NEPC Review: The Effectiveness of Secondary Math Teachers from Teach for America and the Teaching Fellows Programs (Institute of Education Sciences, September 2013)

Reviewed by:

Andrew Brantlinger and Matthew Griffin
University of Maryland

May 2019

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
Managing Director

Alex Molnar
Publications Director

Suggested Citation: Brantlinger, A. & Griffin, M. (2019). *Review of "The Effectiveness of Secondary Math Teachers from Teach For America and the Teaching Fellows Programs"* (Institute of Education Sciences, September 2013) Boulder, CO: National Education Policy Center. Retrieved [date] from http://nepc.colorado.edu/thinktank/tfa.

The authors thank Kathryn Boonstra, Daniel Chazan, Robert Croninger, David Imig, William Mathis, and Linda Valli for providing critical feedback on this manuscript.

This research was supported by NSF Grant #1535251: Examining the Career Trajectories of Urban Math Teachers from a Selective Alternative Certification Program.

NEPC
NATIONAL EDUCATION POLICY CENTER

NEPC Review: The Effectiveness of Secondary Math Teachers from Teach for America and the Teaching Fellows Programs (Institute of Education Sciences, September 2013)

Reviewed by:

Andrew Brantlinger and Matthew Griffin
University of Maryland

May 2019

Executive Summary

This NEPC review offers a critique of a teacher effectiveness experiment conducted by investigators from the Mathematica Policy Group and published by the Institute of Education Sciences.[1] The Mathematica experiment was designed to provide evidence about the effectiveness of teachers who were themselves high-achieving students and trained by either Teach for America (TFA) or the Teaching Fellows programs. Although it did not show that high achievers are more effective at teaching mathematics than typical mathematics teachers in high-poverty schools, it did show that TFA corps members (CMs)[2] significantly outperformed the secondary mathematics teachers they work with.[3] Based on that, the investigators recommended that principals of high-poverty schools would do well to hire TFA's mathematics teachers. Our critique is based on a secondary analysis of the study data. We show that the oversampling of underqualified comparison teachers, the undersampling of first-year TFA corps members, and experimental contamination (i.e., TFA teachers teaching to the study tests) biased the experiment in favor of TFA. We conclude by arguing for more rigorous and reliable results about teacher preparation programs, inclusive of TFA, before making hiring recommendations and for greater dialogue about the results of these kinds of studies.

NEPC
NATIONAL EDUCATION POLICY CENTER

NEPC Review: The Effectiveness of Secondary Math Teachers from Teach for America and the Teaching Fellows Programs (Institute of Education Sciences, September 2013)

Reviewed by:

Andrew Brantlinger and Matthew Griffin
University of Maryland

May 2019

I. Introduction

Prominent reformers believe that, through recruiting more high achievers to teaching, the U.S. can improve teacher quality and, hence, student achievement.[4] Some observe that, in the U.S., "[l]ittle attention has been paid to altering the value proposition of teaching to draw young people with strong academic backgrounds to the career" while acknowledging that "[p]aradoxically, U.S. research on whether teachers' academic backgrounds significantly predict classroom effectiveness is very mixed."[5] Needed is a rigorous study that will show whether or not high achievers make better teachers than academic low achievers.

A report titled, *The Effectiveness of Secondary Math Teachers from Teach For America and the Teaching Fellows Programs,* published in 2013 by the Institute of Educational Sciences in the U.S. Department of Education, was designed to address this gap. The authors included six researchers from Mathematica Policy Group, including the lead author Melissa Clark, and one from Chesapeake Research Associates.

The Mathematica report presented the results of a two-year experiment that was designed to provide evidence about the effectiveness of high-achieving teachers. The investigators sought participants from Teach for America (TFA) and Teaching Fellows programs precisely because these programs use "highly selective admissions criteria designed to admit only applicants who have demonstrated a high level of achievement in academics or other endeavors."[6] Both programs seek to close achievement gaps through providing "excellent teachers" who use "effective teaching in every classroom."[7] TFA and Teaching Fellows programs both figure prominently in the policy imaginative as a mechanism to attract the best and the brightest to teaching. Over the past two decades, more than 100,000 teachers, including

more than 10,000 secondary mathematics teachers, have entered teaching through TFA and Teaching Fellows programs.[8]

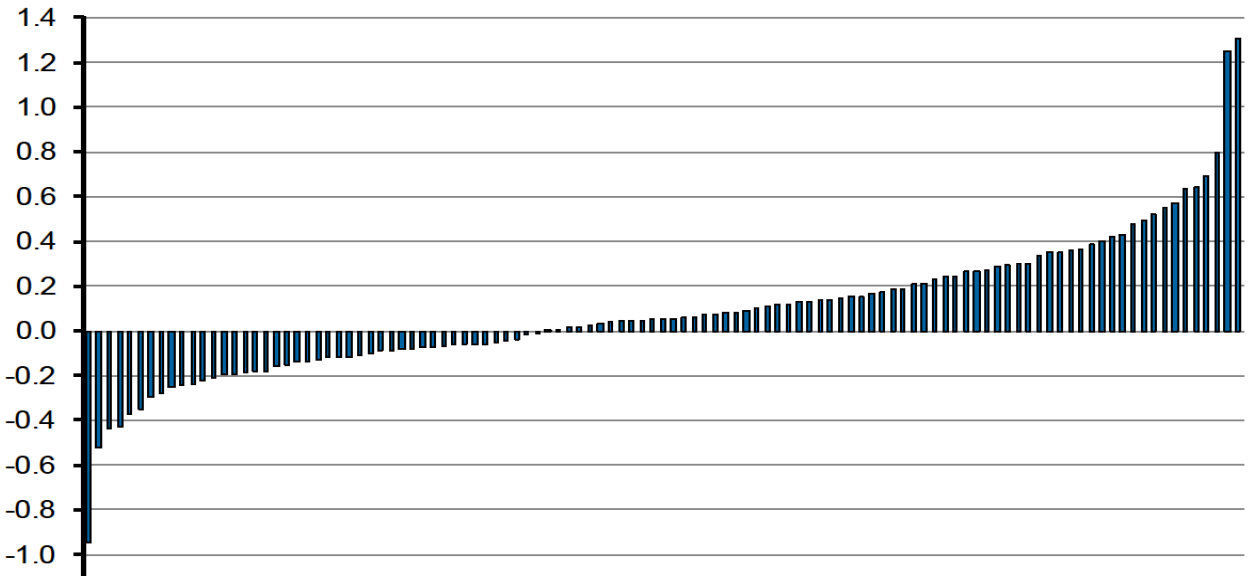
II. Findings and Conclusions of the Report

