memorandum

Social and Economic Policy

Date April 2, 2019 (Revised April 25, 2019)

To Elizabeth Ty Wilde and Ed Pauly, The Wallace Foundation

From Cristofer Price and Barbara Goodson

Subject ESSA Evidence Review of the Principal Pipeline Initiative

Abt Associates conducted an independent Every Student Succeeds Act (ESSA)-informed review of the evidence of effectiveness of the Principal Pipeline Initiative (PPI) evaluation. This research was conducted by RAND Education and Labor, a division of the RAND Corporation, in collaboration with Policy Studies Associates. The work was funded through a subcontract from Policy Studies Associates to RAND on a contract between Policy Studies Associates and The Wallace Foundation. The full citation for the report is:

Gates, Susan M., Matthew D. Baird, Benjamin K. Master, and Emilio R. Chavez-Herrerias, *Principal Pipelines: A Feasible, Affordable, and Effective Way for Districts to Improve Schools*, Santa Monica, Calif.: RAND Corporation, 2019. https://www.rand.org/pubs/research_reports/RR2666.html

The review was conducted by Cristofer Price, Principal Scientist at Abt Associates and Barbara Goodson, Principal Scientist at Abt Associates, who are both certified What Works Clearinghouse (WWC) 4.0 reviewers. The review focused on the three primary outcomes in the study: math and reading achievement and principal retention. The study examined additional secondary outcomes, which were not considered in this review.

The review used the ESSA evidence framework which has four levels, or tiers, of effectiveness: Strong (Tier I), Moderate (Tier II), Promising (Tier III), and a fourth category that has been titled Demonstrates a Rationale (Tier IV). The law provides the basic definitions for the tiers, specifying, for example, that Tier I evidence must come from at least one experimental study showing an improved outcome and that Tier II evidence requires a quasi-experiment. ESSA evidence tiers and their requirements for establishing a cause-and-effect relationship are briefly summarized in the box below. For this review, we used the definitions of these tiers that were developed for the Afterschool Programs: A Review of Evidence Under the Every Student Succeeds Act.[1]

[1] Neild, R.C., Wilson, S.J., & McClanahan, W. (2019). *Afterschool programs: A review of evidence under the Every Student Succeeds Act.* Philadelphia: Research for Action. Detailed definitions for this review are continued in a companion document: Neild, R.C., Wilson, S.J., & McClanahan, W. (2019). *Afterschool evidence guide: A companion to Afterschool programs. A review of evidence under the ESSA Act.* Philadelphia: Research for Action.

The results of the review are summarized below, in terms of the statistical significance of the effect and the evidence tier that was met by each of the impact analyses of the primary outcomes.

Overall Effectiveness: Overall, the PPI was found to have a **positive effect.** There were improved outcomes at each of two time points that were statistically significant and there were no overriding negative effects in the following domains:

- ⊕ Mathematics achievement (two and three years after the initial implementation of the PPI)
- ⊕ Reading achievement (two and three years after the initial implementation of the PPI)
- ⊕ Principal retention (two and three years after new principals are placed after the initial implementation of the PPI)

ESSA Evidence Rating: The study of the effectiveness of the PPI was found to have a **Tier II** ESSA evidence rating for the following domains:

- ⊕ Mathematics achievement
- ⊕ Reading achievement

The study of the effectiveness of the PPI was found to have a **Tier III** ESSA evidence rating for the following domain:

- ⊕ Principal retention

The Tier II rating for the math and reading achievement outcomes means that the treatment group population (the pipeline schools) and the comparison group population (the non-pipeline schools in the same state) were shown to be statistically equivalent on math and reading achievement at baseline, before the introduction of the PPI. This increases our confidence that differences in the two groups two and three years later are attributable to the PPI and not to initial differences between the groups. The Tier III rating for principal retention is the result of finding that the treatment group population (the pipeline schools where a new principal has been placed) and the comparison group population (the non-pipeline schools in the same state with new principals placed in the same year) were not shown to be statistically equivalent at baseline. The schools compared on principal retention did not meet criteria for baseline equivalence on two of the five baseline measures (principal tenure and school average math achievement); the schools met criteria for baseline equivalence on school average reading achievement, proportion of students eligible for free or reduced price lunch, and school average principal retention prior to the intervention. Because of the observed baseline differences, we have less confidence that differences between the groups in principal retention rates two and three years after principal placement can be attributed solely to the effect of the new principal. These observed differences could also be the result of initial differences in principal experience or school-level math achievement.

