The Impact of Voluntary Youth Service on Future Outcomes: Evidence from Teach For America*

Will Dobbie
Princeton University and NBER

Roland G. Fryer, Jr.
Harvard University and NBER

March 2015

Abstract

This paper provides causal estimates of the impact of service programs on those who serve, using data from a web-based survey of former Teach For America applicants. We estimate the effect of voluntary youth service using a discontinuity in the Teach For America application process. Participating in Teach For America increases racial tolerance, makes individuals more optimistic about the life prospects of poor children, and makes them more likely to work in education.

"We need your service, right now, at this moment in history.... I'm asking you to help change history's course. Put your shoulder up against the wheel. And if you do, I promise you - your life will be richer, our country will be stronger, and someday, years from now, you may remember it as the moment when your own story and the American story converged, when they came together, and we met the challenges of our new century."

President Barack Obama, at the signing of the Edward M. Kennedy Serve America Act

Over the past half century, nearly one million American youth have participated in national service programs such as the Peace Corps, AmeriCorps, and Teach For America (TFA).[1] These organizations have two stated objectives. The first is to provide services to communities in need. Peace Corps sends volunteers to work in education, business, information technology, agriculture, and the environment in more than 70 countries. Volunteers in Service to America (VISTA), an AmeriCorps program, enlists members to serve for a year at local nonprofit organizations or local government agencies. Teach For America recruits recent, accomplished college graduates to teach in some of the most challenging public schools.

There is emerging empirical evidence that service organizations benefit the individuals that they serve. Decker et al. (2006) find that students randomly assigned to classrooms with Teach For America corps members score 0.04 standard deviations higher in reading and 0.15 standard deviations higher in math compared to students in classrooms with traditional teachers. Moss et al. (2001) find that students enrolled in an AmeriCorps tutoring program experience larger than expected gains in reading performance.

The second objective of these service organizations is to influence the values and future careers of those who serve. The Peace Corps' stated mission includes helping "promote a better understanding of other peoples on the part of Americans." VISTA hopes to encourage its members to fight poverty throughout their lifetimes. Teach For America aims to develop a corps of alumni dedicated to ending educational inequity even after their two-year commitment is over. Advocates of service organizations point to notable alumni such as Christopher Dodd (Peace Corps), Reed Hastings (Peace Corps), and Michelle Rhee (Teach For America), as evidence of the long term impact on individuals who serve.

Despite nearly a million service program alumni and annual government support of hundreds of millions of dollars, there is no credible evidence of the causal impact of service on

---

[1]This includes approximately 200,000 Peace Corps volunteers, 637,000 AmeriCorps Volunteers, and 28,000 Teach For America corps members.

those who serve.[2] This is due, in part, to the fact that service alumni likely had different values and career goals even before serving. As a result, simple comparisons of service program alumni and non-alumni are likely to be biased.

To our knowledge, this paper provides the first causal estimate of the impact of service programs on those who serve, using data from a web-based survey of TFA applicants – both those that were accepted and those that were not – administered for the purposes of this study.[3] The survey includes questions about an applicant's background, educational beliefs, employment, political beliefs, and racial tolerance. The section on educational beliefs asks about the extent to which individuals feel that the achievement gap is solvable, and the importance of teachers in reaching that goal. Employment variables measure whether individuals are interested in working in education in the future, are currently employed in education, and prefer to work in an urban school. Political beliefs is captured through a series of questions such as whether the respondent self-identifies as liberal, and whether America should spend more money on specific social policies. Racial tolerance is captured using an Implicit Association Test. For a complete list of questions, see Online Appendix C.

Our identification strategy exploits the fact that admission into TFA is a discontinuous function of an applicant's predicted effectiveness as a teacher, calculated using a weighted average of scored responses to interview questions. As a result, there exists a cutoff point around which very similar applicants receive different application decisions. The crux of our identification strategy is to compare the average outcomes of individuals just above and below this cutoff. Intuitively, we attribute any discontinuous relation between average outcomes and the interview score at the cutoff to the causal impact of service in TFA.

One threat to our approach is that interviewers may manipulate scores around the cutoff point. However, the cutoff score is not known to the interviewers or applicants at the time of the interview. Individuals are scored months before TFA knows how many slots they will have for that year. Therefore, it seems unlikely that interviewers could accurately manipulate scores around the cutoff point and the density of scores should be smooth at the cutoff. Indeed, a McCrary (2008) test – which, intuitively, is based on an estimator for the discontinuity at the cutoff in the density function of the interview score – fails to reject that

---

[2]The 2010 federal budget included $373 million for Peace Corps and $98 million for AmeriCorps VISTA. Decker et al. (2006) report that school districts typically contribute about $1,500 per TFA member to offset recruiting costs, or about $8 million per year in total.

[3]There are several studies examining the correlation between service and later outcomes. McAdam and Brandt (2009) compare 1993-1998 TFA alumni to TFA applicants who were admitted but chose not to serve. Haan (1974) surveyed 220 Peace Corps members bound for service, comparing those admitted to those that returned. Yamaguchi et al. (2008) surveyed individuals who "expressed interest" in AmeriCorps to individuals who applied to estimate the impact of AmeriCorps on civic engagement, employment, and educational attainment. All of these studies suffer from the same issues of self-selection and thus do not provide credible causal impacts of the effects of service programs on future outcomes of those that serve.

the density of scores is the same immediately above and below the cutoff point (p-value = 0.551).

Another threat to the causal interpretation of our estimates is that applicants may selectively respond to our survey. In particular, one may be concerned that TFA alumni will be more likely to respond, or that the non-alumni who respond will be different in some important way. Such selective response could invalidate our empirical design by creating discontinuous differences in respondent characteristics around the score cutoff. We evaluate this possibility in two ways. First, we test whether the survey response rate changes at the admissions cutoff to see if TFA alumni are more likely to respond to our survey. Second, we test whether the observable characteristics of survey respondents trend smoothly through the admissions cutoff score to see if alumni and non-alumni respondents are similar. In both cases, we find no evidence of the type of selective survey response that would invalidate our research design.

Our empirical analysis finds that serving in Teach For America increases an individual's faith in education, involvement in education, and racial tolerance. One year after finishing their TFA service, TFA alumni are 35.5 percentage points more likely to believe that the "achievement gap is a solvable problem" and 38.2 percentage points more likely to believe that teachers are the most essential determinant of a student's success. TFA alumni are also 36.5 percentage points more likely to work for a K-12 school and 43.3 percentage points more likely to work in an education related career one year after their service ends. Finally, serving in TFA increases implicit black-white tolerance, as measured by an Implicit Association Test (IAT), by 0.8 standard deviations. TFA service is also associated with statistically insignificant increases in explicit black-white tolerance and implicit white-Hispanic tolerance.

These effects are quite large. For instance, TFA service leads to Implicit Association Test scores jumping from the 63rd percentile to the 87th percentile in the distribution of over 700,000 black-white IAT scores collected by Project Implicit in 2010 and 2011.[4] In a 2010 Gallup poll, 22 percent of a nationally representative sample of individuals aged 18 and older reported that school was the most important factor in determining whether students learn. Thirty-eight percent of respondents below the cutoff point in our sample said that teachers were the most important in determining how well students perform in school. The impact of TFA service on this belief was 38 percentage points.

Subsample results reveal that the impact of TFA service on educational beliefs is larger for white and Asian applicants and non-Pell Grant recipients, while the impact of TFA on

---

[4]Typical IAT scores have higher measures associated with a more anti-black response. However, we multiply our IAT measure by negative one so that higher values indicate less anti-black bias. We also multiply the scores collected by Project Implicit by negative one so that the distributions are comparable.

educational involvement is larger for men, white, and Asian applicants. However, some of these differences are statistically insignificant after correcting for multiple hypothesis testing.

TFA service typically involves sending a college educated young adult, whose parental income is above the national average, into a predominantly poor and minority neighborhood to teach. Eighty percent of corps members in our survey sample are white, and eighty percent have at least one parent with a college degree. The average parental income of a corps member while in high school is $118 thousand, compared to the national median family income of approximately $50 thousand (U.S. Census Bureau 2000). In sharp contrast to this privileged upbringing, roughly 80 percent of the students taught by corps members qualify for free or reduced-price lunch and more than ninety percent are African-American or Hispanic. To the best of our knowledge, our analysis is the first to estimate the impact of contact with poor and minority groups on the beliefs of more advantaged individuals.

There are five potentially important caveats to our analysis. First, because TFA introduced its discontinuous method of selecting applicants in 2007, our primary analysis includes only one cohort of TFA applicants surveyed roughly a year after their service commitment ended. To address this issue, we also collected data on TFA applicants from the 2003 to 2006 cohorts. Applicants in these cohorts were admitted only if they met prespecified interview subscore requirements. For example, TFA admitted applicants with the highest possible interview score in perseverance and organizational ability as long as they had minimally acceptable scores in all other areas. In total, there were six separate combinations of minimum interview subscores that met the admissions requirements. We estimate the impact of service for the 2003 to 2006 cohorts by instrumenting for TFA service using these admissions criteria. The impact of TFA service is therefore identified using the interaction of the subscores in these cohorts. Our key identifying assumption is that the interaction of interview subscores only impacts future outcomes through TFA service after controlling for the impact of each non-interacted subscore. These instrumental variable estimates suggest that the impacts of service are persistent, with older TFA alumni more likely to believe in the power of education, more likely to be employed in education, and more racially tolerant.

