

The Effectiveness of Teach for America: a Review of the Literature

Xiwen Wang

**Econ 495
Professor Smith
March 22, 2009**

I. Introduction:

Teach for America (TFA) is a national service corps through which recent graduates from top colleges and universities commit a minimum of 2 years to teach in K-12 schools situated in low-income communities. Since its inception in 1989, TFA has placed over 20,000 teachers across the country, impacting the lives of some 3 million students. With an operating budget of \$75 million in 2007, TFA now serves 29 regions, both rural and urban. In 2008, TFA selected and placed approximately 3,700 corps members from a pool of 24,718 applications; these individuals will join other current corps members to reach approximately 400,000 students this year. TFA's rapid expansion and rise to prominence has drawn critical attention from educators and researchers alike. Recent studies on the effects of TFA have yielded drastically differing conclusions, which range from unilaterally positive even when comparing TFA teachers to veteran teachers (Xu, Hannaway & Taylor, 2008) to unrelentingly negative (Darling-Hammond, Holtzman, Gatlin & Heilig, 2005). These discrepancies and the sizeable scale of the program motivate an in-depth assessment of the body of research that has been conducted on the effectiveness of TFA.

This analytical literature survey will examine a variety of studies that seek to compare the effectiveness of TFA teachers to non-TFA teachers as a function of two key differences, certification and experience. Effectiveness is determined by evaluating student achievement outcomes, which are typically measured using standardized testing scores. The earliest study included in this survey was conducted by Raymond, Fletcher, and Luque (2001), who examined student outcomes of TFA corps members and traditionally certified teachers in middle and elementary schools in Houston, Texas. They found that TFA teachers were as good as, if not better than, traditionally certified teachers. Laczko-Kerr and Berliner (2002) matched "under-

certified” teachers (including TFA corps members) with fully certified teachers in low-income school districts in Arizona and compared students’ state achievement test scores within each pair. They concluded that students of certified teachers outperformed those of under-certified teachers in every subject. Decker, Mayer and Glazerman (2004) conducted the first national evaluation of TFA. In this study, students were randomly assigned to teachers at the start of a school year, and then evaluated on their math and reading skills at the end of the school year. It was found that TFA teachers were on par with both novice and veteran teachers in reading, and more effective in math. Darling-Hammond, Holtzman, Gatlin and Heilig (2005) adopt a different analytical approach to re-evaluate the findings of Raymond et al. They concluded that TFA teachers are generally less effective than their traditionally certified counterparts, but TFA teachers who become certified are able to close the gap in teaching performance. Pilcher and Steele (2005) compared regularly certified first-year elementary school teachers with first- and second-year TFA corps members in Atlanta, Georgia using a combination of surveys that assessed teaching efficacy and classroom achievement tests. Their findings suggested that regularly certified teachers are more effective than TFA teachers, although this difference becomes smaller when comparing them to second-year TFA teachers. The most recent study (Xu, Hannaway & Taylor, 2008) undertakes the first examination of TFA effectiveness in high schools and conclude that on average, TFA teachers are more effective than non-TFA teachers, both veteran and novice teachers alike.

The remainder of this paper will proceed with a brief discussion of background information regarding TFA’s history, program structure, as well as its recruitment, training, and placement methods, followed by a close examination of the methodologies of the six studies outlined above. Incidentally, three of the studies described above reflect positively on TFA’s

performance while the other three reflect negatively, and so discussion will be structured around these two opposing bodies of evidence. Each study will be evaluated on both the quality of data obtained as well as its analytical approach. I will conclude with an overall assessment of TFA based on the existing body of evidence and outline recommendations for future research and consideration.

II. Background:

Teach for America was founded by Wendy Kopp in 1989 to address the stark educational inequities that plague urban and rural communities nationwide. Kopp, who was at that point a recent graduate from Princeton University, envisioned an organization that would draw top college graduates away from more lucrative careers in management consulting or investment banking for at least two years to teach in the most challenging schools around the country. The organization's early years were marked by financial difficulties, structural problems, and high attrition rates among its corps members, which immediately drew sharp criticism from educators and scholars. Linda Darling-Hammond (1994), the author of one of the studies discussed in this literature survey and a leading scholar in education at Stanford University, condemned TFA's eager, bright-eyed recruits for leaving a "trail of failure" behind them, citing as evidence the 58% attrition rate among those in the inaugural class. However, TFA has appeared to overcome many of its early difficulties. Its operating budget was \$75 million in 2007, compared to \$2.5 million in its first year of operation. On the recruiting front, TFA has also experienced success in attracting top graduates from the best institutions around the country. Between 2000 and 2003, the number of applicants to TFA increased almost fourfold, and it is estimated that some 11% of Yale University's graduating class applies to TFA each year, while 9% of Harvard's graduating

seniors apply. Over the years, TFA has continually refined its selection methods in an attempt to best match the criteria necessary to succeed in the classroom.