The Mathematica experiment showed that teachers from selective colleges were no more effective at teaching mathematics than other teachers in high-poverty secondary schools; that is, there was, "no significant difference in effectiveness between teachers from selective colleges or universities and those from all other educational institutions" and that, "in sensitivity analyses, … teachers from *highly* selective colleges or universities did not differ in effectiveness from teachers whose colleges or universities had lower levels of selectivity."[9] The experiment also showed that, "Teaching Fellows teaching secondary math were neither more nor less effective than comparison teachers."[10]

However, there was a statistically significant effect in the TFA component with "[s]tudents assigned to TFA teachers scored 0.07 standard deviations higher on end-of-year math assessments than students assigned to comparison teachers."[11] The report included a figure that illustrated the estimated effectiveness of each CM relative its comparison teacher for the 111 classroom matches.[12] Included as Figure 1, the report observed, "without regard to statistical significance, the estimated difference in effectiveness between TFA and comparison teachers was positive in 60 percent of classroom matches (67 out of 111) and negative in the remaining 40 percent."[13] [14]

Figure 1

Differences Between the Effectiveness (z-score units) of TFA and Comparison Teachers from Individual Classroom Matches as Reported in Mathematica Study



The report also stated that the 0.07 SDs was equivalent to "2.6 months of additional learning for the average student nationwide."[15] It concluded with the recommendation that, "on average, principals of the secondary schools in the study would raise student math achievement by hiring a TFA teacher rather than a teacher from a traditional or less selective alternative route to teach the math classes examined in the study."[16]

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

The Mathematica experiment tested the hypothesis that high-achieving mathematics teachers from TFA (and Teaching Fellows) programs are more effective than the secondary mathematics teachers they work alongside in high-poverty schools. The experimental design, with random assignment of students to matched pairs of TFA and comparison teachers teaching the same mathematics course at the time in the same school, meant that "any difference in the outcomes of students taught by a TFA [teacher] and the outcomes of students taught by a comparison teacher can be attributed to differences in the teachers' effectiveness."[17] The report demonstrated that the matched classrooms "were statistically similar in terms of their baseline characteristics [which] suggests that random assignment was properly implemented and that student attrition after random assignment (due to lack of outcome test score data) did not lead to differences in observable baseline characteristics between the two groups."[18] Thus, the TFA effect of 0.07 SDs, significant at the 0.01 level, appeared to show that TFA's mathematics teachers are more effective than their mathematics teacher colleagues from less selective programs.

IV. The Report's Use of Research Literature

The report included a brief review of prior research on the effectiveness of teachers from TFA and Teaching Fellows programs. It devoted one paragraph to three non-experimental studies and a second to an experimental study that addressed the effectiveness of TFA's elementary teachers. In a different section, the report referenced a study by Darling-Hammond and her colleagues that examined the effectiveness of TFA teachers.[19] However, it did not address the critique leveled in this article, namely, that: "[t]wo studies have found evidence that TFA recruits' students achieve comparable or better gains in student learning when compared to other similarly experienced teachers in similar schools … but in both of these studies the comparison group teachers were also disproportionately untrained and uncertified teachers."[20]

V. Review of the Report's Methods

Conducted over a two-year period with two waves of teachers, the main Mathematica experiment could be divided into two: one including TFA's CMs and their mathematics teacher

colleagues and one including Teaching Fellows and their colleagues. To determine the relative effectiveness of these teachers, the investigators "estimated regression models in which student math scores were the dependent variable and the characteristics of the students' math teachers were the key independent variables of interest."[21] The regression included sandwich standard errors to account for the nesting of students in teachers in the matched classrooms. (The unit of analysis was student test scores.)

The TFA component included 45 schools, in 11 school districts, in 8 states. There were 111 matched classrooms taught by 66 CMs and 70 comparison teachers, with some participating in two or more matched pairs. Thirty-one matches included sixth grade students, 33 seventh grade students, and 19 eighth grade students. Twenty-eight matches were at the high school level, across Algebra I, General Math, Geometry, and Algebra II classrooms. Middle school student growth was measured using state assessments of mathematics. As no state exams were available for the high school matches, the investigators administered pre- and post-exams developed by the Northwest Evaluation Association (NWEA). In total, the Mathematica experiment included student scores from more than 60 different exams – exams that largely were not designed to measure student growth. Although not scaled and equated, student growth was standardized for the study. To ensure comparable levels of prior mathematics achievement, students were randomly assigned to matched classrooms.

The teacher sample ostensibly was sufficiently representative to provide generalizable results about the relative effectiveness of TFA's secondary mathematics teachers;[22] that is, "the comparison teachers represented a meaningful and appropriate counterfactual—the types of teachers that would have been available had teachers from TFA … not been available in a particular school."[23] To support this claim, the report included descriptive statistics for the participating schools and those that TFA "partners with" nationally, asserting that "[e]ven though study schools were not randomly selected from the full set of secondary schools employing TFA teachers nationwide, the study schools were similar to secondary schools employing TFA teachers nationwide along many dimensions."[24]

However, the report did not demonstrate that the participating teachers were, in fact, representative of the CMs and comparison teachers working in TFA's partner schools nationally. Specifically, while it provided descriptive information about the teachers' demographic, educational, and professional backgrounds, it included no information about the certification status of the comparison teachers and how they compared with non-TFA teachers working alongside CMs in "high-poverty" schools nationally. Thus, it was impossible to know how many of the comparison teachers were uncertified or teaching on emergency certificates, certified to teach but not to teach secondary mathematics, or certified to teach secondary mathematics. In an email exchange, the lead investigator informed us:

> We asked some of the certification questions … in the cohort 1 survey, but unfortunately these questions seemed to create some confusion—there were a variety of inconsistencies in teachers' responses indicating they did not understand the different types of certification we asked about. For that reason, for the second cohort, we revised the questions on this topic to correspond to the actual certificates available in each state, and then we recoded these to correspond to broader categories of certification. Because of our concerns

about the quality of the cohort 1 data we decided not to include them in the report.

