



NEPC REVIEW: EVALUATION OF THE TEACHER INCENTIVE FUND: FINAL REPORT ON IMPLEMENTATION AND IMPACT OF PAY-FOR-PERFORMANCE ACROSS FOUR YEARS (INSTITUTE OF EDUCATION SCIENCES, DECEMBER 2017)



Reviewer:

Francesca López
University of Arizona

February 2018

National Education Policy Center

School of Education, University of Colorado Boulder
Boulder, CO 80309-0249
(802) 383-0058
nepc.colorado.edu

Acknowledgements

NEPC Staff

Kevin Welner
Project Director

William Mathis
Project Manager

Patricia Hinchey
Academic Editor

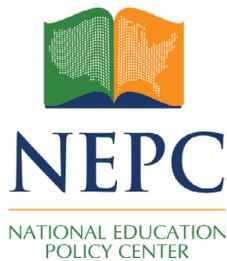
Alex Molnar
Publications Director

Suggested Citation: López, F.A. (2018). *NEPC Review: “Evaluation of the Teacher Incentive Fund: Final Report on Implementation and Impact of Pay-for-Performance across Four Years”* (Institute of Education Sciences, December 2017). Boulder, CO: National Education Policy Center. Retrieved [date] from <http://nepc.colorado.edu/thinktank/review-teacher-incentive>.

Funding: This review was made possible in part by funding from the Great Lakes Center for Educational Research and Practice.



This publication is provided free of cost to NEPC’s readers, who may make non-commercial use of it 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.



NEPC REVIEW: EVALUATION OF THE TEACHER INCENTIVE FUND: FINAL REPORT ON IMPLEMENTATION AND IMPACT OF PAY-FOR-PERFORMANCE ACROSS FOUR YEARS (INSTITUTE OF EDUCATION SCIENCES, DECEMBER 2017)

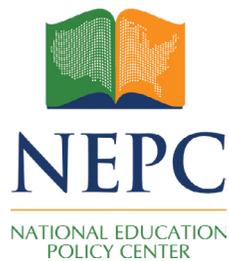
Reviewer:

Francesca López
University of Arizona

February 2018

Summary of Review

This review examines a report evaluating four years of implementation of the 2010 Teacher Incentive Fund (TIF) to determine whether pay-for-performance (PFP) bonuses to teachers and principals improve student achievement. The report notes that the goal of PFP is to motivate improved educator performance and to attract and retain more effective teachers, thereby increasing student achievement. The report found that PFP marginally increased student test scores in reading and math by year two and that it was associated with slightly higher classroom observation ratings; it also found that classroom observation ratings were not associated with student test scores and that PFP had no effect on filling teacher vacancies. Problematic issues in the report involve combining non-comparable sites, employing different outcome measures, and generating a misleading cost-effectiveness analysis. Although several thorough analyses are presented, the data do not support the report's suggestion that PFP holds promise to address achievement disparities. Overall, because of the report's incorporation of a polyglot of outcome measures of questionable validity, this review finds that the report lacks utility for policymakers.



NEPC REVIEW: EVALUATION OF THE TEACHER INCENTIVE FUND: FINAL REPORT ON IMPLEMENTATION AND IMPACT OF PAY-FOR-PERFORMANCE ACROSS FOUR YEARS (INSTITUTE OF EDUCATION SCIENCES, DECEMBER 2017)

Reviewer:

Francesca López
University of Arizona

February 2018

I. Introduction

Congress established the Teacher Incentive Fund (TIF) in 2006 to provide funding for pay-for-performance (PFP) systems in high-need schools. Sixty-two TIF grants were awarded to states or school districts, nonprofits, charter schools or charter management organizations, and universities, with implementation taking place over four years (2011 to 2015) in a total of 183 schools. The participating schools were randomly assigned to treatment or control groups. Cohort 1 included 10 school districts and a total of 138 schools, with random assignment taking place spring/summer of 2011. Cohort 2 included three school districts and a total of 45 schools, with random assignment taking place spring/summer of 2012.