A second caveat is that the response rate of the 2007 cohort to our web-based survey is only 31.2 percent. While there is no evidence that alumni and non-alumni selectively responded to our survey, we cannot rule out unobserved differences among respondents. We note that low response rates are typical in web-based surveys. A web-based survey of University of Chicago Business School alumni conducted by Bertrand, Goldin and Katz (2010) had a response rate of approximately 26 percent, while a web-based survey of individuals receiving UI benefits in New Jersey survey conducted by Krueger and Mueller (2011) had a response rate of six to ten percent.

Third, TFA alumni and non-alumni may be differentially primed by the survey questions. For example, alumni may feel the need to answer in a way that reflects well on TFA, while non-alumni may feel the need to justify their non-participation. We note that measures based on objective outcomes (e.g. current employment) or implicit attitudes (e.g. IAT) are less likely to be influenced in this way.

Fourth, although TFA is broadly similar to other service organizations, it differs in important ways that limit our ability to generalize our results. To the extent that TFA's impact on alumni is driven by factors that all service organizations have in common, the results of our study will be informative about the effects of service programs more generally. If one believes that the unique attributes of TFA, such as its selectivity or focus on urban teaching, drive its impact, the results of our study should be interpreted more narrowly.

Fifth, there is no easy way to distinguish between the impacts of the TFA program and the impacts of becoming a teacher. Some of the effects that we detect could potentially just be things that people believe after becoming a regular teacher. However, it is important to note that the vast majority of TFA applicants would not get involved in teaching but for TFA. In that sense, our lack of ability to distinguish between the impacts of TFA and teaching does not prevent us from estimating the impact of TFA service even if that effect might be similar for first year teachers.

The paper proceeds as follows. Section I provides a brief overview of Teach For America and its relationship to other prominent service programs around the world. Section II describes our web-based TFA survey and sample. Section III details our research design and econometric framework for estimating the causal impact of TFA on racial and educational beliefs, employment outcomes, and political beliefs. Section IV describes our results. The final section concludes. There are three online appendices. Online Appendix A provides additional results and robustness checks of our main analysis. Online Appendix B provides further details of how we coded variables used in our analysis and constructed the samples. Online Appendix C provides implementation details and the complete survey administered to TFA applicants.

I. A Brief Overview of Teach For America

A. History

Teach For America, a non-profit organization that recruits recent college graduates to teach for two years in low-income communities, is one of the nation's most prominent service programs. Based on founder Wendy Kopp's undergraduate thesis at Princeton University, TFA's mission is to create a movement that will eliminate educational inequity by enlisting

our nation's most promising future leaders as teachers. In 1990, TFA's first year in operation, Kopp raised $2.5 million and attracted 2,500 applicants for 500 teaching slots in New York, North Carolina, Louisiana, Georgia, and Los Angeles.

Since its founding, TFA corps members have taught more than three million students. Today, there are 8,200 TFA corps members in 125 "high-need" districts across the country, including 13 of the 20 districts with the lowest graduation rates. Roughly 80 percent of the students reached by TFA qualify for free or reduced-price lunch and more than 90 percent are African-American or Hispanic.

B. Application Process

Entry into TFA is highly competitive; in 2010, more than 46,000 individuals applied for just over 4,000 spots. Twelve percent of all Ivy League seniors applied. A significant number of seniors from historically black colleges and universities also applied, including one in five at Spelman College and one in ten at Morehouse College. TFA reports that 28 percent of incoming corps members received Pell Grants, and almost one-third are people of color.

In its recruitment efforts, TFA focuses on individuals who possess strong academic records and leadership capabilities, regardless of whether or not they have had prior exposure to teaching. Despite this lack of formal teacher training, students assigned to TFA corps members score about 0.15 standard deviations higher in math and 0.04 standard deviations higher in reading than students assigned to traditionally certified teachers (Decker et al. 2006).

To apply, candidates complete an online application, which includes a letter of intent and a resume. After a phone interview, the most promising applicants are invited to participate in an in-person interview, which includes a sample teaching lesson, a group discussion, a written exercise, and a personal interview. Applicants who are invited to interview are also required to provide transcripts, obtain two on-line recommendations, and provide one additional reference.

Using information collected through the application and interview, TFA bases their candidate selection on a model that accounts for multiple criteria that they believe are linked to success in the classroom. These criteria include achievement, perseverance, critical thinking, organizational ability, motivational ability, respect for others, and commitment to the TFA mission. TFA conducts ongoing research on their selection criteria, focusing on the link between these criteria and observed single-year gains in student achievement in TFA classrooms.

As discussed above, between 2003 and 2006 TFA admitted candidates who met pre-specified interview subscore requirements. In 2007, TFA conducted a systematic review of their admissions measures to improve the correlation between interview scores and an inter-

nal TFA measures of classroom success. This review resulted in TFA calculating a single predicted effectiveness score using a weighted average of each interview subscore. These predicted effectiveness scores were meant to serve as a way to systematically rank candidates during the admissions process. TFA officials do not publicly reveal certain details about the model, such as the exact number of indicators, what they measure, or how they are weighted in constructing an overall score, in order to prevent "gaming" of the system by applicants.[5] Section III details how we use the predicted effectiveness scores to estimate the causal impact of service.

C. Training and Placement

TFA cohorts included in our study were required to take part in a five-week TFA summer institute to prepare them for placement in the classroom at the end of the summer. The TFA summer institute includes courses covering teaching practice, classroom management, diversity, learning theory, literacy development, and leadership. During the institute, groups of participants also take full teaching responsibility for a class of summer school students.

At the time of their interview, applicants submit their subject, grade, and location preferences. TFA works to balance these preferences with the needs and requirements of districts. With respect to location, applicants rank each TFA region as highly preferred, preferred, or less preferred and indicate any special considerations, such as the need to coordinate with a spouse. Over 90 percent of the TFA applicants accepted are matched to one of their "highly preferred" regions (Decker et al., 2006).

TFA also attempts to match applicants to their preferred grade levels and subjects, depending on applicants' academic backgrounds, district needs, and state and district certification requirements. As requirements vary by region, applicants may not be qualified to teach the same subjects and grade levels in all areas. It is also difficult for school regions to predict the exact openings they will have in the fall, and late changes in subject or grade-level assignments are not uncommon. Predicted effectiveness scores are not used to determine the placement region, grade, or school, and the scores are not available to districts.

TFA corps members are hired to teach in local school districts through alternative routes to certification. Typically, they must take and pass exams required by their districts before they begin teaching. Corps members may also be required to take additional courses to meet state certification requirements or to comply with the requirements for highly qualified teachers under the No Child Left Behind Act.

TFA corps members are employed and paid directly by the school districts for which they

---

[5]See Dobbie (2011) for additional details on the admissions process and correlation between each subscore and classroom effectiveness.

work, and generally receive the same salaries and health benefits as other first year teachers. Most districts pay a $1,500 per corps member fee to TFA to offset screening and recruiting costs. TFA gives corps members various additional financial benefits, including "education awards" of $4,725 for each year of service that can be used for past or future educational expenses, and transitional grants and no-interest loans to help corps members make it to their first paycheck.

II. Teach For America Survey and Sample

To understand the impact of TFA on racial and educational beliefs, employment outcomes, and political beliefs, we conducted a web-based survey of the 2003-2007 TFA application cohorts between April 2010 and May 2011[6]. The survey contained 87 questions and lasted approximately 30 minutes. As an incentive to complete the survey, every individual was entered into a lottery for a chance to win $5,000. The complete survey is available in Online Appendix C.

A. Contacting TFA Applicants

Applicants were first contacted using the email addresses they supplied to TFA in their initial applications. Between April 2010 and June 2010, applicants received up to three emails providing them with information about the survey and a link to the survey. Each email reminded applicants that by completing the survey they would be automatically entered in a lottery for $5,000. Thirty-nine percent of the 2,573 2007 TFA alumni and 14.1 percent of the 4,795 2007 non-alumni started the survey during this phase. To increase the response rate among 2007 non-alumni, we contacted individuals using phone numbers from TFA application records. We began by contacting all 2007 non-alumni who had not responded to the survey using an automated call system that included a brief 30 second recording with information about the survey. We then contacted the remaining non-respondents using personal calls from an outsourced calling service. Voicemails were left for those who did not answer the phone. Those who did not answer the phone were also called again a few weeks later. We used a similar outreach process for the 2003 - 2006 cohorts, though we made fewer follow-up calls than with the 2007 cohort, as the 2007 cohort was the priority for our analysis. Online Appendix C provides additional details on each step of this process.

These strategies yielded a final response rate of 39.8 percent among 2007 TFA alumni and

---

[6]We also collected data on the 2008 and 2009 TFA application cohorts for other research purposes. This data was not used in this analysis since these cohorts had not completed their TFA service at the time of the survey.

26.6 percent among 2007 non-alumni. The response rate is lower for older cohorts and non-alumni. The difference in the response rate between alumni and non-alumni is smallest in the 2007 cohort, likely due to the additional phone calls to non-alumni in this cohort. Response rates are presented for all cohorts in Appendix Figure 1. Section III examines differences in survey response around the TFA selection cutoff, finding no evidence of selective survey response.

