# Teaching for All?
# Teach For America's Effects on the Distribution of Student Achievement[1]

May 11, 2013


## Abstract

This paper examines the effect of Teach For America (TFA) on the distribution of student achievement in elementary school, extending previous experimental research that finds a positive average effect in math but no average effect in reading. I build on this research by examining how student achievement in TFA and non-TFA classrooms differs across the broader distribution of student achievement by estimating quantile treatment effects. Distributional results reveal a positive impact of TFA teachers across the distribution of math achievement. However, students at the bottom of the reading distribution scored the same or worse in TFA classrooms, while in the upper half of the distribution, students in TFA classrooms outperformed students in non-TFA classrooms. This heterogeneity in the effect of TFA on student achievement is obscured in research examining only mean impacts, and has important implications for teacher recruitment and training. These results suggest that traditional teacher certification programs may be able to learn from TFA practices in math, and that TFA may be able to learn from traditional certification programs how to better equip teachers to help students who struggle with reading.


*Key words:* Teach For America, Quantile Treatment Effects, student achievement

"At Teach For America, we define and measure our teachers' success in terms of how much their students learn." (Farr, 2010, p. 2)

Teach For America (TFA) is a rapidly expanding, influential alternative teacher certification program that seeks to radically improve educational opportunities and achievement for low income youth in underserved urban and rural communities across the US (Kopp, 2011). To do so, TFA recruits recent college graduates and returning professionals from selective backgrounds, and provides them with five weeks of intensive training and ongoing support during a two-year commitment in their placement classrooms. In 2012, more than 10,000 TFA corps members taught 750,000 students in 46 regions across 36 states and the District of Columbia (Barahona, 2012). Lauded and critiqued for its impact on a variety of educational domains (Miner, 2010; Rotherham, 2011; Veltri, 2010), TFA judges its own success and failure based on student academic outcomes in its teachers' classrooms.

In focusing on student academic outcomes, TFA specifically orients its teachers towards goals of mastery, closing gaps, and helping each student to become academically successful and engaged (Farr, 2010). Yet, the impact of TFA on student achievement is one of the most contested aspects of TFA's record. A growing body of work examines the average effect of TFA on student achievement in a variety of contexts, yielding contradictory results, with some finding that TFA teachers outperform their non-TFA counterparts (Glazerman, Mayer, & Decker, 2006) and others finding the opposite (Darling-Hammond, Brewer, Gatlin, & Vasquez Heilig, 2005). These mixed results suggest that TFA may not have a uniform impact for all types of students in all contexts, but might instead generate heterogeneous effects. Developmental science suggests that many types of interventions affect students differently (Duncan & Vandell, 2011), and existing examinations of TFA have only begun to consider how its effects might vary across different types of students and schools (Glazerman et al., 2006; Xu, Hannaway, & Taylor, 2011).

I contribute to the growing literature on the impacts of TFA, and seek to inform larger conversations about teacher training and effectiveness in underserved communities by examining how the effect of TFA teachers varies across the distribution of student achievement. Although the average effect of TFA on student achievement is helpful for summarizing the impact of this complex intervention, TFA does not necessarily strive to have the same impact on all of the students it serves. By motivating its teachers to pursue goals related to closing achievement gaps and having students master grade-level content, TFA teachers might boost achievement the most among low achieving students. This leads to an interest in determining how TFA affects the distribution of achievement more broadly, and whether TFA teachers affect different parts of the distribution equally.

**Prior Research on Teach For America**

The relationship between TFA and student achievement has been examined using a variety of experimental, quasi-experimental, and descriptive research designs, in many different contexts, and over a wide range of grades. To date, only one study uses an experimental research design to isolate the causal effect of TFA on student achievement. Glazerman et al. (2006) evaluate the effect of TFA on student achievement in grades 1-5 using a random assignment design at six TFA sites. They find that students assigned to TFA teachers outperform students in non-TFA classrooms in mathematics, but not reading. TFA corps members outperform both novice and veteran teachers, though there are no statistically significant differences when first year TFA corps members are compared to all control teachers.

These findings are replicated by a quasi-experimental study of high school students by Xu et al. (2011). Exploiting the fact that students may have TFA teachers in some subjects but not others, Xu et al. (2011) compare students' performance across high school subjects in the same school year. They find that students with TFA science teachers score higher than students with non-TFA science teachers, regardless of the certification and experience level of the comparison teachers. They find similar, but smaller, effects in English, while in math students with TFA teachers only outperform students of similarly-experienced non-TFA teachers, but score no differently than students in classrooms with more experienced math teachers.

Several other studies use a variety of quasi-experimental and descriptive designs to evaluate the relationship between TFA and student achievement, finding mixed results across different settings. One group compares TFA teachers only to novice teachers trained through other traditional and alternative certification pathways. These include Boyd, Grossman, Lankford, Loeb, & Wyckoff's (2006) study of teachers in grades 3-8 in New York City, Henry et al. (2010) who analyze teachers in grades 3-12 in North Carolina, Ware et al. (2011) who compare teachers in grades 3-11 in Texas, and the Strategic Data Project (2012) which evaluates teachers in grades 3-9 in Los Angeles. All four studies find that TFA teachers outperform other novices in math (and in science in North Carolina), with mixed effects for reading. A second group of studies compares TFA teachers to both novice and veteran teachers. In New York, TFA teachers outperform all traditionally certified teachers in math, but not reading (Kane, Rockoff, & Staiger, 2008). In Houston, TFA teachers outperform both novice and veteran teachers in math and reading (Raymond, Fletcher, & Luque, 2001), while in Texas middle schools there is a TFA advantage in math among novice teachers as well as a benefit in math and reading when comparing TFA alumni teachers to other veteran teachers (Turner, Goodman, Adachi, Brite, & Decker, 2012). Others, however, find that TFA teachers outperform novice teachers, but are either statistically indistinguishable from (Noell & Gansle, 2009; Schoeneberger, Dever, & Tingle, 2009) or worse than certified veteran teachers (Darling-Hammond et al., 2005). Finally, research from Arizona suggests that TFA teachers may be less effective than veteran teachers and no different than other novice teachers (Laczko-Kerr & Berliner, 2002).