Before assessing the relative merits of studies on the effects of TFA, it is useful to examine the programmatic structure of TFA, including its selection and training of teachers. TFA recruits graduating college seniors or recent graduates of all academic majors, primarily at top universities and colleges around the country. The recruitment process involves an application and two rounds of interviews. The application consists of an online application, a letter of intent, a resume, and an essay. Selected candidates are invited to participate in a phone interview, and those who proceed in the selection process are evaluated over the course of a day-long interview, after which offers are made. TFA conducts a five-week training session for all of its newly inducted corps members during the summer prior to the start of their 2-year commitment. The training includes formal education courses, teaching summer school students, as well as other learning experiences such as content- and grade-specific workshops. In terms of deployment, TFA corps members are typically placed in some of the lowest-performing school districts in the country, particularly those that historically have had trouble attracting and retaining teachers. Each district typically only pays a fee to offset recruiting costs, while TFA pays for the teachers' salaries. The placement of each teacher in a particular region, grade level, and subject area is determined by a combination of his or her personal preferences and the relative supply and demand for teachers in the regions for which the individual expressed a preference.

As expected, TFA teachers systematically differ from their traditionally certified colleagues. For one, a higher percentage of TFA teachers hold Bachelor's degrees (Raymond, Fletcher & Luque, 2001), while a higher percentage of non-TFA teachers hold advanced degrees. A significantly higher percentage of TFA teachers graduate from selective colleges and

universities compared to their colleagues. The racial profile of TFA corps members is also quite different from other new hires; non-TFA teachers tend to be older and more ethnically diverse, with greater proportions of Latinos and African Americans. In one study, it was found that there was a higher proportion of male teachers in TFA than in the general population of teachers (Decker, Mayer & Glazerman, 2004). Xu et al. (2008) found that in North Carolina, where their study was conducted, TFA teachers scored slightly higher on the PRAXIS test (a standardized exam for admission into teacher education programs) and a higher percentage of TFA teachers were licensed in the subject they taught compared to non-TFA teachers. Finally, as mentioned before, TFA teachers are typically assigned to schools with lower socioeconomic conditions than new non-TFA teachers on average, as indicated by the higher proportions of TFA-taught students receiving free or reduced lunch.

III. Discussion of Literature:

The Case for TFA:

Three studies in this literature survey yield conclusions that suggest TFA teachers are just as, if not more, effective than traditionally certified teachers with varying years of experience. The first of these studies (Raymond, Fletcher & Luque, 2001) was conducted by the Center for Research on Education Outcomes (CREDO), an independent, non-partisan research group based at the Hoover Institution at Stanford University, and marks the first independent evaluation of TFA. The study examined elementary and middle school student and teacher data for the Houston Independent School District (HISD) from 1996-2000. Two comparison groups were constructed; one consisted of all non-TFA teachers in the district and the other consisted of only new non-TFA teachers. Student outcomes were measured by improvement in reading, languages arts, and math scores on the Texas Assessment of Academic Skills (TAAS), which were

standardized to have a mean of zero and a standard deviation of one for each year to control for test variation across different years.

Although random assignment of students to classrooms could not be conducted, or at least not to the degree of rigor required in experimental designs, the researchers attempted to control for variation in students, teachers, and schools through regression analysis. Several systematic differences were found between TFA teachers and non-TFA teachers in HISD. Firstly, all TFA teachers were considered to be under-certified and enrolled in an Alternative Certification Program (ACP) that took place during their first year of teaching; however, there was no data on what, if any, proportion of novice non-TFA were under-certified and by extension, what proportion of those individuals chose to enroll in an ACP. As expected, it was found that TFA teachers were assigned to higher-poverty schools than non-TFA teachers. Finally, the authors noted that school-wide average TAAS scores were consistently lower for TFA teachers than non-TFA teachers during their first and second years, while they are often higher during their third and fourth years (for those who elect to extend their time teaching). Although the statistical significance of this difference is not established, they attribute it to the reduced mobility of TFA teachers, who typically only move to better school districts after their two-year commitment.

The underlying model used in this study assumes that current performance is a linear function of prior testing performance and the characteristics of each student and those of his or her classmates and teacher. To separate the effects of experience and being trained through TFA, years of experiences were explicitly controlled for in the model. In sum, these student, class, and teacher control measures allow for isolation of the effects of a given school year, in order to construct what is commonly called a “value-added” model. The study uses both a “General”

model, which compares average achievement gains among classes with TFA teachers versus those with non-TFA teachers, as well as a “TFA Fixed Teacher Effects” model, which looks at student achievement gains across all students taught by each teacher. While the “General” model simply utilizes a dummy variable for TFA teachers, the “Fixed Teacher Effects” model utilizes a dummy variable for each teacher, thus eliminating the potentially unobserved heterogeneity that arises when multiple classes are taught by the same teacher. The Fixed Teacher Effects model also reveals information about the distribution of teacher performance among TFA and non-TFA teachers.

