effectiveness casts in doubt the over-arching conclusion that New Leaders preparation results in more effective principals. Only when used in combination with a wealth of other qualitative and quantitative data would results from evaluations such as this one potentially be useful to PPP personnel in improving practice. Thus, ultimately, this study misses a very important opportunity to discuss these issues and inform policymakers about the problems and prospects of using the strategies it employs.

Notes and References

- 1 Branch, G., Hanushek, E., & Rivken, S. (2012). *Estimating the Effect of Leaders on Public Sector Productivity: The Case of School Principals*, Working paper 17803. Cambridge, MA: National Bureau of Economic Research. Retrieved January 7, 2013, from <http://www.nber.org/papers/w17803>;
Chiang, H., Lipscomb, S., & Gill, B. (2012). *Is School Value-Added Indicative of Principal Quality?* (Working Paper). Princeton, NJ: Mathematica Policy Research. Retrieved January 7, 2013, from http://mathematica-mpr.com/publications/pdfs/education/value-added_principal_quality.pdf;
Coelli, M., & Green, D. (2012). Leadership Effects: School Principals and Student Outcomes. *Economics of Education Review*, 31(1): 92–109. Retrieved October 19, 2012, from <http://www.sciencedirect.com/science/article/pii/S0272775711001488>;
Dhuey, E., & Smith, J. (2012). *How Important Are School Principals in the Production of Student Achievement?* Working paper. Toronto, ON: University of Toronto. Retrieved January 7, 2013, from http://homes.chass.utoronto.ca/~edhuey/datastore/files/docs/dhuey_smith_princ_aug2013.pdf;
Grissom, J., Kalgoides, D., & Loeb, S. (2014, March). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, online first. Retrieved June 6, 2014, from <http://epa.sagepub.com/content/early/2014/03/10/0162373714523831.full.pdf+html>.
- 2 Fuller, E.J. & Hollingworth, L. (in press). Evaluating principal preparation programs based on placement Rates: Problems and Prospects for Policy Makers. *Journal of Research on Leadership Evaluation*.
- 3 Shelton, S. (2012). *Preparing a Pipeline of Effective Principals*. Denver, CO: National Conference of State Legislatures. Retrieved March 14, 2013, from <http://www.ncsl.org/documents/educ/PreparingaPipelineofEffectivePrincipalsFINAL.pdf.a>
- 4 Cheney, G. R., & Davis, J. (2011). *Gateways to the Principalship: State Power to Improve the Quality of School Leaders*. Washington DC: Center for American Progress. Retrieved October 12, 2012, from <http://americanprogress.org/issues/education/report/2011/10/24/10461/gateways-to-the-principalship/>.
- 5 Fuller, E.J. & Hollingworth, L. (in press). Evaluating principal preparation programs based on placement Rates: Problems and Prospects for Policy Makers. *Journal of Research on Leadership Evaluation*.
- 6 Gates, S. M., Hamilton, L. S., Martorell, P., Burkhauser, S., Heaton, P., Pierson, A., Baird, M., Vuollo, M., Li, J.J., Lavery, D.C. Harvey, M., & Gu, K. (2014). *Preparing Principals to Raise Student Achievement: Implementation and Effects of the New Leaders Program in Ten Districts*. Pittsburgh, PA: RAND Education. Retrieved June 25, 2014, from http://www.rand.org/pubs/research_reports/RR507.html.
- 7 New Leaders (n.d.) About (website page). New York: Author. Retrieved June 25, 2014, from <http://www.newleaders.org/about/>.
- 8 Branch, G., Hanushek, E., & Rivken, S. (2012). *Estimating the Effect of Leaders on Public Sector Productivity: The Case of School Principals*, Working paper 17803. Cambridge, MA: National Bureau of Economic Research. Retrieved January 7, 2013, from <http://www.nber.org/papers/w17803>;
Chiang, H., Lipscomb, S., & Gill, B. (2012). *Is School Value-Added Indicative of Principal Quality?* (Working Paper). Princeton, NJ: Mathematica Policy Research. Retrieved January 7, 2013, from http://mathematica-mpr.com/publications/pdfs/education/value-added_principal_quality.pdf;