One potential concern is that we recruited non-alumni using both email and phone call strategies, while we recruited alumni using email strategies only. If phone calls induce different individuals to respond to the survey, our results may be biased. Appendix Table 1 presents summary statistics for the 2007 application cohort separately by survey strategy. Non-alumni respondents from the email strategy are 3.0 percentage points more likely to be black, have college GPAs that are 0.039 points lower, and are 4.1 percentage points more likely to have a math or science major in college. There are no other statistically significant differences among the 12 background variables available. Non-alumni respondents from the email strategy are 6.4 percentage points less likely to believe that teachers are the most important determinant of student success and 5.7 percentage points less likely to believe that teachers can ensure most students achieve, but do not differ on the other 19 outcome measures collected. Appendix Table 2 further examines this issue by estimating results controlling for survey strategy. The results are nearly identical to our preferred specification.

B. The Survey

Data collected in our online survey of TFA applicants is at the heart of our analysis. We asked applicants about their demographics and background, educational beliefs, employment outcomes and aspirations, political beliefs, and racial beliefs. Whenever possible, survey questions were drawn from known instruments such as the College and Beyond Survey, the Harvard and Beyond Survey, The National Longitudinal Study of Adolescent Health Teacher Survey, the Modern Racism Scale, and the General Social Survey. In this paper, we use only a small fraction of the data we collected. For further details on these variables or those omitted from our analysis, see Online Appendix C.

The set of questions on educational beliefs was designed to measure the extent to which individuals feel that the achievement gap is solvable and that schools can achieve that goal, and the importance of teachers in increasing student achievement. Survey respondents were asked whether they agreed or disagreed with a series of statements on a five point Likert scale ranging from "agree strongly" to "disagree strongly." The questions used are similar to those asked in The National Longitudinal Study of Adolescent Health Teacher Survey. Other, more open-ended questions include "what fraction of blacks can we reasonably expect

to obtain a college degree," and "who is the most important in determining how well students perform in school?" For questions with answers that do not have clear cardinality, we create indicator variables equal to one if the response was "favorable" (e.g. strongly agree that the achievement gap is a solvable problem).

Employment variables measure whether individuals are interested in working in education in the future, whether they are currently employed in education, and whether they prefer to work in an urban or suburban school. Political beliefs are captured by a series of questions such as whether or not the respondent self identifies as liberal or whether the government should spend more or less on issues such as closing the achievement gap, welfare assistance, and fighting crime. For political beliefs, we create indicator variables equal to one if the response is more liberal.

In the final portion of the survey, we asked participants to take a ten minute Implicit Association Test that measured black-white implicit bias. Previous research suggests that the IAT is the best available measure of unconscious feelings about minorities (Bertrand, Chugh, Mullainathan 2005).[7] The IAT is more difficult to manipulate than other measures of racial bias (Steffens 2004), and a recent meta-analysis found that black-white IAT scores are better at predicting behaviors than explicit black-white attitudes (Greenwald et al. 2009). IAT scores also correlate well with other implicit measures of racial attitudes and real-world actions. For instance, individuals with more anti-black IAT scores are more likely to make negative judgments about ambiguous actions by blacks (Rudman and Lee 2002); more likely to exhibit a variety of micro-behaviors indicating discomfort with minorities, including less speaking time, less smiling, fewer extemporaneous social comments, more speech errors, and more speech hesitations in an interaction with a black experimenter (McConnell and Leibold 2001); and are more likely to show greater activation of the area of the brain associated with fear-driven responses to the presentation of unfamiliar black versus white faces (Phelps et al. 2000). IAT scores also predict discrimination in the hiring process among managers in Sweden (Rooth 2007) and certain medical treatments among black patients in the U.S. (Green et al. 2006), though the latter finding has been questioned (Dawson and Arkes 2008).

We use a brief format IAT, developed by Sriram and Greenwald (2009), to assess the relative strength of automatic associations between "good" and "bad" outcomes and white and black faces. The brief format IAT performs similarly on test-retest and implicit-explicit correlations as the standard format IAT, with the brief format version requiring only one third the number of trials. We standardize the IAT scores to have a mean of zero and a

---

[7]Some critics argue that the IAT may be assessing shared norms, familiarity, perceptual salience asymmetries, or cultural knowledge that does not correspond to personal endorsement of that knowledge (e.g. Karpinski and Hilton 2001; Rothermund and Wentura 2004).

standard deviation of one, with higher values indicating less anti-black bias.

To complement the IAT measure of implicit bias, individuals were also asked about explicit racial bias.[8] Our first measure of explicit bias comes from the General Social Survey. Individuals were asked to separately rate the intelligence of Asians, blacks, Hispanics, and whites on a seven point scale that ranged from "almost all are unintelligent" to "almost all are intelligent." We recoded this variable to indicate whether individuals believe that blacks and Hispanics are at least as intelligent as whites and Asians. Our second measure of explicit bias is the Modern Racism Scale (McConahay 1983). The Modern Racism Scale consists of six questions with which individuals are asked how much they agree or disagree. Each item was re-scaled so that lower numbers are associated with a more anti-black response, and then a simple average was taken of the six questions. We normalized this scale to have mean zero and standard deviation one across each cohort. The six statements that individuals were presented are: "over the past few years, blacks have gotten more economically than they deserve;" "over the past few years, the government and news media have shown more respect for blacks than they deserve;" "it is easy to understand the anger of black people in America;" "discrimination against blacks is no longer a problem in the United States;" "blacks are getting too demanding in their push for equal rights;" and "blacks should not push themselves where they are not wanted."

Index variables for each survey domain were also constructed by standardizing the sum of individual questions to have a mean of zero and a standard deviation of one in each cohort. Rather than add dichotomous and standardized variables together, we converted all standardized variables to indicator variables equal to one if the continuous version of the variable was above the median of the full sample. Results are qualitatively similar if we combine the original dichotomous and continuous variables. Details on the coding of each measure are available in Online Appendix B.

C. The Final Sample

Our final sample consists of data from our web-based survey merged to administrative data from Teach For America. The administrative records consist of admissions files and placement information for all TFA applicants who attended the in-person interview in the 2003 to 2007 application cohorts. A typical applicant's data include his or her name; undergraduate institution, GPA, and major; admissions decision; placement information; and interview score. We matched the TFA administrative records with our web-based survey using name, application year, college, and email address. Our primary sample consists of all 2007 appli-

---

[8]The IAT directly followed the questions on explicit racial bias, which could influence the IAT measure. However, any potential bias is the same for all survey respondents.

cants who responded to our survey. Our secondary sample consists of survey respondents from all cohorts.

Summary statistics for the 2007 survey cohort are displayed in Table 1. Eighty-one percent of TFA alumni are white, 6.1 percent are Asian, 6.3 percent are black, and 5.0 percent are Hispanic. Among non-alumni, 79.1 percent are white, 6.7 percent are Asian, 7.3 percent are black, and 5.0 percent are Hispanic.[9] TFA alumni have an average college GPA of 3.58 while non-alumni have an average GPA of 3.48. The parents of both the typical alumni and non-alumni are highly educated. Forty percent of alumni have a mother with more than a BA, and 46.7 percent have a father with more than a BA. Among non-alumni, 32.4 percent have a mother with more than a BA and 41.1 percent have a father with more than a BA. With that said, a significant fraction of TFA applicants come from disadvantaged backgrounds. Twenty percent of TFA alumni in our sample were eligible for a Pell Grant in college, while 22.0 percent of non-alumni were eligible.

Appendix Table 3 presents summary statistics for the 2007 survey cohort and the full sample of 2007 TFA applicants. The 2007 alumni survey sample is 3.6 percentage points more likely to be white, 1.0 percentage points more likely to be Asian, 3.6 percentage points less likely to be black, and 1.0 percentage points less likely to be Hispanic than the full sample of 2007 alumni. Conversely, the 2007 non-alumni survey sample is 5.7 percentage points more likely to be white, 3.8 percentage points less likely to be black, and 1.5 percentage points less likely to be Hispanic than the full sample of 2007 non-alumni. The alumni survey sample is also 2.2 percentage points less likely to have received a Pell Grant compared to the full sample of alumni, while the non-alumni survey sample is 3.7 percentage points less likely to have received a Pell Grant. Both the alumni and non-alumni survey samples also have lower college GPAs than the full sample.

III. Research Design

Our identification strategy exploits the fact that entry into TFA is a discontinuous function of an applicant's interview score. Consider the following conceptual model of the relationship between future outcomes (y_i) and serving in TFA (TFA_i):

$$y_i = \alpha + \gamma TFA_i + \varepsilon_i \tag{1}$$

---

[9]The racial distribution of TFA applicants mirrors that of colleges graduates from selective colleges more broadly. Five percent of graduating seniors at "more selective" or "most selective" colleges are black. Six percent are Hispanic (U.S. Department of Education 2010). The 2011 TFA cohort is more diverse than previous cohorts, with 12 percent of the cohort identifying as black and 8 percent identifying as Hispanic.