The study experienced some limitations in the construction of its data set and difficulties in the subsequent specification of its models. For one, it was difficult in some cases to correctly match students to teachers and class peers. This is because students in middle school often change teachers at the end of each semester, and teachers may teach multiple sessions of the same course each day, making it difficult to accurately determine peer groups for the students of those teachers. Furthermore, the fact that a majority of middle school students had multiple teachers for English makes a simple dummy variable for TFA insufficient for capturing the effects of having a TFA teacher. To address the first issue, the authors dropped class variables when none could be found. To address the second issue, they created a TFA “intensity” variable as a measure of the percentage of teaching that was provided by TFA teachers, which replaced the dummy variable for students who did not have a one-to-one match with a teacher in a particular subject area.

In total, sixteen regression models were specified across grade level, subject, and whether the comparison group consisted of new teachers or all teachers. The authors of the study concluded that TFA teachers were not only more effective on average than their peers, as

revealed through the General Model, but more consistently so as well. For elementary reading, TFA teachers were found to be statistically significantly more effective than new non-TFA teachers and on par with the overall pool of non-TFA teachers after controlling for experience. Students with TFA teachers gained an additional 5.8 percent of a standard deviation of test scores compared to those with new non-TFA teachers. For elementary mathematics, students with TFA teachers gained an additional 12 percent compared to those with new non-TFA teachers and an additional 2.9 percent when compared to students of all non-TFA teachers. For both reading and mathematics, the Fixed Teacher Effects model revealed that approximately 60% of TFA teachers performed better than the median performance for all teachers and followed a narrower distribution, suggesting that TFA teachers tended to vary less in their teaching effectiveness. However, the results for middle school students were less conclusive, likely due to the aforementioned difficulties with obtaining data. For middle school reading, the effects of TFA were statistically significantly positive when using the TFA dummy variable but not when using the TFA “intensity” variable, and this was true for both new teachers and all teachers as comparison groups. A sub-analysis of students with only one English teacher was not performed, as only about 12 percent of students fell under this group. Finally, quite surprisingly, with regard to middle school math, it was found that TFA teachers far outperformed the comparison group with all teachers by 11 percent of a standard deviation, but appeared to be only on par with new teachers. Though this defies the commonly accepted notion that novice teachers are not as effective as veteran teachers, it was found to be true for both the General model and the Fixed Teacher Effects model.

The primary difficulty with this study was the fact that 88% of middle school students had multiple teachers, often a mix of TFA and non-TFA teachers, for English over the course of

a school year. Although the authors attempted to account for this via the creation of a proportion-of-teaching variable, they faced a tradeoff on another front, as this meant that class-level variables could not be included. As expected, the effects of TFA were much more difficult to discern. Furthermore, it is unclear whether there are systematic differences between students who only had one teacher versus those who had multiple teachers, calling into question the validity of the positive effects obtained when the TFA dummy variable was employed. Finally, on a more fundamental level, the complexity of multiple teachers contributing to a single student's achievement growth is something that cannot be readily captured. However, the authors point out that all of the coefficients for TFA teaching were positive, even if some were not statistically significant; there were no instances in which TFA teachers performed worse than their peers, affirming the view that TFA is a viable source of teachers for struggling school districts.

Decker, Mayer & Glazerman (2004) conducted the first evaluation of TFA on the national level. Its geographic scope and unique experimental approach make it the most robust study undertaken thus far on the effects of TFA. A total of six regions were included in the study; a pilot study was launched in Baltimore for 2001-2002, and the remaining five regions (Chicago, Los Angeles, Houston, New Orleans, and the Mississippi Delta) were evaluated for the 2002-2003 school year. Seventeen schools, 100 classrooms, and approximately 2,000 students were included in the study; in general, this is a much smaller sample size than the non-experimental studies discussed in this paper. However, a major distinguishing factor of this study is that unlike the approach taken by Raymond et al. (2001), "control" teachers were restricted to those who taught in the same school and the same grade as the TFA teachers who were assessed, thus creating a much more accurate counterfactual for the absence of TFA. Furthermore, for schools in each region participating in TFA, students in grades 1 through 5 were randomly assigned to

classrooms prior to the start of the school year. These measures ensured that TFA teachers and non-TFA control groups had statistically comparable pools of students. In line with the previous study, TFA teachers were compared to control groups that included novice and veteran teachers, as well as control groups that only consisted of novice teachers (defined as those within their first three years of teaching during the years of the study). However, this study further controls for the effects of experience by comparing novice TFA teachers to novice non-TFA teachers. As a result of random assignment, no statistically significant treatment-control differences were found.