The U.S. Department of Education requested an evaluation of the 2010 TIF grants with a focus on two goals. The first was to “inform program development and improvement by describing how grantees implemented the PFP and the challenges they faced.”¹ The second goal was to examine whether PFP was associated with student achievement outcomes, as well as teacher effectiveness defined by student growth and classroom observation scores.

This review assesses the claims and findings of the Department of Education’s Institute of Education Sciences (IES) report evaluating four years’ experience of Cohort 1 2010 TIF grantees, including the extent to which PFP was associated with student and teacher outcomes. Although the report states that “IES evaluation reports do not include conclusions or recommendations or views with regard to actions policymakers or practitioners should take in light of the findings of the report,” numerous implicit recommendations suggest PFP holds promise to address achievement disparities in high-need schools. To the contrary, the evidence is negligible.²

II. Findings and Conclusions of the Report

The study addresses four research questions in a 120-page body and includes an equal number of pages for appendices. The four questions probe the characteristics of the TIF districts, comparisons of teachers and principals, PFP's effects on educator effectiveness and retention, and its effects on student achievement. The latter two areas are those with greatest implications for policy.

One of the report's key findings is that PFP "led to slightly higher student achievement in reading and math by the second year of implementation."³ Effects ranged between .02 and .06 standard deviations, conventionally considered low.

The report employs a change theory that asserts PFP should influence educator behaviors, which in turn should influence student achievement. The report finds, however, that there is "no consistent evidence of any relationship between impact on educator behaviors and impacts on student achievement."⁴

The report also finds that PFP "led teachers to earn slightly higher classroom observation ratings by year 3."⁵ However, it also notes that the "impacts remained small, never amounting to more than 10 percent of the increment between the two performance levels."⁶

The report further asserts that PFP "raised the performance ratings of returning teachers in years 3 and 4."⁷ Despite the absence of effects in years 1 and 2 and the miniscule effect in year 3 (.06 points on a 1-4 scale), the report asserts, "Given the consistent evidence that pay-for-performance led to higher performance ratings among returning teachers, it is likely that pay-for-performance either helped schools retain more effective teachers or motivated returning teachers to be more effective."⁸ The report claims that PFP "had little impact on principals' success in filling teacher vacancies."⁹

The report then claims that "pay-for-performance was generally more cost-effective than class-size reduction."¹⁰

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

The report asserts that "because the study used random assignment, any differences in educators' or students' outcomes between the treatment and control group can be attributed to pay-for-performance and not some other characteristic of the districts or school."¹¹ To be eligible for the TIF grant, applicants had to demonstrate core elements that included

The involvement and support of teachers, principals, and other personnel (including input from teachers, principals, and other personnel in the schools and LEAs to be served by the grant) and the involvement and support of unions in participating LEAs (where they are the designated exclusive representatives for the purpose of collective bargaining) that is needed to carry out the grant.¹²

As will be addressed below, the rationale assumes that random assignment of schools and pairing of virtual twins erased other factors that could have influenced the outcomes. A total of approximately 3,000 to 4,000 teachers¹³ across both treatment and control groups were included in the evaluation. Data for analyses were derived from eight sources: Common Core of Data, a district survey, district interviews, principal surveys, teacher surveys, technical assistance documents, student administrative data, and educator administrator data. Response rates for evaluation districts ranged between 89-100%.¹⁴

IV. The Report's Use of Research Literature

The report cites 24 PFP-focused studies in one footnote but limits explication of them to a single page concluding “Although some studies have found positive impacts of these programs on student achievement, most have not.”¹⁵ The report then cites additional research literature that points to limitations in prior studies that informed the framework of the 2010 TIF evaluation. These limitations included the absence of a randomized assignment design, issues with the amount of bonuses or dissemination of information about the bonuses, and the lack of “a more comprehensive reform package that would help teachers change their teaching practices.”¹⁶