In sum, while research design rigor varies greatly, these results suggest that TFA may not have a uniform effect across different sites, across different subjects, and across all students. In general, TFA teachers appear to have a more positive effect on math and science, and less impact on reading or language arts. Finally, in most cases, TFA teachers outperform other novice teachers, and the most rigorous research suggests that they outperform veteran teachers as well.

**A Distributional Perspective on Teach For America**

Distributional studies show that many education policy interventions do not have a uniform effect on all intervention recipients. For example, research on a school voucher incentive program in Colombia suggests that vouchers are effective for the lower tail of the achievement distribution, but not elsewhere (Lamarche, 2007). In contrast, distributional research on Project STAR suggests that the largest test score gains from small classes are at the top of the achievement distribution (Jackson & Page, 2013). Similarly, Arulampalam, Naylor, & Smith (2012) find that the effects of absences on achievement are largest for students in the upper tail of the achievement distribution. These studies highlight the ways in which evaluations of average impact miss heterogeneous effects throughout the distribution.

We might hypothesize several different distributional patterns for the effect of TFA. Consider Glazerman et al.’s (2006) finding of a positive average effect of TFA teachers on math achievement. It could be the case that TFA teachers are equally effective at improving student achievement at all points of the achievement distribution, relative to counterfactual teachers in the same schools. This might occur because TFA teachers are more effective than non-TFA teachers, and they are equally more effective across high, low, and average-achieving students.

Alternatively, the effect of having a TFA teacher might be larger at the top of the distribution if TFA teachers are especially good at teaching the engaged, higher-performing students working on more challenging material because they can draw from their elite educational background and experiences. In this case, TFA teachers might excel at teaching students like themselves while other students might not benefit from having a TFA teacher.

But the opposite pattern could occur as well. Suppose that the highest-performing students do well regardless of the teacher they have because they always work hard and their learning is not affected as much by their teacher. Or suppose that TFA’s emphasis on helping all students meet mastery goals will disproportionately impact the bottom of the distribution where students are furthest from these goals, as opposed to the students near accountability thresholds (Booher-Jennings, 2005). In either of these cases, we might expect low-performing students to gain the most from being in TFA classrooms.

These scenarios are all plausible, and it is possible that these and other alternatives are occurring simultaneously. Given the variation in average effectiveness between reading and math in previous work, it is also possible that different patterns could hold across these two subjects which may highlight that experience matters differentially for these subjects in elementary school. It is thus helpful to know how the entire distribution of student achievement is affected by having a TFA teacher. Identifying variation in the impact of TFA can help to identify which areas of TFA teacher practice and training may be most fruitful for other teacher preparation programs to emulate, as well as highlighting where TFA might need to improve.

## Data

To understand whether the effect of TFA varies across the distribution of student achievement, this study estimates the Quantile Treatment Effects (QTE) of being randomly assigned to a TFA teacher versus a non-TFA teacher. Data were collected by Mathematica Policy Research during the 2001-2002 and 2002-2003 school years, and come from six of the 15 regions that were active in 2001, including Baltimore, Chicago, Los Angeles, Houston, New Orleans, and the Mississippi Delta.[2] These six regions were randomly selected to represent the mix of districts served by TFA (predominately black vs. Hispanic, urban vs. rural), and within each region schools were randomly chosen from among the schools with at least one TFA and one non-TFA teacher at the same grade level. The experiment was restricted to teachers in grades 1 to 5. The final sample included 100 classrooms at 17 schools, and a total of 1,969 students.

Students in selected classrooms took the Iowa Test of Basic Skills (ITBS) in reading and math in the fall and spring of the given school year. For my dependent variable, I use the Normal Curve Equivalent (NCE) math and reading scores, which are age-adjusted, and nationally normed to have a mean of 50 and a standard deviation of 21.06.[3] Given that first grade students' reading scores are available from only a portion of the reading test, I do not include first graders in my analysis.[4] I also restrict my analysis to students who have at least one fall and one spring test score. In addition, I treat previously unidentified invalid test scores (a raw score of 99), given to 9.5 percent of 2$^{nd}$ through 5$^{th}$ graders in the public use data, as missing.[5] Descriptive statistics for the analytic sample are presented in Table 1, separately for TFA and non-TFA classrooms. The two groups are well balanced on their demographic characteristics, as can be seen from the statistically insignificant joint test of equality across all measures.

[Insert Table 1 Here]

## Method

To examine the effect of TFA on the distribution of student achievement, this paper estimates quantile treatment effects (QTE) (Firpo, 2007). QTE allow for unconditional comparisons of the achievement distributions of TFA and non-TFA students, thus providing much more information on the nature of treatment effects than mean differences. [6] In the context of experimental data, QTEs are estimated by calculating the difference in the two marginal distributions (cumulative distribution functions, or CDFs). Using these CDFs, I examine the difference between these two distributions at various percentiles of the outcome variable, ITBS reading or math test scores. For example, I estimate the QTE at the 0.50 quantile by subtracting median test score of non-TFA students from the median test score of TFA students.

As an example, Figure 1 and Figure 2 Panel A show the CDFs and QTE for unweighted baseline math normal curve equivalent scores. Figure 1 shows the CDF for baseline math scores in TFA and non-TFA classrooms. The CDF presents math NCE scores on the x-axis with the cumulative percent of the sample on the y-axis. The horizontal distance between these CDFs at each point in the distribution, which equals the difference in NCE scores, is the quantile treatment effect at that percentile. Included on Figure 1 are two vertical lines indicating the ITBS national mean (at 50 NCE points) and the unweighted mean for the non-TFA classrooms (for math fall scores this is at 31.5 NCE points) to enable comparisons of this sample relative to the national average. These lines underscore that the majority of the sample scores below the national average.