The parameter of interest is γ, which measures the causal effect of TFA on future outcomes y_i. The problem for inference is that if individuals select into service organizations because of important unobserved determinants of later outcomes, such estimates may be biased.[10] In particular, it is plausible that people who select into service organizations had different beliefs and outcomes before they served: $E[\varepsilon_i | y_i^{pre}] \neq 0$, where y_i^{pre} is a vector of an individual's beliefs and outcomes before they applied to a service organization.[11] Since TFA_i may be a function of past beliefs and outcomes, this can lead to a bias in the direct estimation of γ using OLS. The key intuition of our approach is that this bias can be overcome if the distribution of unobserved characteristics of individuals who were just below the bar for TFA and the distribution for those who were just above the bar are drawn from the same population:

$$E[\varepsilon_i | score_i = c^* + \Delta]_{\Delta \to 0^+} = E[\varepsilon_i | score_i = c^* - \Delta]_{\Delta \to 0^+} \qquad (2)$$

where $score_i$ is an individual's interview score, and c^* is the cutoff score below which very few applicants are admitted to TFA. Equation (2) implies that the distribution of individuals to either side of the cutoff is as good as random with respect to unobserved determinants of future outcomes (ε_i). In this scenario, we can control for selection into TFA using an indicator variable for whether an individual has an interview score above the cutoff as an instrumental variable. Since service in TFA_i is a discontinuous function of interview score, whereas the distribution of unobservable determinants of future outcomes ε_i is by assumption continuous at the cutoff, the coefficient γ is identified. Intuitively, any discontinuous relation between future outcomes and the interview score at the cutoff can be attributed to the causal impact of service in TFA under the identification assumption in equation (2).

Formally, let TFA placement (TFA_i) be a smooth function of an individual's interview score ($score_i$) with a discontinuous jump at the eligibility cutoff c^*:

$$TFA_i = f(score_i) + \eta \cdot \mathbb{1}\{score_i \geq c^*\} + \varepsilon_i \qquad (3)$$

In practice, the functional form of $f(score_i)$ is unknown. We approximate $f(score_i)$ with a local quadratic in interview score that is allowed to vary on either side of the cutoff. Estimation with a local linear or local cubic regressions yield similar results, as do a variety of bandwidths (see Appendix Table 4). To address potential concerns about discreteness in

---

[10]Previous studies that examine the association between service and future outcomes, such as Yamaguchi et al. (2008) and McAdam and Brandt (2009), estimate equations such as (1).

[11]It is also possible that individuals who select into service organizations prefer to *report* different beliefs even when their actual beliefs are similar (Bertrand and Mullainathan 2001). In our context, this is problematic to the extent that TFA causally impacts the reporting of beliefs independent of underlying beliefs. All of our results using subjective measures should be interpreted with this caveat in mind. Results based on objective outcomes or implicit beliefs are less likely to be affected by this issue.

the interview score in both our first and second stage results, we cluster our standard errors at the interview score level throughout the paper (Card and Lee 2008).

One problem unique to our setting is that the cutoff score c^* must be estimated from the data. TFA does not specify a cutoff score each year. Rather, they select candidates using the interview score as a guide until a prespecified number of teaching slots are filled. Our goal is to identify the unknown score cutoff that best fits the data. We identify this optimal discontinuity point using a technique similar to those used to identify structural breaks in time series data and discontinuities in the dynamics of neighborhood racial composition (Card, Mas, and Rothstein 2008). Specifically, we regress an indicator for TFA selection on a constant and an indicator for having an interview score above a particular cutoff c in the full sample of applicants. We then loop over all possible cutoffs c in 0.0001 intervals, selecting the value of c that maximizes the R^2 of our specification. Hansen (2000) shows that this procedure yields a consistent estimate of the true discontinuity. A standard result in the structural break literature (e.g., Bai 1997) is that one can ignore the sampling error in the location of the discontinuity when estimating the magnitude of the discontinuity. Using different cutoff points around the optimal c^* yield very similar results.

One potential threat to a causal interpretation of our estimates is that survey respondents are not distributed randomly around the cutoff. Such non-random sorting could invalidate our empirical design by creating discontinuous differences in applicant characteristics around the score cutoff. In particular, one may be concerned that former TFA alumni will be more likely to respond than non-alumni, or that the non-alumni who respond will be different in some important way from the alumni that respond. We evaluate this possibility by testing whether the frequency and characteristics of applicants trend smoothly through the cutoff among survey respondents. Figure 1 plots the response rate for 2007 TFA applicants around the cutoff. We also plot fitted values from a regression of an indicator for answering at least one survey question on an indicator for being above the cutoff and a quadratic in interview score interacted with the indicator for being above the cutoff. Consistent with our identifying assumption, the response rate does not change at the cutoff (p-value = 0.921). Panel A of Appendix Table 5 presents analogous results for each survey section, finding no evidence of selective survey attrition around the cutoff.

Figure 2 tests whether the observable characteristics of survey respondents trend smoothly through the cutoff. Following our first stage and reduced form regressions, we plot actual and fitted values from a regression of each characteristic on an indicator for being above the cutoff and a quadratic in interview score interacted with the indicator for being above the cutoff. Respondents above the cutoff have lower College GPAs, but are no more or less likely to be white or Asian, black or Hispanic, male, a Pell Grant recipient, or a math or science

major. Panel B of Appendix Table 5 presents results for the full sample of applicants and for each survey section separately, finding nearly identical results as those reported in Figure 2.

Finally, Figure 3 tests for continuity in the interview subscores which make up the interview score following the same quadratic specification. Respondents above the cutoff have higher critical thinking subscores and marginally lower respect subscores, but have similar scores for achievement, commitment, motivational ability, organizational ability, and perseverance. Panel C of Appendix Table 5 presents analogous results for the full sample and each survey domain separately. None of the results suggest that our identifying assumption is systematically violated.

IV. Results

A. First Stage

First-stage results of the impact of the score cutoff on TFA selection and TFA service are presented graphically in Figure 4. The sample includes all 2007 applicants to TFA who answered at least one question on our survey. Results are identical for the full sample of applicants. TFA selection is an indicator for having been offered a TFA slot. TFA placement is an indicator for having completed the two year teaching commitment. Each figure presents actual and fitted values from a regression of the dependent variable on an indicator for having a score above the cutoff and a local quadratic interacted with having a score above the cutoff.

TFA selection increases by approximately 42 percentage points at the cutoff, while TFA service increases by approximately 36 percentage points. The corresponding estimates are significant at the one percent level, suggesting that our empirical design has considerable statistical power.

However, it is worth emphasizing that the interview score is not perfectly predictive of TFA selection or service due to the nature of the selection process. Applicants with very high interview scores are almost always selected for TFA with little additional review, while applicants with very low scores are rejected without further consideration. Conversely, candidates near the score cutoff for that year will have their application reviewed a second time, with the original interview score playing an important but not decisive role in the selection decision. Moreover, the effect of the cutoff on TFA service is further attenuated by approximately 20 percent of selected applicants turning down the TFA offer. Thus, the score cutoff is only a "fuzzy" predictor of TFA service.

15

B. Main Results

Figure 5 summarizes our main results, and Figures 6 through 9 present results for each set of questions separately. The sample includes all 2007 applicants that answered at least one question in the indicated domain. Following our first stage results, each figure presents actual and fitted values from a regression of the dependent variable on an indicator for having a score above the cutoff and a local quadratic interacted with having a score above the cutoff. Appendix Table 6 reports the corresponding first stage, reduced form, and two-stage least squares effects for each outcome.

Figure 5 suggests that serving in Teach For America increases an individual's faith in education, an individual's involvement in education, and an individual's racial tolerance. The corresponding two stage least squares estimates show that TFA service increases faith in education by 1.315 standard deviations and educational employment by 0.961 standard deviations. TFA service also increases racial tolerance as measured by the black-white IAT by 0.801 standard deviations. Political beliefs remains essentially unchanged.

These effects are quite large. Put differently, TFA service leads to Implicit Association Test scores jumping from the 63rd percentile to the 87th percentile in the distribution of over 700,000 black-white IAT scores collected by Project Implicit in 2010 and 2011. In a 2010 Gallup poll, 22 percent of a nationally representative sample of individuals aged 18 and older reported that school was the most important factor in determining whether students learn. 38 percent of respondents below the cutoff point in our sample said that teachers were the most important in determining how well students perform in school. The impact of TFA service on this belief was 38 percentage points. Similarly, in the first follow-up of the National Education Longitudinal Study of 1988, 16 percent of teachers interviewed strongly disagreed with the statement "there is really very little I can do to ensure that most of my students achieve at a high level". In our sample, 53 percent of respondents below the cutoff point strongly disagreed with the same statement. The impact of TFA service lead to essentially everyone strongly disagreeing with the statement.

Figure 6 presents results for each faith in education variable separately. TFA service increases one's faith in the ability of poor children to compete with more advantaged children, and belief in the importance of teachers in raising student achievement. Two stage least squares estimates suggest that individuals who serve are 44.6 percentage points more likely to believe that poor children can compete with more advantaged children, 35.5 percentage points more likely to believe that the achievement gap is solvable, 38.2 percentage points more likely to believe that teachers are the most important determinant of success, and 65.0 percentage points more likely to disagree that there is little teachers can do to ensure that students succeed. On an open ended question on the percent of minorities we can reasonably

expect to graduate from college, individuals who serve provide answers that are an average of 22.4 percentage points higher than individuals who do not serve.