Data was obtained across 100 classrooms in 37 “comparison blocks”, or groups of TFA and non-TFA teachers within a given grade in a particular school. Student outcomes were measured using the Iowa Test of Basic Skills (ITBS), which was administered in the fall and the spring for each class included in the study. Overall, average student test scores for math were found to be significantly higher in TFA classrooms. Students with TFA teachers moved from the 14th to the 17th percentile, while students with non-TFA teachers remained at the 15th percentile for both administrations of the test. This roughly translates into an additional month of instruction, which suggests that the TFA effect is also meaningful in real terms. Average scores for reading were comparable between TFA and non-TFA classrooms; both groups improved by one percentile point. These results were robust across a variety of test and assumptions, as well as across subgroups of students and teachers. With regard to the effect of experience, it was found that the impact of TFA increased when compared only to novice control teachers, and decreased when the sample of TFA teachers was restricted to those in their first year of teaching. However, it was found that such impacts were sensitive to model specification. Due to the limited sample size of only 25 classrooms across 11 comparison blocks, no clear conclusions could be drawn.

In addition to its national scope and random assignment, a key strength of this study is its attempt to explicitly assess student outcomes in terms of what would have happened in the absence of TFA, as well as its discussion of a variety of factors not previously considered, such as the possibility of nonrandom attrition among students. Given the systematic differences in demographic profiles between TFA and non-TFA teachers, the study's assessment of differential impacts across student subgroups is also well-taken, despite discovering a general lack of differences. However, there was a fair amount of variation among the regions studied, which the authors readily attributed to imprecision as a result of the reduced sample size when considering each region individually. This is not altogether convincing, given the lack of baseline scores for each region in the study, and future research should strive to more rigorously assess potential systematic differences in TFA effectiveness across various geographic regions. Finally, it is noteworthy that this study takes a clear stance in the debate over whether TFA teachers should be assessed only in comparison to teachers in the same school or be subjected to more national standards, to which later studies would respond.

Most recently, Xu, Hannaway & Taylor (2008) conducted the first evaluation of TFA at the high school level; all previous studies had focused on either elementary or middle school students. Using data from North Carolina over 2000-2006, the study estimates the effect of TFA teachers compared to traditionally certified teachers on student achievement, which was measured using End-of-Course (EOC) test outcomes in 8 core subjects across 23 Local Education Agencies (LEAs) that hired at least one TFA teacher at any point over this time period. Annual data sets of student and teacher demographic information and student performance were merged into longitudinal files at the student, teacher, and classroom level. Student-level data contained information on ethnicity, gender, grade level, district and school identifiers, and score

achievement levels for EOC exams taken in a given year. Teacher-level data contained education, experience, licensure and certification, and salary. Class-level data contained subject, semester, grade level, ethnicity and gender composition of students, and teacher experience. Due to inconsistencies and omissions in the data, only about 84% of students in the initial sample were successfully matched to a teacher, either via a unique teacher ID or class demographic variables. To ensure that estimated TFA effects are robust, the authors performed an analysis on a second sample consisting of teachers who are matched to classes through their teacher ID, thus providing a higher level of accuracy in the construction of class-level data.

As Raymond et al. (2001) discovered, there were systematic differences between classes assigned to TFA teachers and those assigned to non-TFA teachers. Thus, the authors opted to use a student fixed-effects model as a means to counter any potential non-random sorting that occurs among students across schools, as well as across different teachers within schools. Past evidence suggests that there is a correspondence between observed teacher qualifications and student achievement, possibly as a result of factors such as parental pressure that cause a particular type of student to be assigned to certain teachers. Veteran teachers, in some instances, are also more likely to be assigned “easier” classrooms due to seniority. If TFA teachers are systematically given more or less challenging classrooms than their peers, the OLS estimates of TFA effects will naturally be biased. By looking at within-student variation over time, a student fixed-effect model removes any unobserved characteristics that may exist, so long as they are assumed to be constant over time. Although the availability of longitudinal data made the use of this model possible, many modifications had to be made. One modification is due to the particular structure of the typical high school curricula in schools included in the study, which does not allow for the data to provide repeated measures of student achievement in the same subject area over time.

Thus, the fixed-effects model looked at within-student variation across subjects, rather than temporally. That is, the difference between overall performance and performance in individual subject areas is attributed to the effects of teaching. Finally, data regarding initial student performance was not available, and so the model could not control explicitly for prior performance or construct a gain score. Rather, it assumed that prior academic preparation has a negligible effect on EOC scores.