Figure 2 Panel A shows the corresponding QTE for the CDF shown in Figure 1, so that the x-axis in represents the cumulative percentiles of the distribution, while the y-axis represents the difference in NCE scores between TFA and non-TFA classrooms at each percentile. The difference between the scores (solid line) is plotted along with pointwise 95% confidence intervals (dashed lines), which are calculated by stratifying on block and treatment status and bootstrapping the estimates 999 times. Figure 2 Panel A shows that the bulk of the QTE point estimates are zero or close to zero for the baseline math scores. Two exceptions are at the upper and lower tails. Between the 6th and 10th percentiles, there is a negative and significant difference between treatment and control. Above the 90th percentile there is also a negative difference between treatment and control, but it is not statistically significant. The confidence intervals do not always include 0, particularly at the lower tail, suggesting some imbalance across the distribution in random assignment. Two vertical lines for the national average and the test-specific non-TFA classroom average are also included on Figures 2-5. As the x-axis now shows cumulative percentiles rather than NCE scores, these lines now indicate the percentile of the distribution at which the relevant NCE score occurs.

[Insert Figures 1 and 2 here]

Figure 2 Panel B mirrors Panel A, assessing the degree to which randomization successfully balanced fall scores across the distribution of reading achievement. Panel B shows distributional differences between TFA and non-TFA classrooms on fall reading achievement. These differences between TFA and control classrooms are negative beginning above the 64th percentile, and /are larger and significant above the 90th percentile. This suggests that randomization was even less successful for reading than math, with non-TFA classrooms having more higher-performing students at the outset.

To address the lack of balance on fall scores, I use an inverse propensity-score weighting approach as a nonparametric first step (Firpo, 2007), which allows me to balance baseline test scores across the two groups and to account for differences in the likelihood of being assigned to a TFA or non-TFA teacher in different grade levels and schools.[7] This also allows me to adjust for differences in the presence of non-response and invalid test scores by including indicator variables for whether students had missing or invalid test scores. To implement this approach, I first use a logistic regression model to predict assignment to a TFA or non-TFA teacher among the full sample included in the public release data as a function of randomization block, baseline test score deciles for math and reading, whether the student had valid or invalid missing values for fall or spring scores, and the baseline demographic characteristics listed in Table 1.[8] From this model, I calculate the predicted probability of being in the treatment group $\hat{p}$, and then construct weights of 1 /$\hat{p}$ for those in the treatment group and 1 /(1-$\hat{p}$) for the control group. As shown by the p-values for mean comparisons in Table 1, these weights balance the treatment and control groups on these observable dimensions. Further, a test of joint significance for these characteristics is not significant, suggesting that, when using the inverse propensity score weights, there are no differences across random assignment groups. Propensity score weighting is useful because allows me to obtain unconditional estimates while still adjusting for any post-randomization imbalance in baseline test scores, and missing values.

The fall QTE, adjusted using the inverse propensity score weights, are shown in Figure 3 Panels A and B. Confidence intervals for the point estimates are calculated by bootstrapping within randomization block and treatment status 999 times, re-estimating test score deciles and the logistic regression for the inverse propensity scores in each sample.

[Insert Figure 3 here]

As shown in Figure 3 Panels A and B, the inverse propensity score weight described above is quite successful at balancing differences between treatment and control across the entire distribution for both math and reading. For math, the confidence intervals always include zero, and for most of the distribution, the point estimates are at or near zero.

Likewise, for reading, the point estimates stay close to zero for most of the distribution, and the confidence intervals always include zero. At the lower tail, the confidence intervals are fairly wide. [9] At the very upper tail, above the 90th percentile, the point estimates become negative fairly large, but are still not significant. Overall, the inverse propensity score weight adjusted QTE show balanced samples across the fall distributions in both reading and math. The same weights and procedure are used to estimate the spring QTE for reading and math.

## Results

The QTE results for the spring math and reading tests are presented in Figures 4 Panels A and B. As with Figures 2 and 3, Figure 4 plots the differences between students in TFA and non-TFA classrooms (y-axis), for each percentile of the distribution (x-axis). Thus, when the solid line is above zero, students in TFA classrooms are scoring at higher levels than in non-TFA classrooms, and when the solid line is below zero students in TFA classrooms are scoring lower than those in non-TFA classrooms. These differences are statistically significant when the area between the two dashed lines (representing the 95 percent confidence intervals) does not include the line marking zero on the y-axis. We see, for example, in Panel A of Figure 4 that the median (50th percentile) score in TFA classrooms is two points higher than the median score in non-TFA classrooms, while the 80th percentile score in TFA classrooms is 6 points higher than the 80th percentile score in non-TFA classrooms.

Overall, the QTE results suggest that TFA has impacts that vary across the distribution of math and reading, but in different ways. Across nearly the entire distribution of math, the estimates of the effect of TFA are positive, and for most of the distribution, the confidence intervals show that these differences are statistically significant. Even though we cannot rule out an effect of zero for some portions of the distribution, we can largely rule out a negative effect on math achievement, except at the very upper tail. Thus, writ large, the effect of math can be characterized as positive and shared throughout most of the distribution, though in some parts of the distribution it might be more accurate to characterize it as non-negative.

The magnitude of the differences is quite striking: the TFA effect is around 5 points around the 75th percentile, corresponding to .24 of a standard deviation. Relative to the average fall to spring gain of 2.1 NCE points (.10 of a standard deviation) for the full sample, this difference is substantial.[10] This impact is in line with effects observed in other teacher-related interventions that are much more intensive, such as KIPP.[11]

[Insert Figure 4 Here]

Figure 4 Panel B shows the QTE for spring reading scores. Here the QTE plot suggests that whether having a TFA teacher is beneficial, no different, or detrimental varies across the distribution. At the lower tail, the point estimates of the effect of TFA are negative. Although the confidence intervals indicate that this negative effect is only statistically significant for a very small part of the distribution, they provide strong evidence that the effect of TFA on the bottom of the reading distribution is not positive. In contrast, above the 40th percentile the point estimates are positive, and with a few exceptions the confidence intervals indicate that the effect of TFA on this part of the distribution is not negative, and in some places is positive. Combined, this pattern of results yields an average effect that is near zero and statistically insignificant; however, these differences have potentially important ramifications for teacher recruitment and training.