The effect of TFA on involvement in education is depicted in Figure 7. An important criticism of TFA is that corps members frequently depart before or just after their two-year commitment has been fulfilled (Darling-Hammond et al. 2005). Our results do not address the question of whether TFA teachers are more likely to stay in education compared to other teachers. Instead, we ask whether TFA leads individuals to stay in education longer than they otherwise would have without TFA.

Figure 7 suggests that those who serve in TFA are more likely to be employed in a K - 12 school or in education more generally one to two years after their commitment ends. Our two stage least squares estimates suggest that TFA service increases the probability of being employed in a K - 12 school by 36.5 percentage points and in education more broadly by 43.3 percentage points. TFA alumni are also 31.5 percentage points more likely to believe that service is an important part of their career, and 30.3 percentage points more likely to prefer an urban teaching job over a suburban teaching job. Interestingly, there is not a statistically significant effect of service on wanting to work in education in the future, though the point estimate is economically large. There is also no effect of service on the preference of an urban teaching job over a finance job at the same salary, though this may be because almost all survey respondents prefer teaching.

The effect of TFA on political beliefs is depicted in Figure 8. TFA service does not have a significant impact on political beliefs, at least as we have measured it here. However, we cannot rule out moderate size effects in either direction.

Our final set of outcomes, which measure racial tolerance, are presented in Figure 9. Remarkably, serving in TFA increases implicit black-white tolerance by 0.801 standard deviations. To put this in context, black applicants score 0.558 standard deviations higher than Asian applicants on the black-white IAT, while white and Hispanic applicants score 0.084 and 0.253 standard deviations higher than Asian applicants on the black-white IAT, respectively.

TFA service is also associated with statistically insignificant increases in explicit black-white tolerance in the Modern Racism Scale and the probability of believing that blacks and Hispanics are at least as intelligent as whites and Asians. One interpretation of these results is that while there is little treatment effect on measures of explicit tolerance, TFA increases the unconscious tolerance of its members.

C. Analysis of Subsamples

Table 2 investigates heterogeneous treatment effects across gender, ethnicity, and whether or not a TFA applicant received a Pell Grant in college (a proxy for poverty) controlling for a common quadratic of interview scores. Results are similar, although less precise, if we allow each subgroup to have separate quadratics of interview scores as controls. The impact of service on educational involvement is larger for men and white and Asian applicants, while the impact on faith in education is larger for white and Asian applicants, and applicants who did not receive Pell Grants. TFA service increases a male applicant's educational involvement by 1.100 standard deviations, while increasing a female applicant's educational involvement by 0.863 standard deviations. White and Asian applicants increase their educational involvement by 1.004 standard deviations and faith in education by 1.289 standard deviations compared to 0.481 and 0.989 standard deviations, respectively, for black and Hispanic applicants. Applicants who did not receive Pell Grants also increase their faith in education by 1.398 standard deviations compared to 0.930 standard deviations for applicants who did.

One concern of the above subsample analysis is that we may be detecting false positives due to multiple hypothesis-testing. To address this, we also present results controlling for the Family-Wise Error Rate, which is defined as the probability of making one or more false discoveries - known as type I errors - when performing multiple hypothesis tests, using the Holm step down method described in Romano, Shaikh and Wolf (2010). After correcting for multiple hypothesis-testing, the difference in the impact on faith in education between Pell Grant recipients and non-recipients and the difference in the impact on educational involvement between race groups remains statistically significant.

D. Additional Cohorts

One potential caveat to our analysis is that it includes only one cohort of TFA applicants surveyed roughly a year after their service commitment ended. If there are important longer term impacts of service, our analysis will understate the true impact of TFA. If, on the other hand, the impacts fade over time, our estimates are an upper bound on the true effects of TFA.

To shed some light on this issue, we collected data on TFA applicants in the 2003 to 2006 cohorts. Recall that between 2003 and 2006, TFA admitted candidates who met one of six prespecified interview score requirements. We estimate the impact of service in the 2003 to 2006 cohorts by instrumenting for TFA service using an indicator for whether a candidate meets one of the six subscore criteria for admissions. We include controls for each interview

subscore, with the impact of TFA service identified using the interaction of the subscores. Our key identifying assumption is that the interaction of interview subscores only impacts future outcomes through TFA after controlling for the direct effect of each subscore.[12]

Figure 10 presents results for the impact of service on our summary measures for all available cohorts. We plot reduced form coefficients and associated 95 percent confidence intervals for each cohort. Each estimate comes from a separate regression. The impact of service on educational and racial beliefs and educational involvement is persistent. Alumni from the 2003 to 2006 cohorts are more likely to believe in the power of education, more likely to be employed in education, and are more racially tolerant. Point estimates on the educational beliefs and involvement variables are statistically significant for all alumni cohorts. The racial tolerance point estimate is statistically significant at the 5 percent level for the 2003 cohort, and statistically significant at the 10 percent level for the 2005 and 2006 cohorts. Consistent with our earlier results, there are no systematic impacts on political beliefs.

V. Conclusion

Nearly one million American youth have participated in service programs such as Peace Corps and Teach For America, and annual government spending in support of youth service programs is hundreds of millions of dollars. This paper has shown that serving in Teach For America has a positive impact on an individual's faith in education, involvement in education, and racial tolerance. The impact of service is also quite persistent, with similar effects five years after the completion of the TFA service commitment.

Our results, particularly those on racial beliefs, are broadly consistent with the "Contact Hypothesis," which suggests that contact with other groups will increase tolerance. Changes occur through a combination of increased learning, changed behavior, new affective ties, and reappraisals of one's own group (Pettigrew 1998). A substantial empirical literature suggests that intergroup contact is negatively correlated with intergroup prejudice (Pettigrew and Tropp 2006). Recent research suggests that this correlation may be causal. Van Laar et al. (2005) and Boisjoly et al. (2006) show that white students at a large state university who were randomly assigned black roommates in their first year are more likely to endorse affirmative action, have more personal contact with minority groups, and view a diverse student body as essential for a high-quality education.

---

[12]Appendix Figures 2 and 3 test for selective survey response by cohort. Eligible TFA applicants from the 2003 - 2006 cohorts are approximately ten percentage points more likely to take the online survey. Survey respondents from the 2003 - 2006 cohorts are also more likely to be white or Asian and have higher college GPAs than non-respondents. Our results should be interpreted with these differences in mind.

TFA service typically involves a considerable degree of intergroup contact over a two year period. Eighty percent of alumni members in our sample are white, and 80 percent have at least one parent with a college degree. The average parental income of a corps member is $118 thousand. In stark contrast, roughly 80 percent of the students taught by TFA members qualify for free or reduced-price lunch, and more than 90 percent of these students are African-American or Hispanic.

Note that although our results are consistent with the contact hypothesis, there are other hypotheses that could explain the results without having anything to do with contact with students. For example, TFA and the schools could supply TFA teachers with information that influences the beliefs of these teachers. Indeed, TFA engages in this type of propoganda at summits, annual conferences, and trainings. Unfortunately, we do not have any data that allow us to investigate this.

Taken together, the evidence presented in this paper suggests that TFA service has a significant impact on an individual's values and career decisions. Youth service, particularly service involving extended periods of intergroup contact, may not only help disadvantaged communities, but also help create a more socially conscious and more racially tolerant society.

References

[1] Allport, Gordon W. 1954. "The Nature of Prejudice." Reading, MA: Addison-Wesley.

[2] Barnhardt, Sharon. 2009. "Near and Dear? Evaluating the Impact of Neighbor Diversity on Inter-Religious Attitudes." Unpublished working paper.

[3] Bai, Jushan. 1997. "Estimation of a Change Point in Multiple Regression Models." *The Review of Economics and Statistics*, 79(4): 551-563.

[4] Bertrand, Marianne, and Sendhil Mullainathan. 2001. "Do People Mean What They Say? Implications for Subjective Survey Data." *American Economic Review Papers and Proceedings*, 91(2): 67-72.

[5] Bertrand, Marianne, Dolly Chugh and Sendhil Mullainathan. 2005. "New Approaches to Discrimination: Implicit Discrimination." *American Economic Review*, 95(2): 94-98.

[6] Bertrand, Marianne, Claudia Goldin, and Lawrence Katz. 2010. "Dynamics of the Gender Gap for Young Professionals in the Corporate and Financial Sectors." *American Economic Journal: Applied Economics*, 2(3): 228-255.

[7] Boisjoly, Johanne, Greg J. Duncan, Michael Kremer, Dan M. Levy, and Jacque Eccles. 2006. "Empathy or Antipathy? The Impact of Diversity." *The American Economic Review*, 96(5): 1890-1905.

[8] Brown, Rupert, and Miles Hewstone. 2005. "An Integrative Theory of Intergroup Contact." *Advances in Experimental Social Psychology*, 37: 255-343.

[9] Card, David, and David Lee. 2008. "Regression Discontinuity Inference with Specification Error." *Journal of Econometrics*, 142(2): 655-674.

[10] Card, David, Alexandre Mas, and Jesse Rothstein. 2008. "Tipping and the Dynamics of Segregation." *Quarterly Journal of Economics*, 123(1): 177-218.

[11] Clingingsmith, David, Asim Ijaz Khwaja, and Michael Kremer. 2009. "Estimating the Impact of the Hajj: Religion and Tolerance in Islam's Global Gathering." *Quarterly Journal of Economics*, 124(3): 1133-1170.