Evidence Tiers in ESSA

Programs with Tier I evidence must be supported by at least one experimental study, the "gold standard" for establishing cause-and-effect relationships. In these studies, students are randomly assigned to experience a program or to the control group. The study must show that the program improved at least one outcome, and the improvement must be statistically significant, or unlikely to be the result of chance variation.

Programs with Tier II evidence must be supported by at least one quasi-experimental study that compares outcomes for treatment program participants to outcomes for a comparison group that is closely matched on important characteristics. As with Tier I evidence, the study must show that the program improved at least one outcome, and the improvement must be statistically significant.

Programs with Tier III evidence must be supported by at least one study that the law describes as "correlational… with statistical controls for selection bias." Although not specified in the law, the implication is that Tier II and Tier III studies have many similarities but program and comparison groups in Tier III studies are not as closely matched. For example, compared to Tier II studies, Tier III studies may have larger differences between the program and comparison groups on previous achievement, which raises more doubt about whether the study represents an "apples-to-apples" comparison.

Programs that meet Tier IV requirements provide a rationale for why outcomes are likely to improve based on existing research described only as "high-quality" in the law and are undergoing evaluation of their effectiveness.

Study Details

STUDY DESCRIPTION

This quasi-experiment estimates the impact of a district-wide principal initiative in six large urban school districts. The treatment group for all outcomes is defined as students in grades 3 – 10 in schools in which new principals were placed as part of the initiative. The comparison schools, selected from non-PPI districts in the same states as the study districts, experienced "business as usual," which could have included other district-sponsored principal support activities. The study was conducted in from 2011 to 2016.

MATHEMATICS ACHIEVEMENT

Outcome	ESSA Evidence Tier		Group	Number of Schools	Effect Size (standard error)
	With site and sample criteria	*Without* site and sample criteria			
State mathematics tests: 2 years after principal placement	Tier II	Tier II	Elementary, middle, high schools	966 Treatment/ 4927 Comparison	0.075* (.016)
State mathematics tests: 3 years after principal placement	Tier II	Tier II	Elementary, middle, high schools	819 Treatment/ 4407 Comparison	0.086* (.020)

* Statistically significant at p<.05

READING ACHIEVEMENT

Outcome	ESSA Evidence Tier		Group	Number of Schools	Effect Size (standard error)
	With site and sample criteria	*Without* site and sample criteria			
State reading tests: 2 years after principal placement	Tier II	Tier II	Elementary, middle, high schools	835 Treatment/ 4559 Comparison	0.14* (.016)
State reading tests: 3 years after principal placement	Tier II	Tier II	Elementary, middle, high schools	982 Treatment/ 5179 Comparison	0.17* (.020)

* Statistically significant at p<.05

PRINCIPAL RETENTION

Outcome	ESSA Evidence Tier		Group	Number of Schools	Difference in Percent Retained (standard error)
	With site and sample criteria	*Without* site and sample criteria			
Principal retention: 2 years after principal placement	Tier III	Tier II	Elementary, middle, high schools	1051 Treatment/ 6278 Comparison	5.8%* (2.0%)
Principal retention: 3 years after principal placement	Tier III	Tier III	Elementary, middle, high schools	832 Treatment/ 4910 Comparison	7.8%* (3.1%)