In terms of magnitude, the largest negative difference observed in reading is 4 NCE points, corresponding to a negative effect of .19 standard deviations, with differences of 2 or 3 NCE points being quite common below the 30th percentile. At the upper tail, although the effects are not consistently significant, the positive effects again typically range from 2-4 NCE points, corresponding to positive effects of .09-.19 standard deviations. Relative to the average yearly gain for the sample in reading of .013 SD, this difference is substantial.

## Discussion

This paper extends prior research on Teach For America by examining how student achievement in TFA and non-TFA classrooms varies across the distribution of achievement. It identifies variation in effects that was previously hidden by examining only average impacts of TFA. The distributional findings reveal different patterns for math and reading.

In math, students assigned to TFA teachers outperformed control students throughout most of the distribution, but especially for the upper half. This is consistent with overall positive effects of TFA on math observed in Glazerman et al. (2006), but it highlights that TFA teachers do not have a uniform impact across the distribution, and that their impact is larger among high achieving students. Overall, however, the distributional results speak to fairly wide-spread effectiveness on the part of TFA teachers relative to same-school comparison teachers, and suggest that important insights might be gleaned from observations of TFA mathematics teaching.

The reading results are especially noteworthy, as the previous research on TFA effects in reading suggest that there is no significant benefit of having a TFA teacher on reading achievement (Glazerman et al., 2006; Xu et al., 2011). However, this null mean effect appears to be concealing important distributional differences. In reading, TFA teachers appear to have a positive impact on the middle and top of the distribution. In contrast, students at the bottom of the reading achievement distribution in TFA classrooms if anything are scoring worse than their counterparts in non-TFA classrooms. This suggests that TFA teachers are not as successful at supporting these lowest-level readers as well as traditionally trained teachers. Thus, it may be useful for TFA to consider ways to address literacy instruction to support its lowest-level readers, and compare the practices in counterfactual classrooms that are aimed at this group.

An important consideration when contextualizing these findings is to recall that the distributions presented here are only comparable to schools similar to those in which TFA teachers are placed. The results cannot be generalized to schools that are in higher-performing neighborhoods and districts with no TFA presence. Both the TFA teachers and the control teachers in this sample are working to improve achievement among students well below national averages. As Figure 1 highlights, although some students among this sample perform above the national average, the national mean represents approximately the 87$^{th}$ percentile for this sample, while the mean score for this sample is in the low thirties for both reading and math, which is nearly one standard deviation below the national average. This is particularly relevant in interpreting the results for reading, as the students who are in the 30$^{th}$ percentile and below on reading in this sample are scoring at least 1.5 standard deviations below the national average, suggesting that they might potentially benefit from targeted reading interventions requiring specialized training.

These results highlight how examining treatment heterogeneity can be useful for identifying the portions of the distribution where a particular intervention is and is not successful. However, it is also important to verify these results through replication using other data sources. In particular, while the Mathematica data are able to isolate the effect of TFA through random assignment, few principals have the ability to randomly assign students to classrooms, so replication in administrative data and across other studies would strengthen the external validity of these conclusions.

In sum, when evaluating TFA teachers by TFA’s own rubric—student achievement—they appear to be doing better than non-TFA teachers at boosting elementary math achievement for students across the distribution, while in reading TFA has effects that appear to be positive for much of the top of the distribution, and negative at the bottom. While TFA is highly contentious, my results suggest that both proponents and detractors are partially right: TFA clearly raises math scores throughout the distribution, while in reading it appears to raise scores for high achievers and lower scores for low achievers. This pattern of results suggests that TFA has valuable lessons for traditional certification programs concerning how to more effectively teach mathematics, while traditional certification programs can provide valuable insights into how to effectively teach struggling readers. To this end future work examining how TFA and non-TFA classrooms differ in pedagogy, classroom organization, and targeted intervention would provide valuable insights into how teachers from different backgrounds can work together to meet the needs of our underserved students.

**References**

Angrist, J. D., Dynarski, S. M., Kane, T. J., Pathak, P. A., & Walters, C. R. (2010). Inputs and impacts in charter schools: KIPP Lynn. *The American Economic Review*, *100*(2), 239–243.

Antecol, H., Eren, O., & Ozbeklik, S. (2013). *The effect of teach for America on the distribution of student achievement in primary school: Evidence from a randomized experiment*. IZA Discussion Paper No. 6453.

Arulampalam, W., Naylor, R. A., & Smith, J. (2012). Am I Missing Something? The Effects of Absence from Class on Student Performance. *Economics of Education Review*, *31*(4), 363–375.

Barahona, G. (2012). *Teach For America announces the schools contributing the most graduates to its 2012 teaching corps* (Press Release). New York, N.Y.: Teach For America. Retrieved from http://www.teachforamerica.org/sites/default/files/20120905_Press.Release_Top.Contributers.pdf

Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas Accountability System. *American Educational Research Journal*, *42*(2), 231–268.

Boyd, D. J., 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., Brewer, D. J., Gatlin, S. J., & Vasquez Heilig, J. (2005). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness. *Education Policy Analysis Archives*, *13*(42), 1–42.

Decker, P. T., Mayer, D. P., & Glazerman, S. (2004). *The effects of Teach for America on students: Findings from a national evaluation* (No. 8792-750). Princeton, NJ: Mathematica Policy Research, Inc.

Duncan, G. J., & Vandell, D. L. (2011). *Understanding Variation in the Impacts of Human Capital Interventions on Children and Youth*. Working Paper, Irvine Network on Interventions in Development.

Farr, S. (2010). *Teaching as leadership: The highly effective teacher's guide to closing the achievement gap*. San Francisco, CA: Jossey-Bass.

Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, *75*(1), 259–276.

Glazerman, S., & Grinder, M. (2004). *Mathematica Teach For America Evaluation Public Use Documentation*. Mathematica Policy Research, Inc.

Glazerman, S., Mayer, D., & Decker, P. (2006). Alternative routes to teaching: The impacts of Teach for America on student achievement and other outcomes. *Journal of Policy Analysis and Management*, *25*(1), 75–96.