- Coelli, M., & Green, D. (2012). Leadership Effects: School Principals and Student Outcomes. *Economics of Education Review*, 31(1): 92–109. Retrieved October 19, 2012, from <http://www.sciencedirect.com/science/article/pii/S0272775711001488>;
- Dhuey, E., & Smith, J. (2012). *How Important Are School Principals in the Production of Student Achievement?* Working paper. Toronto, ON: University of Toronto. Retrieved January 7, 2013, from http://homes.chass.utoronto.ca/~edhuey/datastore/files/docs/dhuey_smith_princ_aug2013.pdf.
- 9 Branch, G., Hanushek, E., & Rivken, S. (2012). *Estimating the Effect of Leaders on Public Sector Productivity: The Case of School Principals*, Working paper 17803. Cambridge, MA: National Bureau of Economic Research. Retrieved January 7, 2013, from <http://www.nber.org/papers/w17803>;
- Chiang, H., Lipscomb, S., & Gill, B. (2012). *Is School Value-Added Indicative of Principal Quality?* (Working Paper). Princeton, NJ: Mathematica Policy Research. Retrieved January 7, 2013, from http://mathematica-mpr.com/publications/pdfs/education/value-added_principal_quality.pdf;
- Fuller, E. J., & Hollingworth, L. (2013). A bridge too far? Challenges in evaluating principal effectiveness. *Educational Administration Quarterly*. Available online only; retrieved June 25, 2014, from <http://eaq.sagepub.com/content/early/2013/10/25/0013161X13506595.abstract>;
- Grissom, J., Kalgorides, D., & Loeb, S. (2014, March). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, online first. Retrieved June 6, 2014, from <http://epa.sagepub.com/content/early/2014/03/10/0162373714523831.full.pdf+html>.
- 10 Branch, G., Hanushek, E., & Rivken, S. (2012). *Estimating the Effect of Leaders on Public Sector Productivity: The Case of School Principals*, Working paper 17803. Cambridge, MA: National Bureau of Economic Research. Retrieved January 7, 2013, from <http://www.nber.org/papers/w17803>;
- Chiang, H., Lipscomb, S., & Gill, B. (2012). *Is School Value-Added Indicative of Principal Quality?* (Working Paper). Princeton, NJ: Mathematica Policy Research. Retrieved January 7, 2013, from http://mathematica-mpr.com/publications/pdfs/education/value-added_principal_quality.pdf;
- Fuller, E. J., & Hollingworth, L. (2013). A bridge too far? Challenges in evaluating principal effectiveness. *Educational Administration Quarterly*. Available online only; retrieved June 25, 2014, from <http://eaq.sagepub.com/content/early/2013/10/25/0013161X13506595.abstract>;
- Grissom, J., Kalgorides, D., & Loeb, S. (2014, March). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, online first. Retrieved June 6, 2014, from <http://epa.sagepub.com/content/early/2014/03/10/0162373714523831.full.pdf+html>.
- 11 Branch, G., Hanushek, E., & Rivken, S. (2012). *Estimating the Effect of Leaders on Public Sector Productivity: The Case of School Principals*, Working paper 17803. Cambridge, MA: National Bureau of Economic Research. Retrieved January 7, 2013, from <http://www.nber.org/papers/w17803>;
- Chiang, H., Lipscomb, S., & Gill, B. (2012). *Is School Value-Added Indicative of Principal Quality?* (Working Paper). Princeton, NJ: Mathematica Policy Research. Retrieved January 7, 2013, from http://mathematica-mpr.com/publications/pdfs/education/value-added_principal_quality.pdf;
- Coelli, M., & Green, D. (2012). Leadership Effects: School Principals and Student Outcomes. *Economics of Education Review*, 31(1): 92–109. Retrieved October 19, 2012, from <http://www.sciencedirect.com/science/article/pii/S0272775711001488>;