Given the aforementioned critique of prior TFA effectiveness studies,[25] the lack of valid information on the certification of the comparison teachers in the Mathematica experiment was problematic.

VI. Review of the Validity of the Findings and Conclusions

The National Center for Educational Statistics granted us restricted access to the Mathematica study data. As part of our secondary data analysis, we created analytic files and, from that, descriptive statistics that closely matched those in the Mathematica report. We then used hierarchical linear modeling (HLM), including classroom-match fixed effects and sandwich standard errors, with a goal of matching Mathematica's effectiveness results. Like Mathematica's regression model, our HLM produced estimated effects for the effectiveness of each CM relative to their comparison teacher for the classroom matches. Our HLM came very close to matching their effect sizes and significance levels, including results by teacher subgroups, but were not exact matches and hence we assumed our HLM model was a close match. Using our model, and disaggregating results by teacher subgroup, we were able to explain the experimental results in terms of selection bias and other possible irregularities.

Our secondary analysis demonstrates that the Mathematica experiment was systematically biased in favor of CMs and that this more than explained the TFA effect. Specifically, we show that weak comparison teachers were oversampled, that first-year CMs were under-sampled, and that these sampling issues tilted the experiment to TFA. We then demonstrate that the TFA effect was particularly pronounced at the high school level, which is consistent with the fact that, at that level, CMs were likely to be teaching to the study tests whereas the comparison teachers were not. We conclude by showing that the investigators' translation of 0.07 SDs to 2.6 months of additional learning also was misleading and argue for additional large-scale studies of TFA and its teachers conducted by a diverse range of scholars.

VI.a. Weak Comparison Teachers

The comparison teachers had such weak credentials that they could not have been representative of secondary mathematics teachers that work alongside CMs in high-poverty U.S. schools. Table 1 illustrates this, comparing the qualifications of the comparison teachers with those of a nationally representative sample of eighth grade mathematics teachers from low-income, high-minority schools.[26]

Table 1

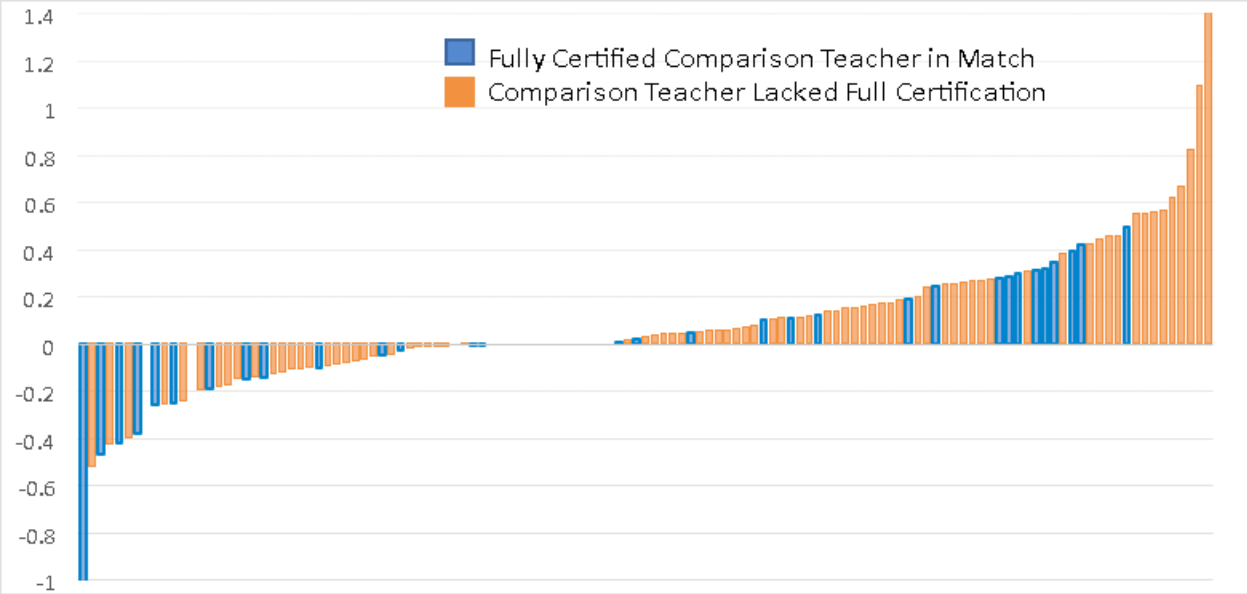
Descriptive Statistics for The Comparison Teachers[27] in the Matched Classrooms Compared with 8th Grade Mathematics Teachers in High-Poverty Schools Nationally

		Fully Certified in Any Subject	Fully Certified in Secondary Math	Degree[28] in Mathematics or Secondary Math Education
8th Grade Teachers High-Poverty Schools Nationally	Yes	> 80	» 80	78.6
	No	< 20	» 20	22.4
8th Grade Comparison Teachers in TFA Experiment	Yes	40.0	40.0	53.3
	No	60.0	6.7	40.0
	Missing	0	53.3	6.7
Comparison Teachers Teachers in TFA Experiment	Yes	42.4	41.6	32.9
	No	57.6	2.4	58.2
	Missing	0	56.0	8.9

Figure 2

Comparative Effectiveness Disaggregated by Comparison Teacher Certification [29]