Henry, G. T., Thompson, C. T., Bastian, K. C., Fortner, C. K., Kershaw, D. C., Purtell, K. M., & Zulli, R. A. (2010). *Portal report: teacher preparation and student test scores in North Carolina*. Chapel Hill, NC: Carolina Institute for Public Policy, The University of North Carolina at Chapel Hill.

Hoover, H. D., Dunbar, S. B., & Frisbie, D. A. (2007). Iowa tests of basic skills (ITBS). *Rolling Meadows, IL: Riverside Publishing*.

Jackson, E., & Page, M. E. (2013). Estimating the Distributional Effects of Education Reforms: A Look at Project STAR. *Economics of Education Review*, (32), 92–103.

Kane, T., Rockoff, J., & Staiger, D. O. (2008). What Does Teacher Certification Tell Us About Teacher Effectiveness? Evidence from New York City. *Economics of Education Review*, *27*(6), 615–631.

Kopp, W. (2011). *One day, all children* (3rd ed.). Cambridge, MA: Public Affairs, NY Perseus Book Group.

Laczko-Kerr, I., & Berliner, D. C. (2002). The effectiveness of "Teach for America" and other under-certified teachers on student academic achievement: A case of harmful public policy. *Education Policy Analysis Archives, 10*(37). Retrieved from http://epaa.asu.edu/epaa/v10n37/

Lamarche, C. (2007). Voucher program incentives and schooling performance in colombia: a quantile regression for panel-data approach. *preprint, University of Oklahoma*.

Miner, B. (2010). Looking past the spin: Teach for America. *Rethinking Schools Online*, *24*(3). Retrieved from http://www.rethinkingschools.org/archive/24_03/24_03_TFA.shtml.

Noell, G. H., & Gansle, K. A. (2009). *Technical report Teach for America teachers' contribution to student achievement in Louisiana in grades 4-9: 2004-2005 to 2006-2007*. Baton Rouge, LA: Louisiana State University.

Raymond, M., Fletcher, S. H., & Luque, J. (2001). *Teach for America: An evaluation of teacher differences and student outcomes in Houston, Texas*. Houston, TX.

Riverside Publishing. (2012). Iowa Tests of Basic Skills (ITBS_ Forms A, B, and C). Retrieved April 2, 2013, from http://www.riversidepublishing.com/products/itbs/details.html

Rotherham, A. (2011, February 20). Teach for America: 5 Myths That Persist 20 Years On - TIME. *Time Magazine*. Retrieved from http://www.time.com/time/nation/article/0,8599,2047211,00.html

Schoeneberger, J. A., Dever, K. A., & Tingle, L. (2009). *Evaluation of Teach For America in Charlotte-Mecklenburg Schools*. Center for Research and Evaluation Office of Accountability: Charlotte-Mecklenburg Schools.

Strategic Data Project. (2012). *SDP Human Capital Diagnostic: Los Angeles Unified*. Center for Education Policy Research: Harvard University.

Turner, H. M., Goodman, D., Adachi, E., Brite, J., & Decker, L. (2012). *Evaluation of Teach For America in Texas schools*. Edvance Research, Inc.

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

Ware, A., LaTurner, R. J., Parsons, J., Okulicz-Kozaryn, A., Garland, M., & Klopfenstein, K. (2011). *Teacher preparation programs and Teach for America research study*. Education Research Center: University of Texas at Dallas.

Xu, Z., Hannaway, J., & Taylor, C. (2011). Making a difference? The effects of Teach For America in high school. *Journal of Policy Analysis and Management*, *30*(3), 447–469.

Table 1

*Baseline characteristics of study sample and missing values*[1]

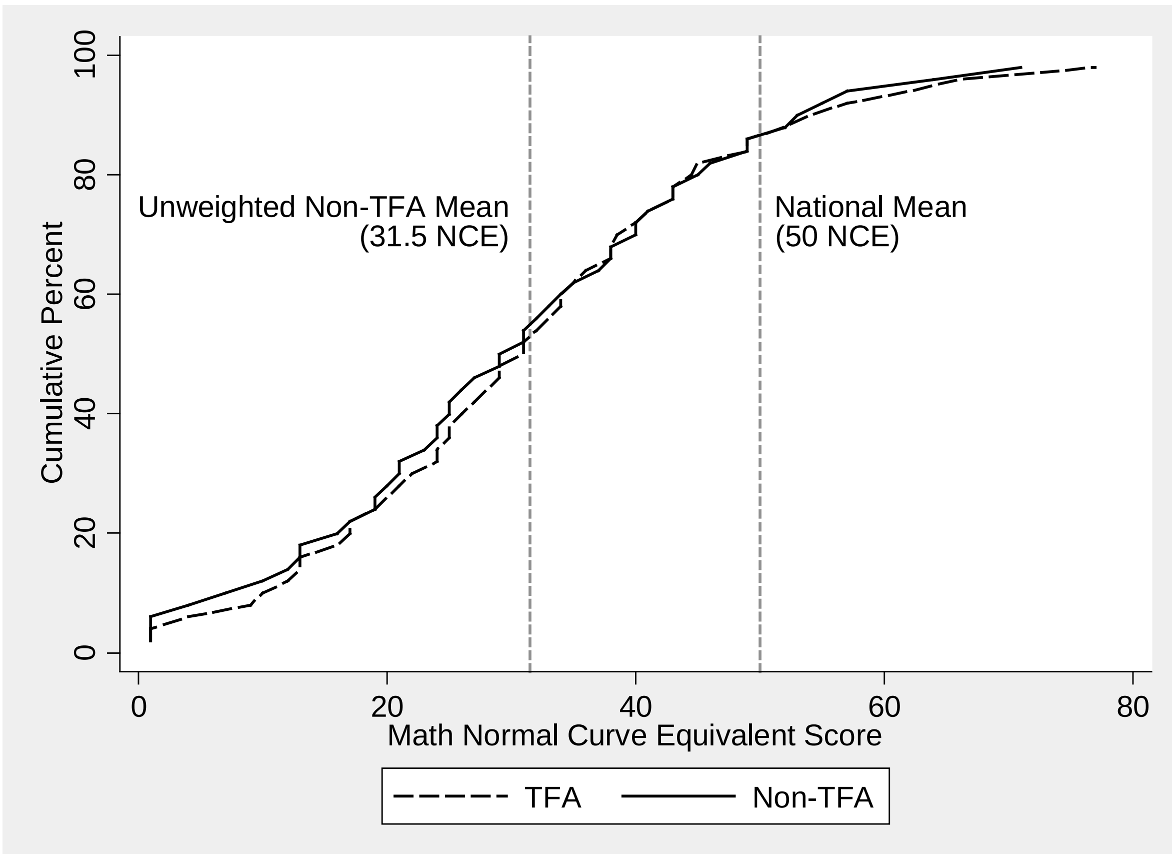
| | Control Mean | T-C Difference | SE | P-value[2] |
|---|---|---|---|---|
| Female | 0.474 | 0.018 | 0.018 | 0.319 |
| Black | 0.720 | -0.008 | 0.034 | 0.809 |
| Hispanic | 0.238 | 0.016 | 0.021 | 0.459 |
| Over age for grade | 0.231 | 0.018 | 0.024 | 0.447 |
| Free/reduced lunch eligible | 0.975 | -0.002 | 0.009 | 0.786 |
| Did not move classes during school year (stayer) | 0.928 | 0.018 | 0.014 | 0.195 |
| Percent of class in research sample at end of year | 0.817 | -0.018 | 0.030 | 0.556 |
| Math pretest Normal Curve Equivalent score | 30.638 | -0.177 | 0.840 | 0.834 |
| Reading pretest Normal Curve Equivalent score | 31.141 | 0.245 | 0.832 | 0.771 |
| Sample size: | N = 1430 | | | |
| Joint test for baseline child characteristics: | p = 0.224 | | | |