- Dhuey, E., & Smith, J. (2012). *How Important Are School Principals in the Production of Student Achievement?* Working paper. Toronto, ON: University of Toronto. Retrieved January 7, 2013, from http://homes.chass.utoronto.ca/~edhuey/datastore/files/docs/dhuey_smith_princ_aug2013.pdf;
- Fuller, E. J., & Hollingworth, L. (2013). A bridge too far? Challenges in evaluating principal effectiveness. *Educational Administration Quarterly*. Available online only; retrieved June 25, 2014, from <http://eaq.sagepub.com/content/early/2013/10/25/0013161X13506595.abstract>;
- Grissom, J., Kalgorides, D., & Loeb, S. (2014, March). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, online first. Retrieved June 6, 2014, from <http://epa.sagepub.com/content/early/2014/03/10/0162373714523831.full.pdf+html>.
- 12 Chiang, H., Lipscomb, S., & Gill, B. (2012). *Is School Value-Added Indicative of Principal Quality?* Working Paper. Princeton, NJ: Mathematica Policy Research Retrieved January 7, 2013, from http://mathematica-mpr.com/publications/pdfs/education/value-added_principal_quality.pdf;
- Fuller, E. J., & Hollingworth, L. (2013). A bridge too far? Challenges in evaluating principal effectiveness. *Educational Administration Quarterly*. Available online only; retrieved June 25, 2014, from <http://eaq.sagepub.com/content/early/2013/10/25/0013161X13506595.abstract>;
- Grissom, J., Kalgorides, D., & Loeb, S. (2014, March). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, online first. Retrieved June 6, 2014, from <http://epa.sagepub.com/content/early/2014/03/10/0162373714523831.full.pdf+html>.
- 13 Chiang, H., Lipscomb, S., & Gill, B. (2012). *Is School Value-Added Indicative of Principal Quality?* (Working Paper). Princeton, NJ: Mathematica Policy Research. Retrieved January 7, 2013, from http://mathematica-mpr.com/publications/pdfs/education/value-added_principal_quality.pdf.
- 14 Chiang, H., Lipscomb, S., & Gill, B. (2012). *Is School Value-Added Indicative of Principal Quality?* (Working Paper). Princeton, NJ: Mathematica Policy Research. Retrieved January 7, 2013, from http://mathematica-mpr.com/publications/pdfs/education/value-added_principal_quality.pdf.
- 15 Chiang, H., Lipscomb, S., & Gill, B. (2012). *Is School Value-Added Indicative of Principal Quality?* (Working Paper). Princeton, NJ: Mathematica Policy Research. Retrieved January 7, 2013, from http://mathematica-mpr.com/publications/pdfs/education/value-added_principal_quality.pdf;
- Grissom, J., Kalgorides, D., & Loeb, S. (2014, March). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, online first. Retrieved June 6, 2014, from <http://epa.sagepub.com/content/early/2014/03/10/0162373714523831.full.pdf+html>.
- 16 Grissom, J., Kalgorides, D., & Loeb, S. (2014, March). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, online first. Retrieved June 6, 2014, from <http://epa.sagepub.com/content/early/2014/03/10/0162373714523831.full.pdf+html>.
- 17 Chiang, H., Lipscomb, S., & Gill, B. (2012). *Is School Value-Added Indicative of Principal Quality?* (Working Paper). Princeton, NJ: Mathematica Policy Research. Retrieved January 7, 2013, from http://mathematica-mpr.com/publications/pdfs/education/value-added_principal_quality.pdf;
- Coelli, M., & Green, D. (2012). Leadership Effects: School Principals and Student Outcomes. *Economics of Education Review*, 31(1): 92–109. Retrieved October 19, 2012, from <http://www.sciencedirect.com/science/article/pii/S0272775711001488>;
- Gates, S. M., Hamilton, L. S., Martorell, P., Burkhauser, S., Heaton, P., Pierson, A., Baird, M., Vuollo, M., Li, J.J., Lavery, D.C. Harvey, M., & Gu, K. (2014). *Preparing Principals to Raise Student Achievement: Implementation and Effects of the New Leaders Program in Ten Districts*. Pittsburgh, PA: RAND Education;