The report explains that although the impacts of PFP were small, “policy decisions on whether to direct resources to pay-for-performance may depend, in part, on whether it is a cost-effective way of raising student achievement compared with other policies.”¹⁷ The report compares the cost-effectiveness of PFP to transfers and class-size reduction. While it does not provide a comprehensive review of studies on transfers or class-size reduction, it does cite four studies used in calculating the cost-per-student for class-size reduction. The report later explains that “the available evidence cannot predict the policies’ cost-effectiveness beyond the limited number of years in which these policies were implemented and evaluated.”¹⁸ However, the absence of key class-size reduction studies is a glaring omission.¹⁹

The report also cites a study²⁰ that provided a formula for cost-effectiveness ratios; however, that formula was not used in the report’s evaluation of cost-effectiveness. Instead, the report’s estimate is based on the class-size reduction costs in the cited study.

V. Review of the Report's Methods

Several issues with the report’s methods suggest that implications of the findings are overstated. These include assertions of causality that cannot be justified; a “Frankenstein” amalgamation of key outcome measures; reporting of “impacts” that are at times statistically “significant” but are nevertheless too miniscule to inform policy; and a misleading cost-effectiveness analysis.

Causality

Given the report's acknowledgement regarding the limitations of prior research (which has generally found PFP does not have positive "impacts" on student achievement), it is particularly problematic for the report to claim that "because the study used random assignment, any differences in educators' or students' outcomes between the treatment and control group can be attributed to pay-for-performance and not some other characteristic of the districts or school."²¹ However, the claim of causality is undermined by issues with attrition and with differences in baseline measures of treatment and control groups.²² That is, baseline measures for achievement and ethnicity across treatment and control groups differed. In addition, the study included no controls for the size of the bonuses, which differed from district to district. Using baseline controls to adjust for differences between treatment and control groups was necessary, but it also requires that the findings be interpreted cautiously, with reservations. Moreover, as the report explains,

Because some students entered or left the study schools during the course of the study, implementing pay-for-performance at a school for a certain number of years did not mean that all students at the school were exposed to pay-for-performance for that duration. In fact, at the end of Year 4, our estimate is that 37 percent of students had been at their schools since the beginning of the study and were exposed to all four years of the program.²³

Identified issues with baseline differences, treatment, and attrition all attenuate the report's causal inferences.

Frankenstein Measures

The measures used to reflect "teacher effectiveness" and student achievement were selected at the discretion of participating school districts. This resulted in substantial variation in what the scores reflected for each school and/or district, compromising the validity of the various outcomes. Scores from required classroom observations of classroom or school practices included those from the Teacher Advancement Program teacher observation rubric; Danielson's Framework for Teaching (or a modified version); Kim Marshall's observation rubric; and district or state observation rubrics. The number of observations was also left to the discretion of participating districts, although there was a required minimum of two.

Similarly, various measures were used to assess student growth as a measure of teacher effectiveness. Estimates of student achievement growth were variously calculated based on the teachers' own students; all students in the same grade, team, or subject area; all students in the school; or a combination of these. Moreover, districts could use value-added models or calculate changes in achievement from one year to the next. The student achievement scores themselves were also highly varied, with some districts using state assessments and others using assessments developed by private vendors. The various measures were converted to z scores and treated as if they could therefore be combined or compared. This approach means that the report completely ignores the differences among the numerous measures used to derive scores for "teacher effectiveness" and "student achievement."

Misleading “Impacts”

One of the key findings in the report is that PFP “led to slightly higher student achievement in reading and math by the second year of implementation.”²⁴ This is not accurate. The achievement in math was statistically higher only in year 3. To explain the very small effects that ranged between .02 and .06 standard deviations for both reading and math, the report describes differences in percentile units that ranged between 1 to 2 percentiles.

In the first two years of the TIF implementations, the average student in control schools earned a math score at about the 33rd percentile of student achievement statewide. After treatment schools had implemented pay-for-performance for two years, their students earned an average math score at about the 35th percentile—a difference of two percentile points.²⁵

The report also claims that the effects in terms of “weeks of learning” translate to the equivalent of three to four weeks of learning in both reading and math. Both of these approaches—converting effect sizes to percentile units and weeks of learning—are egregiously problematic. First, percentiles are not distributed equally along a continuum: they are very close together in the middle, spreading out further apart as they get closer to the tails of the distribution. In the example provided by the report, the 2 percentile difference reported actually diminishes below the 33rd and above the 67th percentiles. In other words, not all students gained two percentiles. Furthermore, a gain of two percentiles coming from the 33rd percentiles can be explained by regression toward the mean.