[1] Sample includes 2nd - 5th graders with at least one valid pre- and post-test

[2] P-values calculated after using inverse propensity score weights and clustering on randomization block

Figure 1.

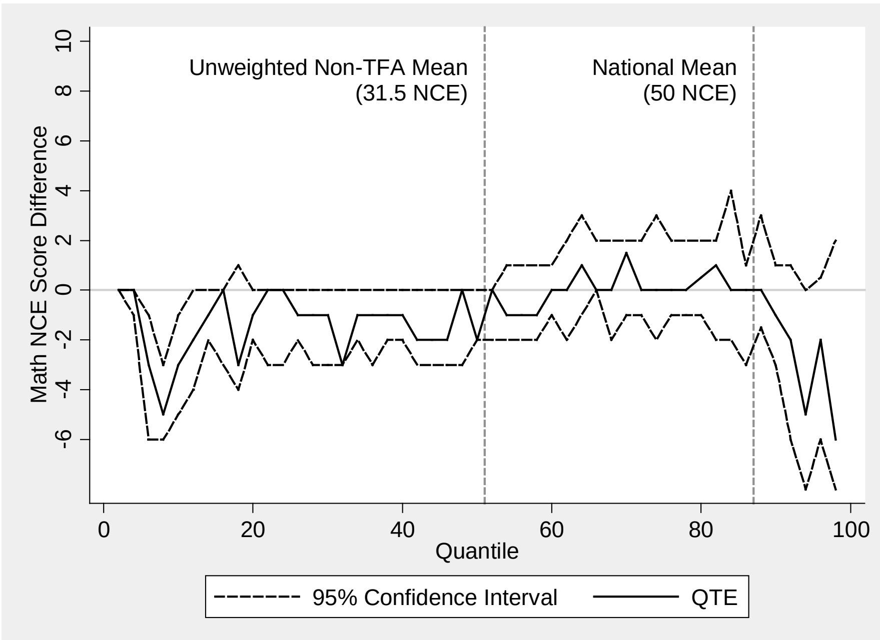
*Cumulative Density Functions (CDFs) for fall math achievement in TFA and non-TFA Classrooms*



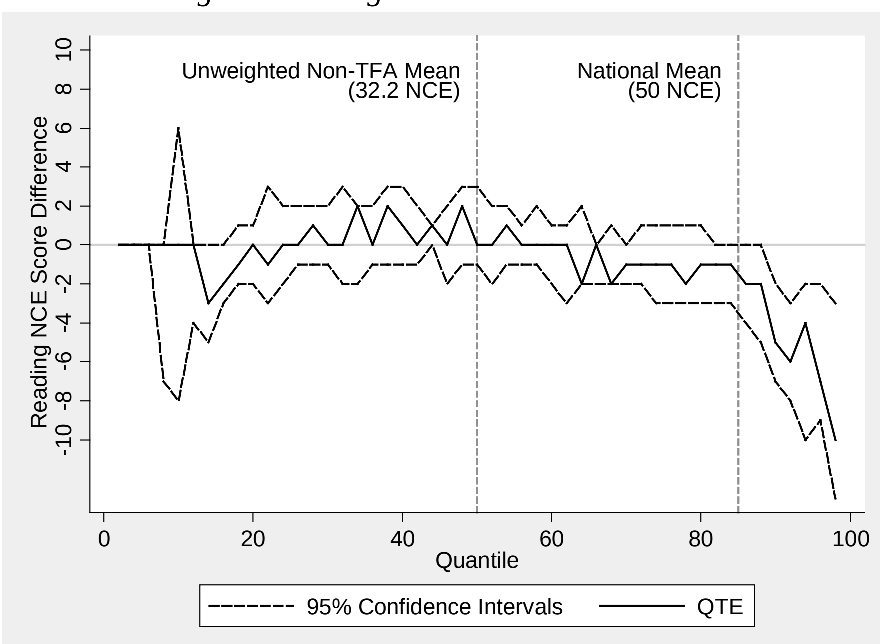

Notes: Figure shows cumulative distribution functions for baseline math Normal Curve Equivalent scores from the Iowa Test of Basic Skills separately for TFA classrooms and non-TFA classrooms. Estimates are unweighted. Data from the Mathematica Policy Research Teach For America Evaluation.