* Statistically significant at p<.05

Research Design Requirements

Outcomes

Domain / Oucome	Measure	Reliability	Validity	Over-Aligned	Same Method in T and C	Meets WWC Outcome Standards
Math achievement: 2 years after principal placement	z-scored state achievement tests within state grade and year	Assumed reliable	Face valid	No	Yes	Meets
Math achievement: 3 years after principal placement	z-scored state achievement tests within state grade and year	Assumed reliable	Face valid	No	Yes	Meets
Reading achievement: 2 years after principal placement	z-scored state achievement tests within state grade and year	Assumed reliable	Face valid	No	Yes	Meets
Reading achievement: 3 years after principal placement	z-scored state achievement tests within state grade and year	Assumed reliable	Face valid	No	Yes	Meets
Principal retention: 2 years after principal placement	Binary	Assumed reliable	Face valid	No	Yes	Meets
Principal retention: 3 years after principal placement	Binary	Assumed reliable	Face valid	No	Yes	Meets

Baseline equivalence

Domain/Outcome	Baseline Measure	Meets WWC Standards for Baseline Measures	Effect Size of Baseline Difference	Meets WWC BE Standards
Math achievement: 2 years after principal placement	Math in year before PPI intervention. Z-scored state achievement tests within state, grade, year	Yes	.022	Meets
Math achievement: 3 years after principal placement	Math in year before PPI intervention. Z-scored state achievement tests within state, grade, year	Yes	.021	Meets
Reading achievement: 2 years after principal placement	Reading in year before PPI intervention. Z-scored state achievement tests within state, grade, year	Yes	.005	Meets

Domain/Outcome	Baseline Measure	Meets WWC Standards for Baseline Measures	Effect Size of Baseline Difference	Meets WWC BE Standards
Reading achievement: 3 years after principal placement	Reading in year before PPI intervention. Z-scored state achievement tests within state, grade, year	Yes	.010	Meets
Principal retention 2 years after principal placement	The study did not have a measure of years of principal of experience	NA	NA	Does not meet
Principal retention 2 years after principal placement	School average retention prior in years before PPI intervention. Z-scored mean of retention rates within sample	Yes	.025	Meets
Principal retention 2 years after principal placement	School-level percentage of students eligible for free or reduced price lunch in year before PPI intervention	Yes	.12 (Cox index)	Meets
Principal retention 2 years after principal placement	Reading in year before PPI intervention. Z-scored state achievement tests within state, grade, year	Yes	.04	Meets
Principal retention 2 years after principal placement	Math in year before PPI intervention. Z-scored state achievement tests within state, grade, year	Yes	.12	Does not meet (Baseline effect size >.05 and <.25, but this measure not included as a model covariate)
Principal retention 3 years after principal placement	The study did not have a measure of years of principal experience	NA	NA	Does not meet
Principal retention 3 years after principal placement	School average retention prior in years before PPI intervention. Z-scored mean of retention rates within sample	Yes	.025	Meets
Principal retention 3 years after principal placement	School-level percentage of students eligible for free or reduced price	Yes	.12 (Cox index)	Meets

Domain/Outcome	Baseline Measure	Meets WWC Standards for Baseline Measures	Effect Size of Baseline Difference	Meets WWC BE Standards
	lunch in year before PPI intervention			
Principal retention 3 years after principal placement	Reading in year before PPI intervention. Z-scored state achievement tests within state, grade, year	Yes	.04	Meets
Principal retention 3 years after principal placement	Math in year before PPI intervention. Z-scored state achievement tests within state, grade, year	Yes	.13	Does not meet (Baseline effect size >.05 and <.25, but this measure not included as a model covariate)

Representativeness

Domain	Measure	Representativeness of Baseline Sample	Representativeness of Outcome Sample	Meets WWC Representativeness Standard
Math achievement	State assessments 2 and 3 years	yes	yes	Yes
Reading achievement	State assessments 2 and 3 years	yes	yes	Yes

Correlation of Pretest and Posttest (required for difference-in-difference and CIT-S models)

Domain	Measure	Pre-Post Correlation 2 Years After Principal Placement	Pre-Post Correlation 3 Years After Principal Placement	Meets WWC Standard for Correlation (> .60)
Math achievement:	z-scored state assessments	.77	.74	Yes
Reading achievement:	z-scored state assessments	.79	.75	yes

Data imputation

No imputation of missing baseline or posttest data. Dummy indicator used for missing covariate values.

https://doi.org/10.59656/EL-LS8403.001