The key question here, however, is whether an effect of .02 to .06 standard deviations is noteworthy. If we consider that TIF targeted high-need schools, suggesting that the program’s ultimate goal was to eliminate achievement disparities, the answer is “no.” To put the apparent gain in perspective: the achievement disparity on the 2015 4th grade reading National Assessment of Educational Progress²⁶ (NAEP) between White (232) and Black (206) students was .76 of a standard deviation; for Hispanic (208) students, it was .68 of a standard deviation. PFP, at best, eliminated only 8-9% of the disparities in reading (7-9% in math). The gain is far too weak to make up the difference.

The strategy of attempting to convert effects to “weeks of learning” is similarly misleading. If achievement disparities were converted to weeks, Black and Hispanic students would need approximately 1 ½ years of learning to eliminate achievement disparities. The miniscule relationship between PFP and student achievement (with the disclaimer that PFP cannot be causally attributed to achievement), whether reported in percentiles or weeks of learning, is hardly worth touting.

One of the explicit goals of PFP is that it would help principals attract and retain more effective teachers. The report asserts that PFP had little impact on principals’ success in filling teacher vacancies, but there were actually *no* differences between control and treatment groups on the teaching offers accepted. Therefore, the report should have asserted that PFP had *no* impact on principals’ success in filling teacher vacancies. In terms of retention, the report maintains that PFP “raised the performance ratings of returning teachers in years 3

and 4.”²⁷ Despite the absence of effects in years 1 and 2 and the miniscule effect in year 3 (.06 points on a 1-4 scale), the report says,

Given the consistent evidence that pay-for-performance led to higher performance ratings among returning teachers, it is likely that pay-for-performance either helped school retain more effective teachers or motivated returning teachers to be more effective.²⁸

The findings, however, do not support this claim. The results were not consistent and the assertions are speculative, at best.

The report also claims that PFP “led teachers to earn slightly higher classroom observation ratings by year 3.”²⁹ It displays “impacts” in raw score units (on a 1-4 scale) that ranged from .03 to .09. Although the report fails to provide effect sizes to examine the magnitude of the “impacts,” it explicitly explains that “these impacts remained small, never amounting to more than 10 percent of the increment between two performance levels.”³⁰ However, the point of the classroom observation ratings was to capture a component of teacher effectiveness. Indeed, the report asserts that the premise of PFP is that it should motivate teachers to be more effective. To that end, teachers must be aware of the possibility of a bonus based on performance. Fewer than half of the teachers in treatment schools, however, reported that they were aware of their eligibility for a performance bonus. The second year had only a slight improvement (from 49% in year 1 to 61% in year 2), with no subsequent improvement in years 3 and 4. It makes no sense for PFP be credited for teachers’ effectiveness when so many teachers were unaware of the program. Moreover, the report states that “no consistent evidence of any relationship between impact on educator behaviors and impacts on student achievement” were found.³¹ These findings, though accurately presented in the report, were not used to attenuate the key finding that PFP led to higher student achievement, despite the fact that the theory of change used by the report explicitly linked PFP to educator behaviors and student achievement.

The report also asserts that compared to principals in control schools, principals in treatment schools were more satisfied with their jobs by the third year of implementation, as measured by the TIF principal survey. However, the findings were not statistically significant and therefore principals cannot be accurately described as “more” or “less” satisfied.

Cost-Effectiveness Analysis

The report asserts that “The goal of a cost-effectiveness analysis is to enable policymakers to compare, and ultimately choose from, alternative policies.”³² The report then claims that “pay-for-performance was generally more cost-effective than class-size reduction.”³³ Thus, the report implicitly endorses PFP as a more cost-effective choice when compared to class-size reduction. The report, however, omits salient research on class-size reduction that compromises this conclusion.