Figure 2. *Unweighted quantile treatment effect estimates of the impact of assignment to TFA classroom on reading Normal Curve Equivalent scores at baseline*

*Panel A. Unweighted Differences in Fall Math Achievement*



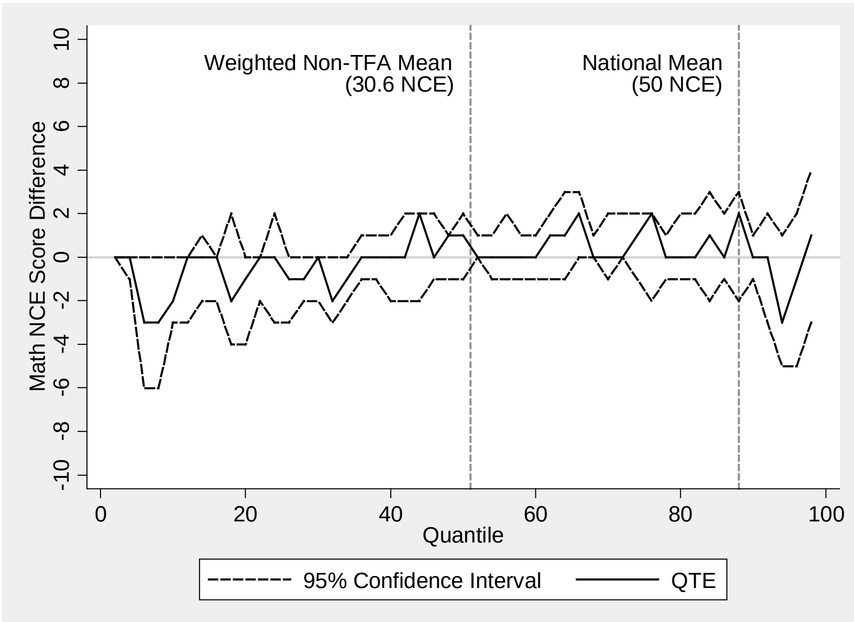

*Panel B. Unweighted Reading Pretest*



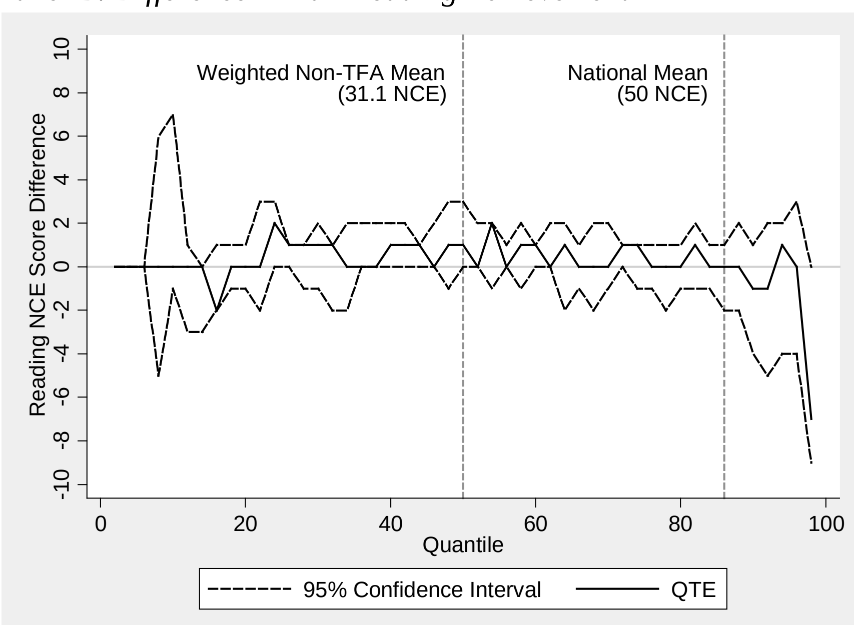

Notes: Panels A & B of figure show QTE for the effect of being assigned to a TFA classroom on math and reading Normal Curve Equivalent scores from the Iowa Test of Basic Skills at baseline. Estimates are unweighted. Data from the Mathematica Policy Research Teach For America Evaluation.

Figure 3. *Inverse propensity score weighted quantile treatment effect estimates of the impact of assignment to TFA classroom on reading Normal Curve Equivalent scores at baseline*