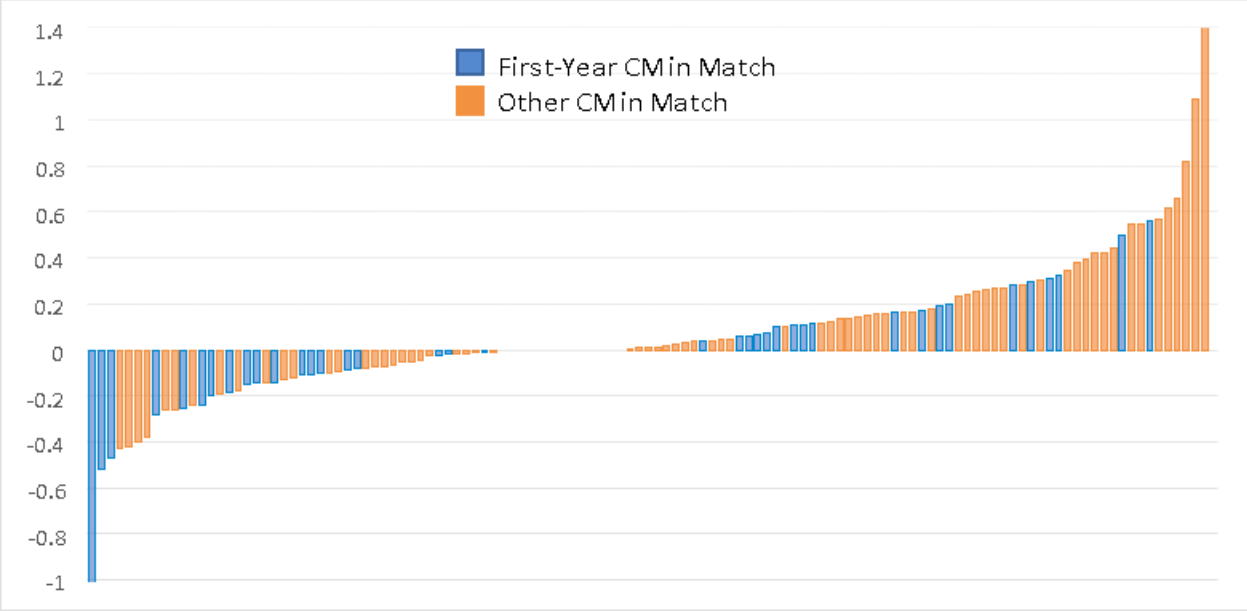
Using our HLM, and differentiating between certified and non-certified comparison teachers, we created Figure 2. It shows that only one of the comparison teachers in the 10 *most positive matches* reported being certified (Figure 2, right side), whereas this held for 5 of the 10 comparison teachers in the *most negative matches* (Figure 2, left side). The implication is that it the TFA effect is largely, if not entirely, explained by an unusually high proportion of uncertified comparison teachers.



VI.b. Too Few First-Year TFA Corps Members

The under-representation of first-year CMs posed a second validity threat to the experiment. In particular, not discussed were precautions taken to avoid the oversampling of more experienced, desirable, and effective CMs. However, the experiment included inclusion-exclusion rules for participation that may have produced those outcomes. As noted, the experiment only included teachers in schools that had hired TFA teachers prior to the study. Thus, teachers in schools that began hiring TFA teachers in the first or second year of the study were excluded. As TFA places first-year CMs, but generally not more experienced CMs, the vast majority of CMs in the excluded schools would be novices.

Figure 3
Comparative Effectiveness Disaggregated by First-Year and More Experienced CMs



The Mathematica report lumped first- and second-year CMs together when reporting descriptive statistics, never presenting disaggregated data. The restricted access data showed that 35.7% of the 111 classroom matches were taught by first-year CMs whereas 42.8% were taught by second-year CMs. To be representative, these percentages should have been reversed. This for two reasons: First, some 5-10% of novice CMs do not return for a second year of teaching.[30] Second, annually over the period from 2003 to 2015, TFA more than tripled the size of its incoming cohorts,[31] meaning, in a representative sample, first-year CMs would have outnumbered second-year CMs. Hence, a conservative estimate is that an additional 8 to 10 classroom matches should have included a first-year CM for the experimental sample to be representative.

The relative proportion of first-year CMs is important as early-career teachers, inclusive of CMs, become more effective as they gain experience.[32] Hence, the fewer first-year CMs in the experiment advantaged TFA. The secondary analysis bears this out. Specifically, Figure

3 shows that only 2 of the 16 *most positive matches* included a first-year CM (Figure 3, right side) whereas 8 of the 16 *most negative matches* included a first-year CM (see Figure 3, left side). Hence, the inclusion of more, and a more representative sample of, first-year CMs certainly would have reduced, if not eliminated, the TFA effect.[33] [34]

VI.c. Teaching to the Study Tests

It was impossible to know the extent to which the participating teachers were teaching to the study tests as the Mathematica study treated teaching as a black box. However, the CMs were more likely than the comparison teachers to have taught to the experimental tests. This because teaching to the tests is part-and-parcel of TFA's theory of action. Specifically, as part of teacher training, TFA requires CMs to learn to teach to the state-required exams and supports them to do so once they are in the classroom.[35] Hence, during the experiment, TFA's instructional coaches – currently called Managers of Teacher Leadership Development – would have met regularly with participating CMs to have them analyze the state-required tests for mathematics, organize their instruction to cover the tested material, and use released items to benchmark student achievement at regular intervals throughout the school year. TFA also would have held participating CMs accountable for producing "measurable gains" using students' pre- and post-performance on exams.[36]

In a 2011 grant proposal to the U.S. Department of Education,[37] TFA elaborated on its test-centric approach:

> We rely on high-quality student achievement data from rigorous tests to drive program improvements and provide tailored support to corps members. ... [W]e recently designed a new student achievement measurement system (SAMS) to generate even better, more useful data for corps members and program staff. SAMS is a transparent system that supports classroom learning by helping corps members and program staff set and manage toward ambitious yet feasible and measurable goals for their students ... The system is grounded in the aspiration that our corps members perform at the level of highly effective teachers; corps members are therefore measured against a benchmark that reflects a high-performing classroom where students achieve high rates of student growth. At the beginning of the year, corps members work with program staff to determine what their specific classroom benchmark should be, based on historical test score data.