The report compares PFP, transfers, and class-size reduction using an impact of a .04 standard deviation increase in achievement per year as a meaningful growth target. To approx-

imate the impacts and costs for PFP, the report uses the Common Core of Data, as well as educator and student administrator data. To approximate the impacts and costs of transfers, the report uses one published study.³⁴ To approximate the impacts of class-size reduction, the report uses two studies,³⁵ and to approximate costs of class-size reduction, it uses one study.³⁶ From this narrow and questionable base, the report concludes that PFP would cost \$153 per student at the end of one year, whereas class-size reduction would cost \$345 per student. Although the report cites a study that provides formulas for cost-effectiveness ratios,³⁷ the report fails to calculate (or disclose) such ratios for performance pay or class-size reduction.

In terms of student achievement, it should be noted that PFP was unable to produce effects larger than .06 over the course of a year; in contrast, the single-year effects in the cited class-size study³⁸ ranged from .20 to .39. In addition, a study on the long-term effects of class-size reduction³⁹ found that by grade 4, the cumulative effects of small classes in all grades was a .34 standard deviation increase in math (.37 standard deviations in reading). Given that achievement disparities in high-need schools are much larger (up to almost .90 standard deviations on 2015 4th grade NAEP⁴⁰), these calculations indicate that it would take 15 years for PFP to eliminate achievement disparities (assuming growth would be cumulative). For class-size reduction, however, it would take only about 2.5 years. The report also ignored the long-term study's finding⁴¹ that reduced class-size effects did not fade over time; effects in grade 8 remained as large as they were in grade 3 (approximately .35 in both reading and math).

VI. Review of the Validity of the Findings and Conclusions

The report is consistent with the U.S. Department of Education's primary goal for the evaluation, which was to inform program development of PFP (rather than evaluate whether it holds utility for policymakers). Although the secondary goal was to examine whether PFP was associated with teacher effectiveness and student achievement outcomes, the findings do not support the assertions that PFP shows promise in promoting achievement in high-need schools and is more cost-effective than reducing class size. The report includes several analyses that cast doubt on PFP's utility, and it does not appropriately temper its findings in light of the fact that reported positive effects are miniscule.

- The findings are based on the assumption of causality, which cannot be sustained given the substantive attrition rate and the weaknesses in baseline controls for schools as well as students.
- The analyses use outcomes that are an amalgamation of numerous measures, resulting in compromised findings.
- Effect sizes are rarely reported. The report relies on descriptions of "impacts" that are so small that they have little or no value for policy making.

- In many cases, impacts are reported even when they are not statistically significant and when they are exaggerated by flawed metrics.
- The assertion that PFP is more cost-effective is egregiously misleading. Class-size reduction has markedly larger, lasting effects that have meaningful implications for policy.

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

The report states that the central goal of PFP “is to increase student achievement by motivating educators to improve their performance and by attracting and retaining more effective teachers.”⁴² The report found that there is “no consistent evidence of any relationship between impact on educator behaviors and impacts on student achievement,” which reflects a clear disconnect between the premise of the study and the findings. In consideration of the flawed methods that include unsupported causality assumptions, a polyglot of outcome measures of questionable validity, exaggerated presentation of the findings and misleading cost-effectiveness analysis, this review finds that the report lacks utility for policymakers.

Notes and Resources

- 1 Institute of Education Sciences (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. 1.
- 2 Institute of Education Sciences. (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. ii.
- 3 Institute of Education Sciences. (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. xxi.
- 4 Institute of Education Sciences. (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. 106.
- 5 Institute of Education Sciences. (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. 93.
- 6 Institute of Education Sciences. (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. 93.
- 7 Institute of Education Sciences. (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. 96.
- 8 Institute of Education Sciences. (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. 96.
- 9 Institute of Education Sciences. (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. 87.
- 10 Institute of Education Sciences. (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. 108.
- 11 Institute of Education Sciences. (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. 23.
- 12 U.S. Department of Education. *Application for the Teacher Incentive Fund* (CFDA Number 84.385). Retrieved February 1, 2018, from <https://www2.ed.gov/programs/teacherincentive/appfinal5192010.doc>, p. 12.
- 13 The report provides ranges given fluctuations in the data available for demographic characteristics. See Institute of Education Sciences (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. 19.
- 14 Institute of Education Sciences. (2017). *Evaluation of the Teacher Incentive Fund: Final report on*

- implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. 20.
- 15 Institute of Education Sciences. (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. 2.
 - 16 Institute of Education Sciences. (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. 2.
 - 17 Institute of Education Sciences. (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. 26.
 - 18 Institute of Education Sciences. (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. 110.
 - 19 Nye, B., Hedges, L.V., & Konstantopoulos, S. (1999). The long-term effects of small classes: A five-year follow-up of the Tennessee class size experiment. *Educational Evaluation and Policy Analysis*, 21(2), 127-142.
 - 20 Harris, D.N. (2009). Toward policy-relevant benchmarks for interpreting effect sizes: Combining effects with costs. *Educational Evaluation and Policy Analysis*, 31(1), 3-29.
 - 21 Institute of Education Sciences (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. 23.
 - 22 What Works Clearinghouse Evidence Standards for Reviewing Studies, Version 1.0 (2008, May). Retrieved February 1, 2018, from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_version1_standards.pdf
 - 23 Institute of Education Sciences (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. 25.
 - 24 Institute of Education Sciences (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. xxi.
 - 25 Institute of Education Sciences (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. 99.
 - 26 The Nation's Report Card. Retrieved February 1, 2018, from <https://nces.ed.gov/nationsreportcard/naepdata/dataset.aspx>
 - 27 Institute of Education Sciences (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. 96.
 - 28 Institute of Education Sciences (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. 96.
 - 29 Institute of Education Sciences (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC:

U.S. Department of Education, p. 93.

- 30 Institute of Education Sciences (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. 93.
- 31 Institute of Education Sciences (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. 106.
- 32 Institute of Education Sciences (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. H.14.
- 33 Institute of Education Sciences (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. 108.
- 34 Glazerman, S., Protik, A., Teh, B., Bruch, J., & Max, J. (2013). *Transfer Incentives for High-Performing Teachers: Final Results from a Multisite Randomized Experiment*. NCEE 2013-4003. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.
- 35 Nye, B., Hedges, L.V., & Konstantopoulos, S. (2000). The effects of small classes on academic achievement: The results of the Tennessee class size experiment. *American Educational Research Journal*, 37(1), 123-151.
- 36 Harris, D.N. (2009). Toward policy-relevant benchmarks for interpreting effect sizes: Combining effects with costs. *Educational Evaluation and Policy Analysis*, 31(1), 3-29.
- 37 Harris, D.N. (2009). Toward policy-relevant benchmarks for interpreting effect sizes: Combining effects with costs. *Educational Evaluation and Policy Analysis*, 31(1), 7.
- 38 Nye, B., Hedges, L.V., & Konstantopoulos, S. (2000). The effects of small classes on academic achievement: The results of the Tennessee class size experiment. *American Educational Research Journal*, 37(1), 123-151.
- 39 Nye, B., Hedges, L.V., & Konstantopoulos, S. (1999). The long-term effects of small classes: A five-year follow-up of the Tennessee class size experiment. *Educational Evaluation and Policy Analysis*, 21(2), 127-142.
- 40 The Nation's Report Card. Retrieved February 1, 2018, from <https://nces.ed.gov/nationsreportcard/naepdata/dataset.aspx>
- 41 Nye, B., Hedges, L.V., & Konstantopoulos, S. (1999). The long-term effects of small classes: A five-year follow-up of the Tennessee class size experiment. *Educational Evaluation and Policy Analysis*, 21(2), 127-142.
- 42 Institute of Education Sciences (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impact of pay-for-performance across four years* (NCEE 2018-4004). Washington, DC: U.S. Department of Education, p. 87.