*Panel A. Difference in Fall Math Achievement*



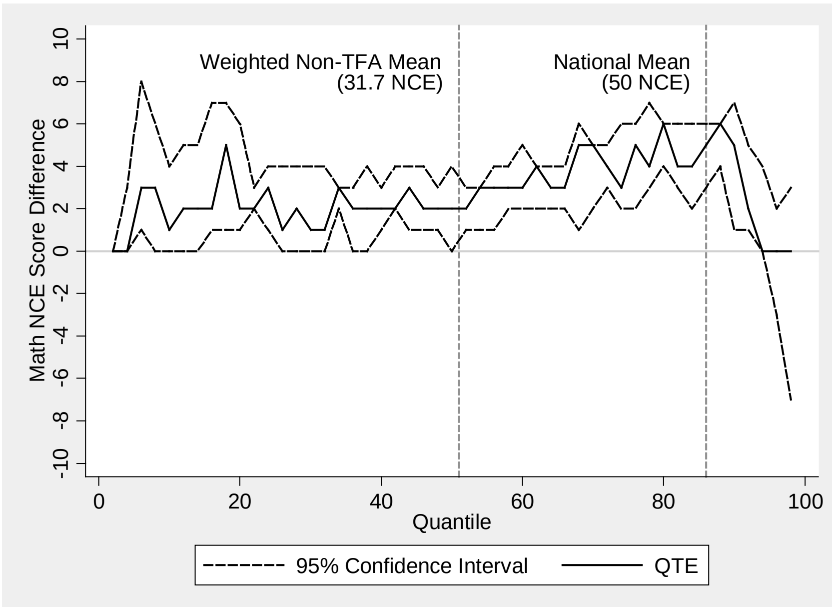

*Panel B. Difference in Fall Reading Achievement*



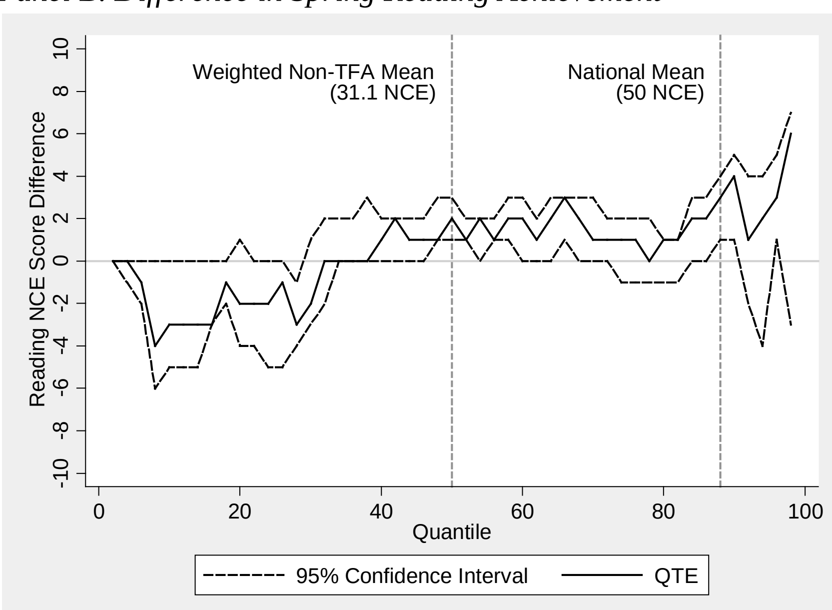

Notes: Panels A & B of figure show QTE for the effect of being assigned to a TFA classroom on math and reading Normal Curve Equivalent scores from the Iowa Test of Basic Skills at baseline. Estimates are weighted using inverse propensity score weights. Weights are 1 /$\hat{p}$ for treatment observations and 1/(1-$\hat{p}$) for control observations, where $\hat{p}$ is generated from a logistic regression of treatment status on baseline demographics, sample design variables, and baseline test score deciles. 95% CIs are obtained by bootstrapping with replacement within randomization block. Data from the Mathematica Policy Research Teach For America Evaluation.

Figure 4. *Inverse propensity score weighted quantile treatment effect of assignment to TFA classrooms on spring test scores*

*Panel A. Difference in Spring Mathematics Achievement*




*Panel B. Difference in Spring Reading Achievement*




Notes: Panels A & B of figure show QTE for the effect of being assigned to a TFA classroom on math and reading Normal Curve Equivalent scores from the Iowa Test of Basic Skills in the spring following random assignment. Estimates are weighted using inverse propensity score weights. Weights are 1 /$\hat{p}$ for treatment observations and 1 /(1-$\hat{p}$) for control observations, where $\hat{p}$ is generated from a logistic regression of treatment status on baseline demographics, sample design variables, and baseline test score deciles. 95% CIs are obtained by bootstrapping with replacement within randomization block. Data from the Mathematica Policy Research Teach For America Evaluation.

[1] The author would like to thank Mathematica Policy Research for data access and Steve Glazerman for invaluable assistance with data questions.

[2] The pilot region, Baltimore, was studied during the 2001-2002 school year. The other five regions were studied from 2002-2003.

[3] Normal Curve Equivalent scores are calculated from raw scores which are then normed based on grade and quarter of the school year (fall, winter, or spring) using the national ITBS sample, and converted into rankings such that the distribution of scores is normal. This allows for cross age and cross-grade comparisons of scores at equal intervals.

[4] While students in grades 2-5 received reading scores that were calculated using responses from tests of both vocabulary and word analysis, first grade reading tests were scored separately as vocabulary and word analysis, and no combined score is available (Glazerman & Grinder, 2004). To facilitate comparison across reading and mathematics achievement, I likewise only examine mathematics achievement among students in grades 2-5. Results presented are robust to the inclusion of first graders; in these analyses I use students' vocabulary scores to match Glazerman et al. (2006).

[5] Below the score of 99, the next highest raw score observed was 41, while the highest possible raw score at any grade level is 44 in reading and 50 in math (Hoover, Dunbar, & Frisbie, 2007; Riverside Publishing, 2012). Raw scores of 99 corresponded to NCE scores of 0. Communication with Riverside Publishing confirmed that 99 is an invalid raw score, and that 0 is an invalid NCE score.

[6] Concurrent with this paper, which was presented in February 2013 at the annual Sociology of Education Association meeting and in March at the spring meeting for the Society for Research on Educational Effectiveness, a working paper released by Antecol, Eren, & Ozbeklik (2013) in April also examines distributional differences in the effects of TFA on student achievement in elementary school using the MPR data. While I use inverse propensity score weights to calculate QTE, Antecol et al. use fixed effects quantile regression models which report differences at conditional quantiles (see Firpo, Fortin, and Lemieux 2009 for a discussion of the shortcomings of conditional quantiles). See the online supplement for further discussion of the distinctions between this paper and Antecol et al. (2013).

[7] Glazerman et al. (2006) include a sample normalization weight in their estimates to account for the fact that each block has a different number of TFA and non-TFA classrooms, with slightly different numbers of students in each

classroom, making the odds of assignment to treatment non-uniform. This variation is accounted for by including block fixed effects in the inverse propensity score weights.

[8] These variables are the control variables used in Glazerman et al. (2006).

[9] See the discussion on floor effects in the methodological supplement for more details.

[10] It is important to note that the ITBS is designed such that a year's worth of growth would lead a student to have the same NCE score from fall to spring. Students with higher scores than the previous year have improved more than a year's worth of growth and students with lower scores than the previous year have grown less than expected in a year (Decker, Mayer, & Glazerman, 2004).

[11] For example, a year of enrollment in KIPP Lynn in Massachusetts resulted in average effects of 0.35 SD in math and 0.12 SD in reading (Angrist, Dynarski, Kane, Pathak, & Walters, 2010).