Xu et al. argue that these modifications are reasonable upon an examination of the data. It was found that relative performance levels were significantly lower among classrooms taught by TFA teachers, suggesting that TFA teachers are typically assigned to more challenging classes. Students of TFA teachers were also more likely to be ethnic minorities and less likely to have parents with Bachelor's degrees or higher. Furthermore, an analysis of the data revealed that student performance tended to vary little across subjects; a high-performing math student was more likely to do well in other subjects as well, suggesting that the effects of any non-random sorting among students will be constant across subject areas. According to the authors, this justifies the use of the fixed-effects model with the interpretation of within-student variation across subject areas as the effects of teaching. Under this model, the effect of TFA was found to be positive and statistically significant compared to that of teachers across a variety of experience levels, and coefficients were robust across a variety of model specifications and data restrictions. In models where classroom characteristics were not controlled for, the effect of having a TFA teacher was roughly equivalent to the effect of having a teacher with 3 or more years of experience instead of a novice teacher; this almost triples in models that include classroom characteristics. In line with past research, the effect of teaching experience on student performance diminishes after the first three years of teaching, leading the authors to conclude

that TFA teachers more than compensate for their relative lack of teaching experience. However, the authors note that the results of this study are not indicative of the value of teacher training, as the effects of TFA could very well be attributed to its recruitment and selection process.

Despite the study's consistency across a range of assumptions and specifications, its methodology is far from perfect. Mainly, the comparison of individual subject area performance to overall performance as a means of evaluating teacher impact is flawed; an improvement in one subject area will also increase the overall level of performance, thus reducing the model's ability to accurately capture gains in achievement. However, it would seem that this would only cause a downward bias on the estimation of teacher impact, and there is no reason to believe that either TFA or non-TFA teachers would be systematically affected by it. It would appear, then, that the results hold.

The Case against TFA:

Three studies in this literature survey conclude that TFA teachers are less effective than traditionally certified teachers, which they attribute to a combination of a lack of specialized education, training, and experience. The first of these studies was conducted by Laczko-Kerr and Berliner in 2002. The study compared the academic achievement of primary school students who were taught by "under-certified" teachers and those taught by regularly certified teachers. Under-certified teachers were defined as those who held emergency, provisional, or temporary teaching certificates, and included TFA corps members, while regularly certified teachers were defined as those who fully met state certification requirements. Academic achievement was measured by mathematics, reading, and language scores on the Stanford Achievement Test 9 (SAT 9), a standardized achievement test commonly utilized in the United States to assess student knowledge in K-12 settings.

To obtain data for the study, Laczko-Kerr and Berliner solicited voluntary participation from school districts in Arizona participating in the Federal Teacher Shortage Loan Deferment Program, a program in which individuals teach in districts designated as teacher shortage areas in exchange for a deferral in payment on student loans. Five school districts agreed to participate in the study; in addition to facing teacher shortages, all five districts served inner-city populations, had large percentages of minority students, and participated in TFA. Each district provided a list of newly hired teachers for 1998-1999 and 1999-2000. From the list of 293 teachers who were eligible to be included in the study, Laczko-Kerr and Berliner matched under-certified teachers with regularly certified teachers to create 109 pairs, based on grade level taught and highest education level obtained. An effort was made to match teachers who taught in the same school, but quite understandably, a portion of the pairs had to be constructed across different schools in the same district or different districts. Teachers for whom no match could be found were removed from the sample. An analysis of test scores among districts revealed systematic and significant differences, suggesting that the matching of teachers from two different districts was detrimental to the study's ability to accurately assess the effect of TFA. Furthermore, the authors assumed, based on conversations with district personnel, that there were no systematic differences in student ability between those taught by certified and under-certified teachers within a given district. Given that previously discussed studies have pointed to the occurrence of non-random sorting of students into classes taught by novice or veteran teachers and that TFA teachers are consistently novice teachers, this assumption may not hold true.

Once pairs were constructed, SAT 9 scores were aggregated on the class level for both individuals within a given pair and compared using a one-way analysis of variance (ANOVA) with certification as the independent variable and student achievement score as the dependent

variable. No significant differences were found among those who held emergency, temporary, or provisional certifications. However, the students of regularly certified teachers outperformed those of under-certified teachers significantly in 2 out of the 3 subject areas in 1998 and all 3 subject areas in 1999. According to the authors, this difference translated roughly to a loss of 2 months' worth of academic growth for students placed with under-certified teachers. However, when the pairs were restricted to those containing TFA teachers, a statistically significant difference in achievement was only found for data collected in 1999. When the pairs were restricted to those with data over the entirety of 1998-2000, no statistically significant differences in achievement were found. A sub-analysis conducted to specifically assess the effect of experience yielded no statistically significant differences from one year to the next. Thus, the study was able to draw some conclusions on the effect of certification on student performance, but not on the effect of experience.