[12] Darling-Hammond, Linda, Deborah Holtzman, Su Jin Gatlin, and Julian Vasquez Heilig. 2005. "Does Teacher Preparation Matter? Evidence about Teacher Certification, Teach for America, and Teacher Effectiveness." *Education Policy Analysis Archives*, 13(42).

[13] Dawson, Neal V. and Hal R. Arkes. 2008. "Implicit Bias Among Physicians. " *Journal of General Internal Medicine*, 24(1): 137-140.

[14] Decker, Paul T., Daniel P. Mayer, and Steven Glazerman. 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.

[15] Dobbie, Will. 2011. "Teacher Characteristics and Student Achievement: Evidence from Teach For America." Unpublished Working Paper.

[16] Green, Alexander R., Dana R. Carney, Daniel J. Pallin, Long H. Ngo, Kristal L. Raymond, Lisa I. Iezzoni, and Mahzarin R. Banaji. 2006. "Implicit Bias among Physicians and its Prediction of Thrombolysis Decisions for Black and White Patients. " *Journal of General Internal Medicine*, 22(9): 1231-1238.

[17] Greenwald, Anthony G., T. Andrew Poehlman, Eric L. Uhlmann, and Mahzarin R. Banaji. 2009. "Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity." *Journal of Personality and Social Psychology*, 97: 17-41.

[18] Haan, Norma. 1974. "Changes in Young Adults After Peace Corps Experiences: Political-Social Views, Moral Reasoning, and Perceptions of Self and Parents." *Journal of Youth and Adolescence*, 3(3): 177-194.

[19] Hansen, Bruce E. 2000. "Sample Splitting and Threshold Estimation." *Econometrica*, 68: 575-603.

[20] Hewstone, Miles and Rupert J. Brown. 1986. "Contact and conflict in intergroup encounters." Oxford: Basil Blackwell.

[21] Karpinski, Andrew and James L. Hilton. 2001. "Attitudes and the Implicit Association Test." *Journal of Personality and Social Psychology*, 81(5): 774-788.

[22] Krueger, Alan, and Andreas Mueller. 2011. "Job Search, Emotional Well-Being and Job Finding in a Period of Mass Unemployment: Evidence from High-Frequency Longitudinal Data." Unpublished Working Paper.

[23] McAdam, Doug and Cynthia Brandt. 2009. "Assessing the Effects of Voluntary Youth Service: The Case of Teach For America. " *Social Forces*, 88(2): 945-970.

[24] McConahay, John B. 1983. "Modern Racism and Modern Discrimination: The Effects of Race, Racial Attitudes, and Context on Simulated Hiring Decisions." *Personality and Social Psychology Bulletin*, 37(2): 551-558.

[25] McConnell, Allen R. and Jill M. Leibold. 2001. "Relations among the Implicit Association Test, Discriminatory Behavior, and Explicit Measures of Racial Attitudes." *Journal of Experimental Social Psychology*, 37(5): 436-442.

[26] McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design." *Journal of Econometrics*, 142(2)698-714.

[27] Moss, Marc, Janet Swartz, Dawn Obeidallah, Gerrie Stewart, and Diane Greene. 2001. "AmeriCorps Tutoring Outcomes Study." Cambridge, MA: Abt Associates.

[28] Pettigrew, Thomas F. 1998. "Intergroup Contact Theory." *Annual Review of Psychology*, 49: 65-85.

[29] Pettigrew, Thomas F. and Linda R. Tropp. 2006. "A Meta-Analytic Test of Intergroup Contact Theory." *Journal of Personality and Social Psychology*, 90(5): 751-783.

[30] Pettigrew, Thomas F. and Linda R. Tropp. 2008. "How Does Intergroup Contact Reduce Prejudice? Meta-analytic Tests of Three Mediators." *European Journal of Social Psychology*, 38(6): 922-934.

[31] Phelps, Elizabeth A., Kevin J. O'Connor, William A. Cunningham, E. Sumie Funayama, J. Christopher Gatenby, John C. Gore, Mahzarin R. Banaji. 2000. "Performance on Indirect Measures of Race Evaluation Predicts Amygdala Activation." *Journal of Cognitive Neuroscience*, 12(5): 729-738.

[32] Romano, Joseph P., Azeem M. Shaikh, and Michael Wolf. 2010. "Hypothesis Testing in Econometrics." *Annual Review of Economics*, Vol. 2: 75-104.

[33] Rooth, Dan-Olof. 2007. "Implicit Discrimination in Hiring: Real World Evidence." IZA Discussion Paper No. 2764.

[34] Rothermund, Klaus and Dirk Wentura. "Underlying Processes in the Implicit Association Test: Dissociating Salience From Associations." *Journal of Experimental Psychology*, 133(2): 139-165.

[35] Rudman, Laurie A. and Matthew R. Lee. 2002. "Implicit and Explicit Consequences of Exposure to Violent and Misogynous Rap Music." *Group Processes and Intergroup Relations*, 5(2): 133-150.

[36] Sriram, N., and Anthony G. Greenwald. 2009. "The Brief Implicit Association Test." *Experimental Psychology*, 56: 283"294.

[37] Steffens, Melanie C. 2004. "Is the Implicit Association Test Immune to Faking? " *Experimental Psychology*, 51(3): 165-179.

[38] U.S. Department of Education. 2010. Institute of Education Sciences, National Center for Education Statistics.

[39] Van Laar, Colette, Shana Levin, Stacey Sinclair, and Jim Sidanius. 2005. "The Effect of University Roommate Contact on Ethnic Attitudes and Behavior." *Journal of*

Experimental Social Psychology, 41: 329-345.