Having entered teaching through a range of pathways, the comparison teachers would not have received the same level of test-centric training, support, and pressure as did the CMs. As new teachers, CMs had coaches and other forms of induction support that the far more experienced comparison teachers did not have access to; 72.2% of CMs reported having a mentor (or coach) during the experiment, whereas only 24.5% of comparison teachers did so.

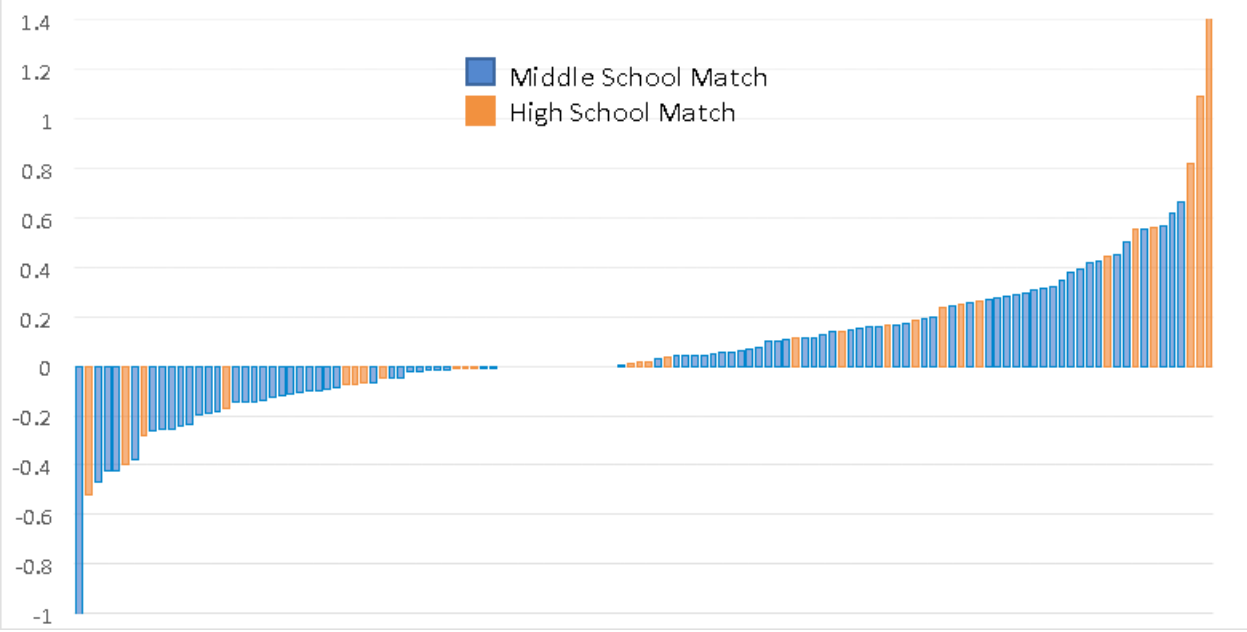
In the absence of state-required exams, TFA's coaches would have helped CMs access validated tests to anchor their instruction and assess student learning.[38] Because the No Child

Left Behind Act mandated state-required tests for Grades 3 through 8, but not high school, TFA staff would have supported its CMs in high schools to access validated exams during the experiment. At the time of the study, TFA's preferred tests for high school mathematics were the NWEA exams.[39] About this, a Consortium for Policy Research in Education document[40] reported that:

> [In 2007], TFA launched a pilot of Northwest Evaluation Association (NWEA) Measures of Academic Progress (MAP) assessments to standardize evaluation of teacher performance to increase student learning. The pilot consisted of 350 teachers and 15,000 students in 10 regions. [In 2008], TFA will double the number of teachers using NWEA growth assessments.

Given the scope of this undertaking, some of thousand-plus CMs trained to use NWEA exams undoubtedly participated in the Mathematica experiment which, again, ran from 2009 to 2011. In contrast, the comparison teachers likely had no exposure to the NWEA exams and would have had no reason to align their instruction to them. In email correspondence, the lead investigator informed us that both the, "treatment and control teachers of [G]rades 9 through 12 knew that their students would be tested, but none were informed that the study would use the NWEA subject tests."

Figure 4
Comparative Effectiveness Disaggregated by High School and Middle School Matches



If CMs were teaching to the NWEA tests, this likely would manifest in larger positive comparative effects for CMs in the high school matches than in the middle grades matches. The secondary analysis supported this hypothesis. Specifically, Figure 4 shows that, although only 29.8% of the total matches were at the high school level, half of the 12 *most positive matches* were at the high school level. Moreover, by restricting our HLM to high school

matches, we found that students of CMs in the high-school matches outperformed students in comparison classrooms by about 0.12 SDs (p-value < 0.05). In contrast, at the middle grades level, students of CMs outperformed those in comparison classrooms by only 0.05 SDs, a result that was not significant at the 0.05 alpha level (p-value of .065).

VI.d. Interpreting The TFA Effect

The Mathematica report asserts that the TFA effect was meaningful enough to make policy and hiring recommendations. Based on the experimental outcomes, the investigators concluded, "our study suggests that, on average, principals of the secondary schools in the study would raise student math achievement by hiring a TFA teacher rather than a teacher from a traditional or less selective alternative route to teach the math classes examined in the study."[41]

However, the TFA effect of 0.07 SDs is small. Randomized control studies of educational interventions at the middle grades level produce effect sizes of 0.51 SDs of student growth on average.[42] Reviews of research comparing effective and more typical teachers show that a one-unit increase in teacher effectiveness is on the order of 0.20 to 0.40 SDs.[43] The Mathematica report would have the reader believe otherwise, reporting that the TFA effect of 0.07 SDs was equivalent to "2.6 months of additional learning for the average student nationwide."[44]

To explain this translation, the investigators briefly referenced a table compiled by Hill, Bloom, Black and Lipsky.[45] Reproduced as Table 2, this table presents the standardized values for students' average annual gains (with margins of error) on nationally normed tests of mathematics for grade-level transitions from fifth-to-sixth grade through eleventh-to-twelfth grade. It provides empirical benchmarks for interpreting student growth using nationally normed tests at different grade levels. Through email correspondence, the lead Mathematica investigator confirmed that the team made the following calculation: (0.07÷0.27)*10 months ≈ 2.6 months, where the 0.27 is the unweighted average of the six mean annual gains shown in Table 2 and 10 months is the approximate length of a school year. However, Hill et al. (2008) actually recommend that, for experimental studies, researchers use a different set of benchmarks for interpreting effect sizes. Table 3 reproduces those benchmarks. Using the Table 3 values and a weighted average for students in the matches, the 0.07 SDs effect would be equivalent to 1.55 months of additional learning for the average student nationwide. But even this estimate is problematic, as the study tests were not specifically designed to measure annual growth.