The major shortcomings of the data used in this study are the small and geographically concentrated sample and short time span, while its analytical methods can be characterized as also somewhat lacking in rigor. With particular attention to TFA teachers among those who are considered under-certified, there were only 8 pairs identified for 1998 and 22 pairs identified for 1999. Among all 109 pairs, only 6 had data across multiple years, yielding no readily apparent conclusions on the isolated effect of teaching experience on teacher effectiveness. While the authors' conclusion is somewhat bolstered by the statistical significance of results in 1999, when there were more pairs identified, the sample size is still hardly sufficient to allow the study to make generalizations about TFA as a national program. Furthermore, the analysis failed to control for a variety of factors that may systematically bias results. Mainly, the methods employed to construct pairs were crude and did not necessarily result in comparable individuals

being matched together, which was likely the case in particular when individuals across different school districts were matched together. Furthermore, this study does not control for prior achievement or rigorously assess the possibility of systematic biases in the assignment of classes to various teachers based on either certification or experience. Later studies control for a more comprehensive set of teacher characteristics than simply highest degree attained and grade level taught, utilize individual student level data rather than class level data, and either control for previous achievement or examine relative student achievement *gains* rather than relative student achievement levels.

Pilcher and Steele (2005) compared both the relative efficacy and effectiveness of TFA teachers in low-income elementary schools in Atlanta, Georgia through the construction of three groups, which consisted of 19 first- and 18 second-year TFA teachers (henceforth referred to as TFA1 and TFA2) and 21 regularly certified first-year teachers. Effectiveness was measured through a one-time administration of classroom achievement tests in mathematics, reading, and English/language arts for all classes taught by the three groups, while efficacy was measured through the administration of a revised version of the Hoy and Woolfolk Teacher Efficacy Scale, which assessed teachers' perceptions of general teacher efficacy (GTE) and personal teacher efficacy (PTE). GTE seeks to assess a teacher's belief about the relative influence of external factors versus that of teachers, while PTE seeks to assess the teacher's confidence in his or her own ability to overcome obstacles in the classroom. It was found that regularly certified and fully trained teachers generally scored higher on personal efficacy and general efficacy than TFA teachers, while TFA1 and TFA2 teachers did not differ from each other significantly. In terms of effectiveness, it was found that regularly certified teachers had a greater proportion of successful (at or above grade level) students as well as higher achievement scores than TFA1 teachers, but

there were not always statistically significant differences when compared with TFA2 teachers. Across the three subject areas tested, the proportion of successful students differed by more than 10% between TFA1 and TFA2 teachers. Based on this information, the authors conclude that both training and experience influence teacher effectiveness.

This study is significantly flawed and cannot be generalized to any great extent for a number of reasons. Firstly, the sample size of teachers is quite small, and they are confined to an even smaller number of schools. Furthermore, this study fails to rigorously control for school characteristics, classroom demographics, teacher demographics, grade levels, or prior student achievement, as both previous and subsequent studies have done. It has been shown that TFA teachers are consistently assigned to classes which score lower than that of their peers (Xu, Hannaway & Taylor, 2008), thus rendering a simple comparison of relative student performance inadequate for assessing TFA's impact. Finally, this study overstates the importance of teacher efficacy as an accurate measure of the relative merit of TFA teachers, as responses to statements such as "The amount a student can learn is primarily related to family background" may depend greatly on the teachers' formal education training, but such differences would not automatically correspond to similarly disparate differences in teaching ability. A survey of the literature revealed no recent studies assessing the impact of perceptions of efficacy on student achievement, though past studies have reported a significant relationship between teacher efficacy and achievement (Ashton, 1984).

Darling-Hammond, Holtzman, Gatlin and Heilig (2005) examine data from HISD to re-evaluate the findings of Raymond, Fletcher & Luque (2001) in an effort to address the fact that Raymond et al. compared TFA teachers to what they call an "extraordinarily underqualified pool of teachers," and failed to control for teacher certification status and degree levels. This study

also expands to include a greater number of years, from 1995-2002. Finally, due to the CREDO study's difficulties with assessing middle school students who had multiple teachers in one subject area, Darling-Hammond et al. limited their data set to students in grades 3 through 5. Student achievement was assessed using three separate standardized tests, the Texas Assessment of Academic Skills (TAAS), the Stanford Achievement Test 9 (SAT 9), and the Aprenda, a Spanish-language test. Each test consisted of two subject areas, reading and math, for a total of six outcome areas.