[40] Yamaguchi, Ryoko, Philip Gordon, Christopher Mulvey, Fatih Unlu, Laura Simpson, JoAnn Jastrzab, Christopher Winship, Cristofer Price, Ken Lam, Cay Bradley, Melanie Brown-Lyons, Robert Grimm, Kevin Cramer, LaMonica Shelton, Nathan Dietz, Lillian Dote, and Shelby Jennings. 2008. "Still Serving: Measuring the Eight-Year Impact of AmeriCorps on Alumni. 2008. " Cambridge, MA: Abt Associates.
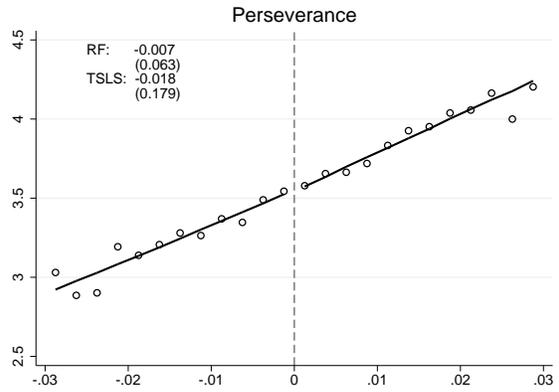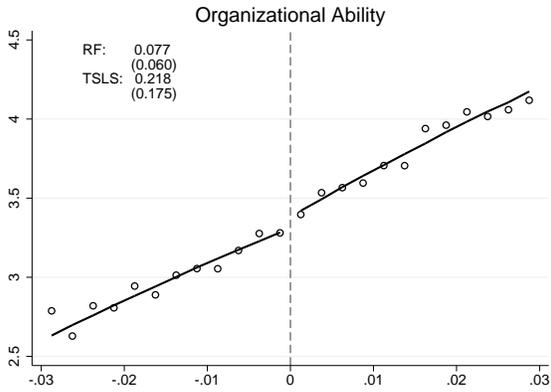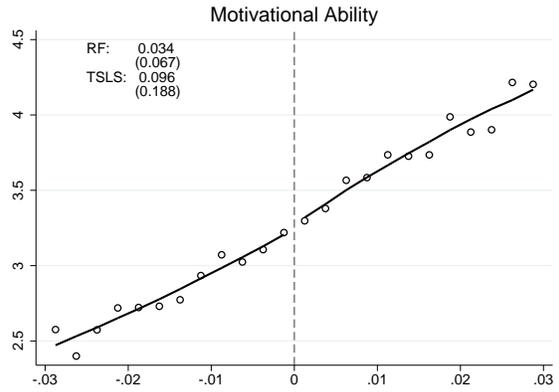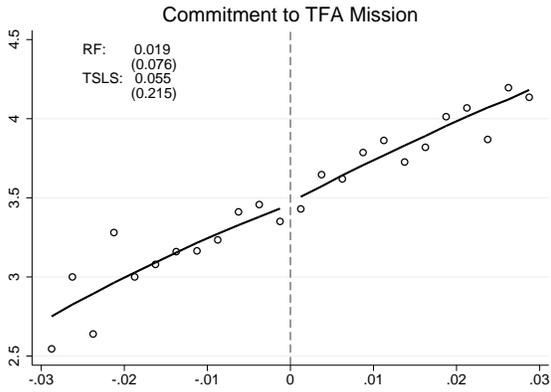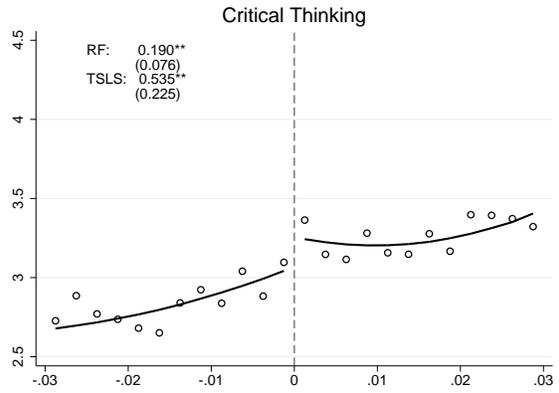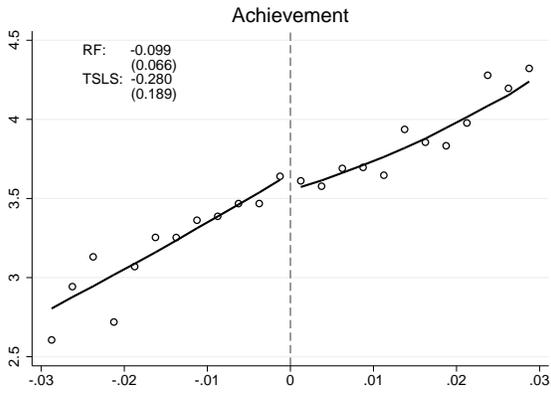
Figure 1
Survey Response
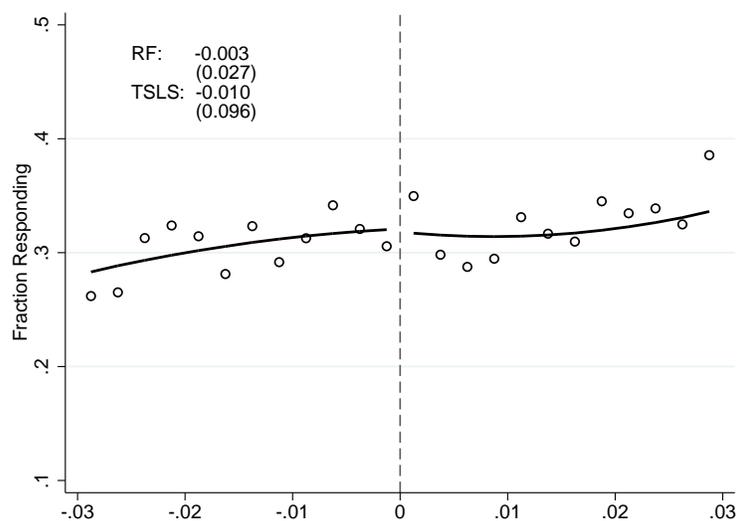
Notes: This figure presents actual and fitted values for 2007 TFA applicants. The actual values are plotted in bins of size 0.0025. The fitted values come from a local regression of survey response on an indicator variable for scoring above the cutoff score and a quadratic in interview score interacted with an indicator variable for scoring above the cutoff score. The reduced form and two stage least squares estimates along with their standard errors are reported on the figure. Standard errors are clustered at the interview score level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.
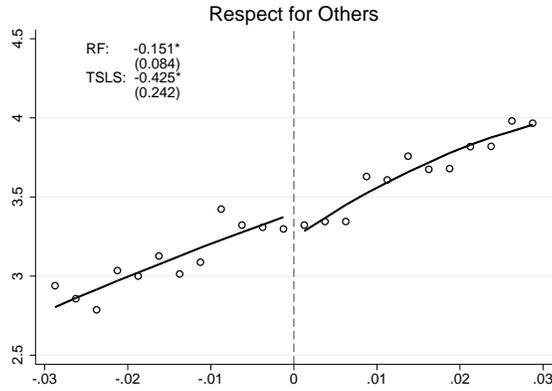
Figure 2
Baseline Characteristics in Survey Sample
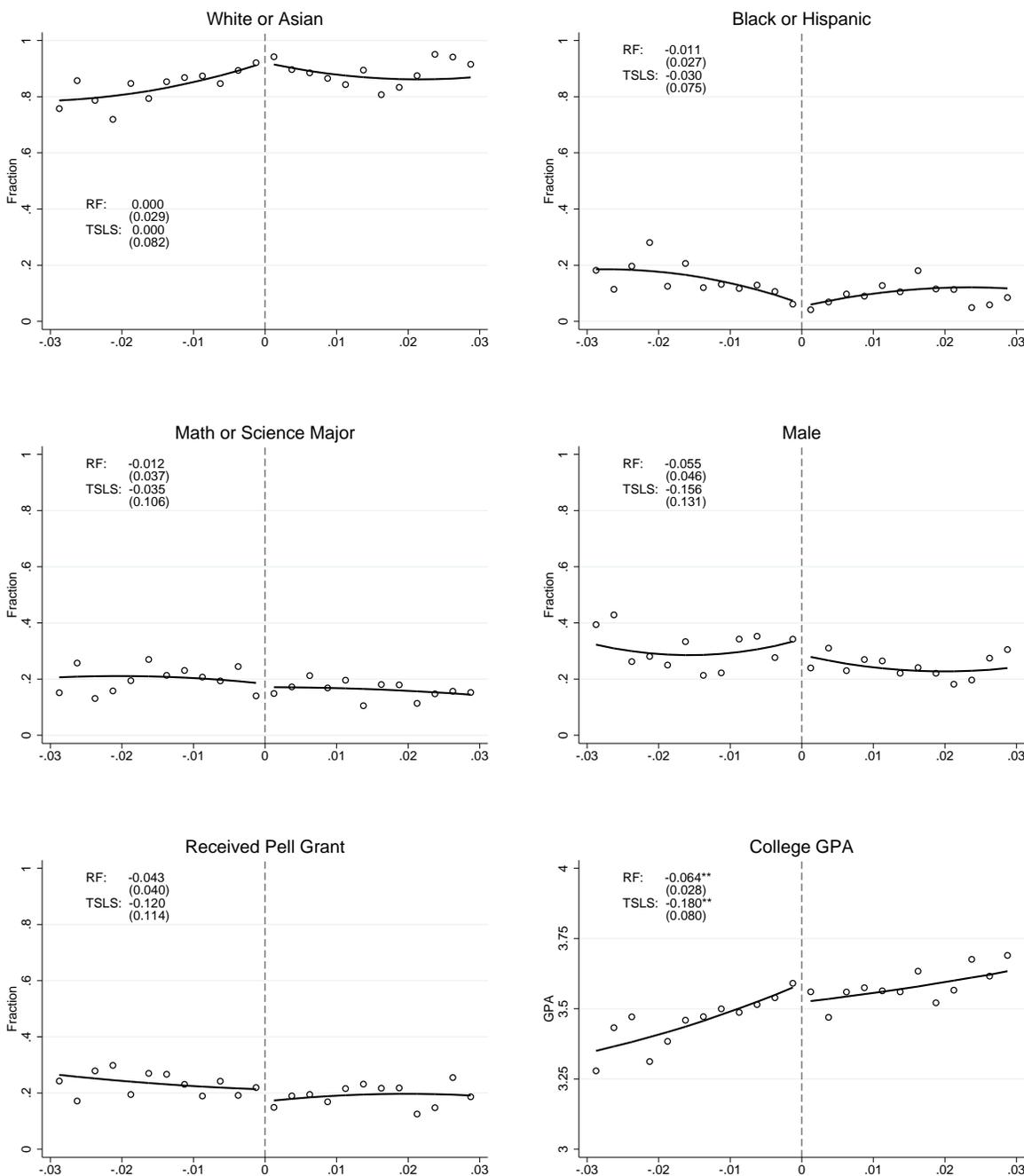
Notes: These figures present actual and fitted values for the survey sample of 2007 TFA applicants. The actual values are plotted in bins of size 0.0025. The fitted values come from a local regression of each baseline characteristic on an indicator variable for scoring above the cutoff score and a quadratic in interview score interacted with an indicator variable for scoring above the cutoff score. The reduced form and two stage least squares estimates along with their standard errors are reported on each plot. Standard errors are clustered at the interview score level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Figure 3
Interview Subscores in Survey Sample

Respect for Others

RF: -0.151*
(0.084)
TSLS: -0.425*
(0.242)