- Grissom, J., Kalgorides, D., & Loeb, S. (2014, March). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, online first. Retrieved June 6, 2014, from <http://epa.sagepub.com/content/early/2014/03/10/0162373714523831.full.pdf+html>.
- 18 Grissom, J., Kalgorides, D., & Loeb, S. (2014, March). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, online first. Retrieved June 6, 2014, from <http://epa.sagepub.com/content/early/2014/03/10/0162373714523831.full.pdf+html>.
- 19 Grissom, J., Kalgorides, D., & Loeb, S. (2014, March). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, online first. Retrieved June 6, 2014, from <http://epa.sagepub.com/content/early/2014/03/10/0162373714523831.full.pdf+html>.
- 20 Grissom, J., Kalgorides, D., & Loeb, S. (2014, March). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, online first. Retrieved June 6, 2014, from <http://epa.sagepub.com/content/early/2014/03/10/0162373714523831.full.pdf+html>.
- 21 Chiang, H., Lipscomb, S., & Gill, B. (2012). *Is School Value-Added Indicative of Principal Quality?* (Working Paper). Princeton, NJ: Mathematica Policy Research. Retrieved January 7, 2013, from http://mathematica-mpr.com/publications/pdfs/education/value-added_principal_quality.pdf.
- 22 Fuller, E. J., & Hollingworth, L. (2013). A bridge too far? Challenges in evaluating principal effectiveness. *Educational Administration Quarterly*. Available online only; retrieved June 25, 2014, from <http://eaq.sagepub.com/content/early/2013/10/25/0013161X13506595.abstract>.
- 23 Burke, M. A., & Sass, T. R. (2013). Classroom peer effects and student achievement. *Journal of Labor Economics*, 31(1), 51-82. Retrieved January 26, 2014, from <http://www.jstor.org/stable/10.1086/666653>;
- Lin, X. (2010). Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics*, 28(4), 825-860. Retrieved Jul 28, 2012, from <http://www.jstor.org/stable/10.1086/653506>.
- 24 Burke, M. A., & Sass, T. R. (2013). Classroom peer effects and student achievement. *Journal of Labor Economics*, 31(1), 51-82. Retrieved January 26, 2014, from <http://www.jstor.org/stable/10.1086/666653>.
- 25 Lockwood, J. R., McCaffrey, D. F., Hamilton, L. S., Stecher, B., Le, V., & Martinez, J. F. (2007). The sensitivity of value-added teacher effect estimates to different mathematics achievement measures. *Journal of Educational Measurement*, 44, 47-67.
- 26 Fuller, E. J., & Hollingworth, L. (2013). A bridge too far? Challenges in evaluating principal effectiveness. *Educational Administration Quarterly*. Available online only; retrieved June 25, 2014, from <http://eaq.sagepub.com/content/early/2013/10/25/0013161X13506595.abstract>;
- Fuller, E. J., & Hollingworth, L. (April, 2014). *Problems and Prospects of Using Outcome Measures to Evaluate Principal Preparation Programs*. Paper presented at the annual meeting of the American Educational Research Association, Philadelphia, PA;
- Grissom, J., Kalgorides, D., & Loeb, S. (2014, March). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, online first. Retrieved June 6, 2014, from <http://epa.sagepub.com/content/early/2014/03/10/0162373714523831.full.pdf+html>.

DOCUMENT REVIEWED:

Preparing Principals to Raise Student Achievement: Implementation and Effects of the New Leaders Program in Ten Districts

AUTHORS:

Susan M. Gates *et al.*

PUBLISHERS/THINK TANKS:

RAND Corporation;
evaluation sponsored by New Leaders

DOCUMENT RELEASE DATE:

May 13, 2014

REVIEW DATE:

June 26, 2014

REVIEWER:

Edward J. Fuller, Penn State University

E-MAIL ADDRESS:

EJF20@PSU.EDU

PHONE NUMBER:

(814) 865-2233

SUGGESTED CITATION:

Fuller, E.J. (2014). *Review of "Preparing Principals to Raise Student Achievement: Implementation and Effects of the New Leaders Program in Ten Districts."* Boulder, CO: National Education Policy Center. Retrieved [date] from <http://nepc.colorado.edu/thinktank/review-preparing-principals>