Table 2
Average Annual Gain in Effect Size from Nationally Normed Mathematics Tests Grades 5-12 Transition Results[46]

Grade Transition	Mean (SDs)	Margin of Error
Grade 5-6	0.41	±0.08
Grade 6-7	0.30	±0.06
Grade 7-8	0.32	±0.05
Grade 8-9	0.22	±0.10
Grade 9-10	0.25	±0.07
Grade 10-11	0.14	±0.16

Table 3[47]
Summary of Effect Sizes from Randomized Studies[48]

Achievement Measure	Number of Effect Size Estimates	Mean Effect Size	SD
Middle school	36	0.51	0.49
High school	43	0.27	0.33

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

The Mathematica experiment has been celebrated in some quarters as proof that TFA improves outcomes for students in high-poverty and high-minority schools.[49] On its current website, TFA reports that, "[a] 'gold standard' study commissioned from Mathematica Policy Research, Inc. by the U.S. Department of Education ... found that students of Teach For America teachers learned **2.6 months more mathematics in a year** than students in the same schools taught by teachers from traditional preparation programs or less selective alternative route programs."[50] In funding proposals, TFA uses the study results to claim that, "TFA teachers have a statistically significant, positive impact on student achievement, in a variety of subject areas and grade levels."[51]

Our secondary analysis of the Mathematica experiment casts serious doubt on the validity of such claims. We demonstrated that, by not safeguarding against threats to internal validity, the Mathematica experiment favored TFA and its CMs. We showed that, rather than superior teachers, the TFA effect was the result of biased teacher sampling and experimental contamination (i.e., CMs likely teaching to the study tests). In the aggregate, these flaws were more than sufficient to produce a statistically significant effect on the order 0.07 SDs.

In sum, the 2013 Mathematica experiment should not be seen as putting to rest debates about the effectiveness of TFA and its CMs. At best, it is premature to recommend that "on average, principals of the secondary schools in the study would do well to hire a corps member rather than a teacher from a traditional or less selective alternative route to teach the math classes examined in the study." We submit that consequential educational policy recommendations should not be based on single studies, even those that are well designed and implemented, and particularly not those with serious flaws.

Notes and References

1 Clark, M., Chiang, H., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). *The effectiveness of secondary math teachers from Teach For America and the teaching fellows programs* (NCEE 2013-4015). Washington, DC: Institute of Education Sciences. Retrieved October, 12, 2013, from https://ies.ed.gov/ncee/pubs/20134015/pdf/20134015.pdf

2 TFA refers to its first- and second-year teachers as corps members. For simplicity, we use the term corps member (CM) to refer to any practicing teacher certified through a TFA program.

3 The results of the TFA experiment were subsequently published:

Chiang, H.S., Clark, M.A., & McConnell, S. (2017). Supplying disadvantaged schools with effective teachers: Experimental evidence on secondary math teachers from Teach for America. *Journal of Policy Analysis and Management*, *36*(1), 97-125.

4 See, for example: Levy, H.O. (2000). Why the best don't teach. *New York Times*, *9*.

5 Auguste, B.G., Kihn, P., & Miller, M. (2010). *Closing the talent gap: Attracting and retaining top-third graduates to careers in teaching: An international and market research-based perspective*. McKinsey.

6 Clark, M., Chiang, H., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). *The effectiveness of secondary math teachers from Teach For America and the teaching fellows programs* (NCEE 2013-4015). Washington, DC: Institute of Education Sciences. Retrieved October, 12, 2013, from https://ies.ed.gov/ncee/pubs/20134015/pdf/20134015.pdf

7 TNTP 2019 Website: https://tntp.org/about-tntp

8 See:

Mead, S., Chuong, C., & Goodson, C. (2015). *Exponential growth, unexpected challenges: How Teach For America grew in scale and impact*. Boston, MA: Bellwether. Retrieved September 1, 2017, from https://bellwethereducation.org/publication/exponential-growth-unexpected-challenges-how-teach-america-grew-scale-and-impact;

TNTP 2019 Website: https://tntp.org/about-tntp

9 Clark, M., Chiang, H., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). *The effectiveness of secondary math teachers from Teach For America and the teaching fellows programs* (NCEE 2013-4015). Washington, DC: Institute of Education Sciences, p. F.3. Retrieved October, 12, 2013, from https://ies.ed.gov/ncee/pubs/20134015/pdf/20134015.pdf

10 Clark, M., Chiang, H., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). *The effectiveness of secondary math teachers from Teach For America and the teaching fellows programs* (NCEE 2013-4015). Washington, DC: Institute of Education Sciences, p. xix. Retrieved October, 12, 2013, from https://ies.ed.gov/ncee/pubs/20134015/pdf/20134015.pdf

11 Clark, M., Chiang, H., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). *The effectiveness of secondary math teachers from Teach For America and the teaching fellows programs* (NCEE 2013-4015). Washington, DC: Institute of Education Sciences, p. xix. Retrieved October, 12, 2013, from https://ies.ed.gov/ncee/pubs/20134015/pdf/20134015.pdf

12 Clark, M., Chiang, H., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). *The effectiveness of secondary math teachers from Teach For America and the teaching fellows programs* (NCEE 2013-4015). Washington, DC: Institute of Education Sciences, p. 57. Retrieved October, 12, 2013, from https://ies.ed.gov/ncee/pubs/20134015/pdf/20134015.pdf

13 Clark, M., Chiang, H., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). *The effectiveness of secondary math teachers from Teach For America and the teaching fellows programs* (NCEE 2013-4015). Washington, DC: Institute of Education Sciences, p. 57. Retrieved October, 12, 2013, from https://ies.ed.gov/ncee/pubs/20134015/pdf/20134015.pdf

14 The comparison teachers in the TFA experiment also had much weaker mathematics backgrounds than those in the Teaching Fellow experiment. Specifically, 80.2% of the comparison teachers in the Teaching Fellows experiment reported having a post-secondary degree in mathematics or secondary mathematics education whereas only 32.9% of the comparison teachers in the TFA experiment reported the same. While programmatic differences could partially account for this (e.g., Teaching Fellows trains a greater proportion of high school mathematics teachers), both programs place new teachers in similar – often the same – high-poverty schools.