In contrast to the CREDO study, this study divided a sample of 1558 teachers into six categories depending on their certification type, which consisted of standard, alternative, emergency/temporary, certified out-of-field, certified no-test, uncertified, and certification unknown. There were TFA teachers in the uncertified, alternative, and standard certification categories. Thus, the study sought to simultaneously assess both the impact of certification and TFA status. Using OLS regression estimates of the predictors of six student outcomes (two subject areas for each of the three tests) at the individual student level, Darling-Hammond et al. were only able to replicate the results of the CREDO study with regard to TAAS scores; TFA teachers had a significant positive effect on math achievement, while they had no significant effect on reading. However, the authors remarked that the TFA effect became non-significant starting in 1999-2000, when TFA teachers were "noticeably" less likely to be certified compared to non-TFA teachers. TFA teachers were found to have a statistically significantly negative effect on math achievement on the SAT 9, and they had a significantly negative effect on both reading and math achievement on the Aprenda. With regard to certification, it was found that teachers with standard certification outperformed those with other types of certification or who lack certification altogether in 22 of the 36 estimates performed. When TFA and certification

were combined, it was found that TFA teachers with standard certification were statistically comparable to other teachers with standard certification. The authors conclude that the effects of TFA teachers depend largely on their certification status, which is a direct reflection of their level of preparation.

However, this study does not simply improve upon methods undertaken in the CREDO study. One caveat to this study is that by assessing both certification status and TFA status, the sample size of teachers within each category becomes quite small. Although no actual breakdown in the numbers of teachers in each category is provided, the authors acknowledge that the share of TFA teachers in the sample was small, and the number of TFA teachers with alternative certification was “very tiny.” The Aprenda and the SAT 9 were also only administered for 3 of the 6 years included in the study. Thus, the authors were unable to provide an explanation for why student achievement results might differ between the TAAS and the SAT9, since it could be the result of a combination of inherent differences in the tests themselves and fluctuations in the strengths of TFA teaching cohorts. Finally, the authors chose to simply omit middle school students from the study rather than address the difficulties faced by the CREDO study in obtaining accurate data. Thus, while this study legitimately addresses some of the shortcomings of the CREDO study, such as the omission of teacher certification in the model, it does not merit a convincing case against the findings of Raymond et al.

IV. Conclusions:

Implications:

As shown above, the studies that have been conducted thus far have been generally limited in scope, either geographically, temporally, or in terms of grade level. To date, there has only been one national and randomized study on the effects of TFA, and even so, it included

only 6 of the 29 regions in which TFA currently operates. Furthermore, efforts to study both experience and certification simultaneously have been plagued by insufficient sample sizes. Thus, it is difficult to confidently generalize the results of any one study. However, it is clear that the best study that has been conducted to date is the one conducted by Decker et al. (2004), while the study that uses the most recent data is the one conducted by Xu et al. (2008). Given TFA's dynamic growth and presumed refinement in its recruitment and training practices, the relative recency of the latter's data lends the study significant credibility in its ability to accurately assess TFA's current impact as an organization, which is the most important piece of information for all stakeholders involved. Both of these studies yielded strong positive results regarding the effects of TFA, tipping the debate over the merits of TFA in the organization's favor. As Xu et al. point out, school districts pay far less to hire a TFA teacher than through traditional recruitment methods, making participation in TFA a low-risk venture, particularly if TFA is found to have a consistently positive or at least non-negative impact across various grades and subject areas. Nevertheless, while TFA may be a beneficial program from a school district's standpoint, it is perhaps also important to consider the overall social cost of TFA, as corps members often forgo far more lucrative professions during the time they spend teaching.

TFA in the Context of Current Debate:

One reason for the difference in conclusion between at least two of the studies can be attributed to the construction of the comparison group. While Darling-Hammond et al. (2005) compare TFA teachers to all regularly certified teachers, including those who teach in school districts that do not participate in TFA and thus may have significantly better socioeconomic conditions, Xu et al. (2008) and Decker et al. (2004) restrict data to only schools that have at least one TFA teacher. In such schools, the comparison group tended to have a much higher

percentage of under-certified teachers or those who lack formal education training.

Unsurprisingly, this approach yielded much more positive results with regard to TFA effectiveness. This introduces an interesting philosophical question: to which standard should TFA hold its teachers? While some argue that the latter constructs a much more accurate counterfactual of conditions in the absence of TFA, others say that comparing TFA teachers to other disproportionately under-certified teachers fails to address the fundamental issues that lie at the core of educational inequality. This reflects the greater policy context in which TFA operates: should educational reform introduce more flexibility to attract a broader base of candidates to address the shortage in teacher supply, or raise educational and certification requirements to place a higher caliber of individuals in public schools?

If future research gravitates toward the former ideological stance and thus adopts the practice of only comparing teachers who teach in the same school or district, analytical steps might be taken to construct a more realistic counterfactual of what students would experience in the absence of TFA. Given the typical financial limitations and chronic teacher shortages of school districts in which TFA operates, it may be overly optimistic to assume that the counterfactual simply differs in that it entails instruction by a non-TFA teacher, either regularly or alternatively certified. Rather, it is more reasonable to assume that a school district will not hire as many new teachers in the absence of TFA, causing the average class size to increase. An analytical model that accounts for this change in class size might more accurately reflect TFA's impact, as research has shown that a reduction in class size significantly increases achievement (Decker, Mayer & Glazerman, 2004) and one can reasonably hypothesize that an increase in class size will do just the opposite. Future research in this direction will connect the debate over

TFA and the debate over whether the effects of class size are significant enough to merit action on the part of education policymakers.