15 Clark, M., Chiang, H., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). *The effectiveness of secondary math teachers from Teach For America and the teaching fellows programs* (NCEE 2013-4015). Washington, DC: Institute of Education Sciences, p. xix. Retrieved October, 12, 2013, from https://ies.ed.gov/ncee/pubs/20134015/pdf/20134015.pdf

16 Clark, M., Chiang, H., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). *The effectiveness of secondary math teachers from Teach For America and the teaching fellows programs* (NCEE 2013-4015). Washington, DC: Institute of Education Sciences, p. xxxii. Retrieved October, 12, 2013, from https://ies.ed.gov/ncee/pubs/20134015/pdf/20134015.pdf

17 Clark, M., Chiang, H., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). *The effectiveness of secondary math teachers from Teach For America and the teaching fellows programs* (NCEE 2013-4015). Washington, DC: Institute of Education Sciences, p. 1. Retrieved October, 12, 2013, from https://ies.ed.gov/ncee/pubs/20134015/pdf/20134015.pdf

18 Clark, M., Chiang, H., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). *The effectiveness of secondary math teachers from Teach For America and the teaching fellows programs* (NCEE 2013-4015). Washington, DC: Institute of Education Sciences, p. 51. Retrieved October, 12, 2013, from https://ies.ed.gov/ncee/pubs/20134015/pdf/20134015.pdf

19 Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., & Heilig, J. V. (2005). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness. *Education Policy Analysis Archives, 13*(42). Retrieved December, 21, 2007, from https://epaa.asu.edu/ojs/article/view/147

20 Decker, P.T., Mayer, D.P., & Glazerman, S. (2004). *The effects of Teach For America on students: Findings from a national evaluatio*n. Princeton, NJ: Mathematica Policy Research, Inc. Retrieved December, 21, 2007, from https://www.mathematica-mpr.com/our-publications-and-findings/publications/the-effects-of-teach-for-america-on-students-findings-from-a-national-evaluation

21 Clark, M., Chiang, H., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). *The effectiveness of secondary math teachers from Teach For America and the teaching fellows programs* (NCEE 2013-4015). Washington, DC: Institute of Education Sciences, p. xix. Retrieved October, 12, 2013, from https://ies.ed.gov/ncee/pubs/20134015/pdf/20134015.pdf

22 There were three restrictions that affected the teacher sample. First, the comparison teachers were not to have entered teaching through a selective teacher certification program. The Mathematica report noted that any excluded teachers would have come from, “five [selective] programs that, collectively, trained only 14 math teachers nationwide in 2007” (p. 9). Second, the study schools had to be sufficiently large to enable the random assignment of students to two or more teachers teaching the same secondary course at the same time. The investigators did not address (e.g., through proximity matching) how this decision might have influenced the experimental outcomes. Third, the study only included schools that had a history of partnering with TFA and, hence, excluded schools that began a partnership in the first or second year of the study. We discuss this

decision later as it likely lead to the undersampling of first-year CMs.

23 Clark, M., Chiang, H., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). *The effectiveness of secondary math teachers from Teach For America and the teaching fellows programs* (NCEE 2013-4015). Washington, DC: Institute of Education Sciences, p. 9. Retrieved October, 12, 2013, from https://ies.ed.gov/ncee/pubs/20134015/pdf/20134015.pdf

24 Clark, M., Chiang, H., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). *The effects of Teach For America on students: Findings from a national evaluatio*n (NCEE 2013-4015). Washington, DC: Institute of Education Sciences, p. 39.

25 See:

Decker, P.T., Mayer, D.P., & Glazerman, S. (2004). *The effects of Teach for America on students: Findings from a national evaluation*. University of Wisconsin-Madison, Institute for Research on Poverty. Retrieved December, 21, 2007, from https://www.ssc.wisc.edu/irpweb/publications/dps/pdfs/dp128504.pdf;

Darling-Hammond, L., Holtzman, D.J., Gatlin, S.J., & Heilig, J.V. (2005). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness. *Education policy analysis archives 13,* 42. Retrieved December, 21, 2007, from https://epaa.asu.edu/ojs/article/view/147

26 Data from:

Rahman, T., Fox, M.A., Ikoma, S., & Gray, L. (2017). *Certification status and experience of US public school teachers: Variations across student subgroups* (NCES 2017-056). Washington, DC: US Government Printing Office. National Center for Education Statistics. Retrieved September 23, 2018, from https://nces.ed.gov/pubs2017/2017056_report.pdf

Note that we were unable to find comparable data from the study years, 2009-2011. However, the qualifications of secondary mathematics teachers likely have not changed much in the period since the experiment was conducted in 2009-2011.

27 These rates are based on individual study participants. Similar distributions to those above held for descriptive statistics based on classroom matches.

28 Degree includes a Master's degree or undergraduate major or minor in either mathematics or secondary mathematics education or both.

29 Missing bars for matches in which the certification information of comparison teachers was missing.

30 Donaldson, M.L., & Johnson, S.M. (2011). Teach For America teachers: How long do they teach? Why do they leave?. *Phi Delta Kappan*, *93*(2), 47-51.