Directions for Future Research:

Future research should generally seek to be more comprehensive in scope, whether geographically, temporally, or across grade levels. Geographically, given TFA's non-random placement of its teachers, it is quite possible that selection biases exist among various regions. One can see how the number of teachers who want to teach in New York City might be significantly greater than that for rural South Dakota, which might cause systematic differences between TFA teachers in the two regions. This idea can be similarly applied to the assignment of teachers to grade levels or subject areas. Thus, an assessment of the supply and demand for TFA teachers across these various criteria should be undertaken. With regard to temporal changes, it is unclear whether the current body of literature reflects a refinement in TFA's recruiting and training methods over time, or business cycles that may affect employment choices among recent college graduates. If discernible changes in the applicant pool have taken place, they would shed light on the relative value of various studies based on the time periods covered by each; thus, longitudinal studies should be undertaken to assess the change in the impact of TFA corps members over time in light of various external factors.

Overall, TFA's rapid expansion can be viewed as an opportunity for more rigorous and multifaceted research into particular aspects of the program. It has resulted in a significantly greater number of students taught by TFA teachers, more TFA teachers per school in many cases, and lower attrition rates (currently only 9% between the first and second year, compared to the aforementioned 58% in 1990). This has led to a greater number of relatively more experienced TFA teachers teaching at any given time, and there is now a significant contingent of teachers

who opt to stay beyond their 2-year commitment. It has been estimated that about a third of alumni choose to continue teaching; thus, one possible research area may be to assess the relative effectiveness of veteran TFA teachers versus traditionally certified veteran teachers, in order to shed light on the effects of certification on more experienced teachers, as well as possible differences in inherent teaching ability.

Despite its history of controversy, TFA has become one of the most prominent service organizations in the nation. Just as its founder Wendy Kopp had envisioned, TFA alumni are continuing their participation in shaping the state of public education by starting alternative education programs, running school districts, and influencing policymakers. Given the organization's increasing entrenchment in our nation's education system, it is more than likely that TFA will continue its work of supplying schools around the country with a corps of recent college graduates each year to serve as teachers. Until now, research has only focused on evaluating the merits of TFA as a whole, but from now on, focus on particular aspects of TFA will be more productive. Given the increasing amount of available data in recent years, future research has the potential to accurately assess the impact of specific factors such as recruiting, training, and experience on teaching effectiveness, with potential implications for meaningful changes in TFA's programmatic structure. Indeed, in light of TFA's growing prevalence, discovering ways to refine and improve upon TFA's existing practices is perhaps one of the best approaches for researchers and policymakers alike to work toward the goal of addressing "our nation's greatest injustice."

References

- Ashton, P. (1984). *Teacher Efficacy: A Motivational Paradigm for Effective Teacher Education*. *Journal of Teacher Education*, Vol. 35, p. 28-32.
- Darling-Hammond, L. (1994). *Who Will Speak for the Children? How 'Teach for America' Hurts Urban Schools and Students*. *Phi Delta Kappan*, Vol. 76, p. 21-34.
- Darling-Hammond, L., Holtzman, D., Gatlin, S., & Heilig, J. (2005). *Does Teacher Preparation Matter? Evidence about Teacher Certification, Teach for America, and Teacher Effectiveness*. Stanford University School of Education. Retrieved March 10, 2009 from <http://www.ncate.org/documents/EdNews/StanfordTeacherCertificationReport.pdf>.
- Decker, P., Mayer, D., & Glazerman, S. (2004). *The effects of Teach for America on Students: Findings from a national evaluation*. Mathematica Policy Research. Retrieved February 1, 2009 from <http://www.mathematica-mpr.com/publications/pdfs/teach.pdf>
- 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 February 8, 2009 from <http://epaa.asu.edu/epaa/v10n37/>.
- Pilcher, Lorene C., & Steele, Donald C. (2005). *Teach for America and regularly certified teachers: teacher efficacy, teaching concerns, career aspirations, and teaching effectiveness*. *Research on Alternative and Non-traditional Education* (Lanham, Maryland: The Association of Teacher Educators), p. 123-142.
- Raymond, M., Fletcher, S. H., & Luque, J. (2001). *Teach for America: An evaluation of teacher differences and student outcomes in Houston, Texas*. Stanford, CA: The Hoover Institution, Center for Research on Education Outcomes. Retrieved February 1, 2009 from <http://credo.stanford.edu/downloads/tfa.pdf>.
- Xu, Z., Hannaway, J., & Taylor, C. (2008). *Making a difference? The effects of Teach for America in high school*. The Urban Institute. Retrieved February 1, 2009 from <http://www.urban.org/publications/411642.html>.