31 Mead, S., Chuong, C., & Goodson, C. (2015). *Exponential growth, unexpected challenges: How Teach For America grew in scale and impac*t. Boston, MA: Bellwether. Retrieved September 1, 2017, from https://bellwethereducation.org/publication/exponential-growth-unexpected-challenges-how-teach-america-grew-scale-and-impact

32 See:

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. *Education Finance and Policy 1*(2): 176-216.

Darling-Hammond, L., Holtzman, D.J., Gatlin, S.J., & Heilig, J.V. (2005). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness. *Education policy analysis archives*, *13*(42). Retrieved December, 21, 2007, from https://epaa.asu.edu/ojs/article/view/147

33 Those first-year CMs in the included schools likely would have received more support comparatively than first-

year CMs in the included schools. As the investigators observed: "[TFA's r]egional staff typically try to place new corps members in schools where current or former TFA corps members are teaching, so that the new teachers can benefit from an added source of support. The programs in the study had typically succeeded in arranging such placements over the three school years preceding our interviews. On average, across fall 2007, 2008, and 2009, according to program representatives, an estimated 92 percent of all new teachers and 93 percent of new secondary math teachers were hired in schools where current or former corps members were teaching" (p. 35).

34 A related source of potential bias, also unaddressed in the report, concerns the likely undersampling of CMs who were involuntarily transferred the year prior to the study. In particular, as they would have been moved out of a historical TFA partner school, such CMs would have been excluded from the experiment. Involuntary transfers often occur when principals dismiss or refuse to rehire untenured, early-career teachers based on observations of their performance. On average, teachers who are involuntarily transferred appear to be less effective than other teachers (Feng and Sass, 2011). Thus, the exclusion of any such CMs from the experiment would have further tilted the experiment towards TFA.

35 See:

Brewer, T.J., & DeMarrais, K.B. (Eds.). (2015). *Teach For America counter-narratives: Alumni speak up and speak out*. New York, NY: Peter Lang.

Veltri, B.T. (2010). *Learning on other people's kids: Becoming a Teach For America teacher*. Charlotte, North Carolina: Information Age Publishing.

36 Blanchard, O. (2013). I quit Teach for America. *The Atlantic*, *23*. Retrieved March 23, 2018, from https://www.theatlantic.com/education/archive/2013/09/i-quit-teach-for-america/279724/

37 Teacher for America. (2015). *Investing in a pipeline of effective educators for high-need students: Sustaining and bringing innovation to Teach for America's summer training institutes.* Proposal for Supporting Effective Educator Development (SEED) Grant, U.S. Department of Education. Retrieved March 23, 2018, from https://www2.ed.gov/programs/edseed/2015/tfanarr.pdf

38 Mead, S., Chuong, C., & Goodson, C. (2015). *Exponential growth, unexpected challenges: How Teach For America grew in scale and impact*. Boston, MA: Bellwether. Retrieved September 1, 2017, from https://bellwethereducation.org/publication/exponential-growth-unexpected-challenges-how-teach-america-grew-scale-and-impact

39 See:

Northwest Evaluation Association (NWEA). (2011). NWEA Returns to Homegrown Roots and Celebrates New Headquarters. Retrieved from: https://www.nwea.org/2011/01/nwea-returns-homegrown-roots-celebrates-new-headquarters/

40 Goetz, M., & Aportela, A. (2008). *Strategic management of human capital: Teach for America*. Madison, WI: University of Wisconsin, Wisconsin Center for Education Research, Consortium for Policy Work in Education.

41 Clark, M., Chiang, H., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). *The effectiveness of secondary math teachers from Teach For America and the teaching fellows programs* (NCEE 2013-4015). Washington, DC: Institute of Education Sciences, p. xxxii. Retrieved October, 12, 2013, from https://ies.ed.gov/ncee/pubs/20134015/pdf/20134015.pdf

42 Hill, C.J., Bloom, H.S., Black, A.R., & Lipsey, M.W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child development perspectives*, *2*(3), 172-177.

43 See:

Nye, B., Konstantopoulos, S., & Hedges, L.V. (2004). How large are teacher effects? *Educational evaluation*

and policy analysis, *26*(3), 237-257;

Rowan, B., Correnti, R., & Miller, R.J. (2002). *What large-scale, survey research tells us about teacher effects on student achievement: Insights from the prospectus study of elementary schools*. CPRE Research Reports. Retrieved February 1, 2019, from https://repository.upenn.edu/cpre_researchreports/31/

44 Clark, M., Chiang, H., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). *The effectiveness of secondary math teachers from Teach For America and the teaching fellows programs* (NCEE 2013-4015). Washington, DC: Institute of Education Sciences, p. xix. Retrieved October, 12, 2013, from https://ies.ed.gov/ncee/pubs/20134015/pdf/20134015.pdf

45 Hill, C.J., Bloom, H.S., Black, A.R., & Lipsey, M.W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child development perspectives*, *2*(3), 172-177.

46 Adapted from: Hill, C.J., Bloom, H.S., Black, A.R., & Lipsey, M.W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child development perspectives*, *2*(3), 172-177.

47 Adapted from: Hill, C.J., Bloom, H.S., Black, A.R., & Lipsey, M.W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child development perspectives*, *2*(3), 172-177.

48 Unweighted means across all effect sizes and samples in each category. Values compiled from 61 existing research reports and publications.

49 See:

Mead, S., Chuong, C., & Goodson, C. (2015). *Exponential growth, unexpected challenges: How Teach For America grew in scale and impact*. Boston, MA: Bellwether. Retrieved September 1, 2017, from https://bellwethereducation.org/publication/exponential-growth-unexpected-challenges-how-teach-america-grew-scale-and-impact;

50 Teach for America (2019). TFA Website: *Our Impact*. Retrieved April 15, 2019, from https://www.teachforamerica.org/what-we-do/impact

51 Teacher for America. (2015). *Investing in a pipeline of effective educators for high-need students: sustaining and bringing innovation to Teach for America's summer training institutes*. Proposal for Supporting Effective Educator Development (SEED) Grant, U.S. Department of Education. Retrieved March 23, 2018, from https://www2.ed.gov/programs/edseed/2015/tfanarr.pdf, p. 8.