Notes: These figures present actual and fitted values for the survey sample of 2007 TFA applicants. The actual values are plotted in bins of size 0.0025. The fitted values come from a local regression of each subscore on an indicator variable for scoring above the cutoff score and a quadratic in interview score interacted with an indicator variable for scoring above the cutoff score. The reduced form and two stage least squares estimates along with their standard errors are reported on each plot. Standard errors are clustered at the interview score level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.
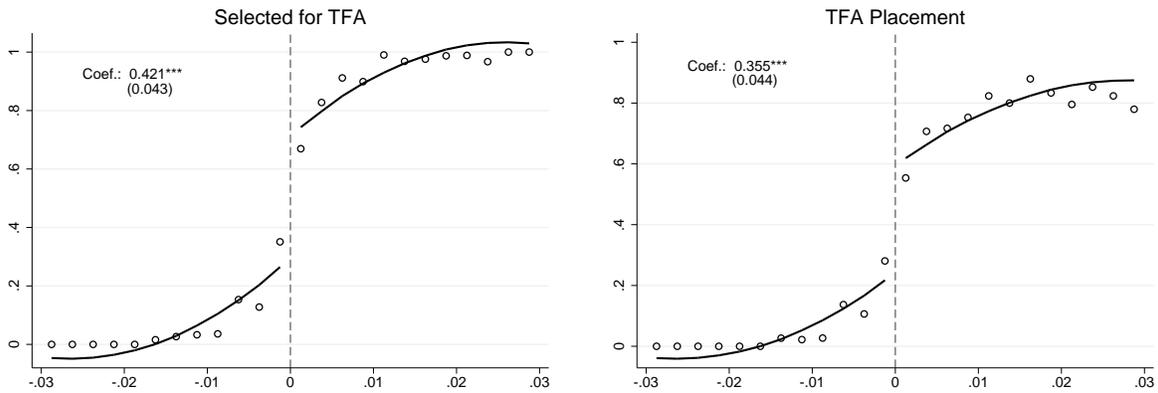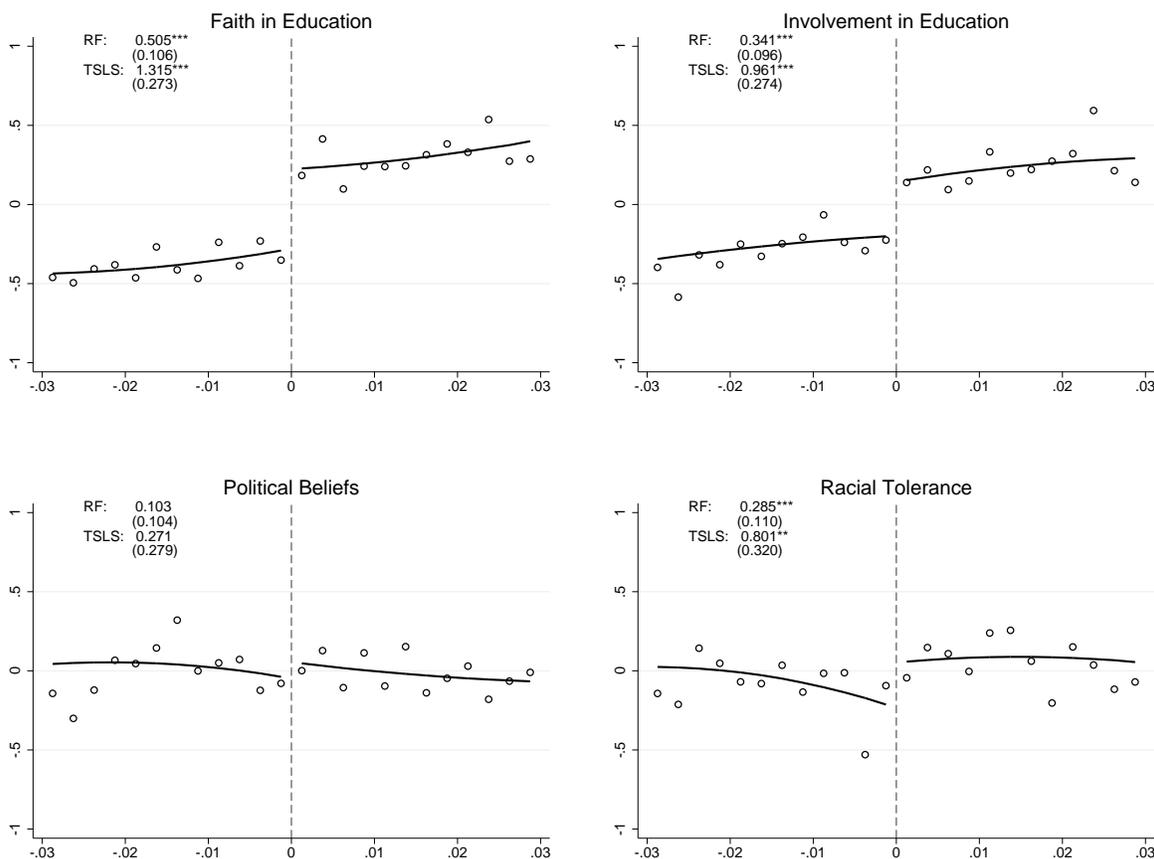
Figure 4
First Stage Results

Notes: These figures present actual and fitted values for the survey sample of 2007 TFA applicants. The actual values are plotted in bins of size 0.0025. The fitted values come from a local regression of each variable on an indicator variable for scoring above the cutoff score and a quadratic in interview score interacted with an indicator variable for scoring above the cutoff score. The first stage estimate along with its standard error is reported on each plot. Standard errors are clustered at the interview score level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Figure 5
Summary of Main Results

Faith in Education

RF: 0.505***
(0.106)
TSLS: 1.315***
(0.273)

Involvement in Education

RF: 0.341***
(0.096)
TSLS: 0.961***
(0.274)

Political Beliefs

RF: 0.103
(0.104)
TSLS: 0.271
(0.279)

Racial Tolerance
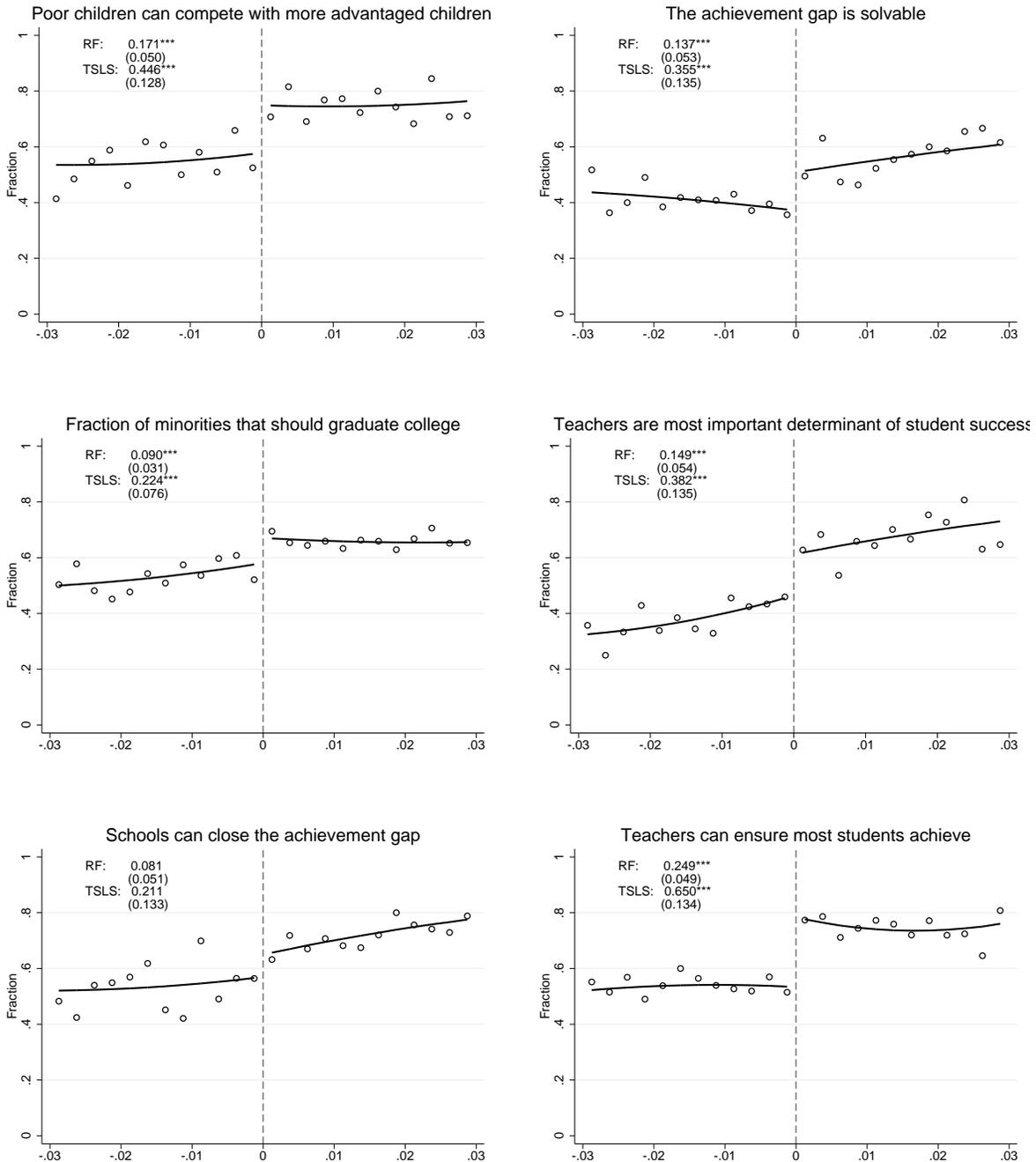
RF: 0.285***
(0.110)
TSLS: 0.801**
(0.320)

Notes: These figures present actual and fitted values for the survey sample of 2007 TFA applicants. The actual values are plotted in bins of size 0.0025. The fitted values come from a local regression of each variable on an indicator variable for scoring above the cutoff score and a quadratic in interview score interacted with an indicator variable for scoring above the cutoff score. Each index was constructed by standardizing the sum of all questions in that area to have a mean of zero and a standard deviation of one. All standardized variables were converted to indicator variables using the median of the full sample. The variables included in each composite variable are available in Online Appendix B. The reduced form and two stage least squares estimates along with their standard errors are reported on each plot. Standard errors are clustered at the interview score level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.
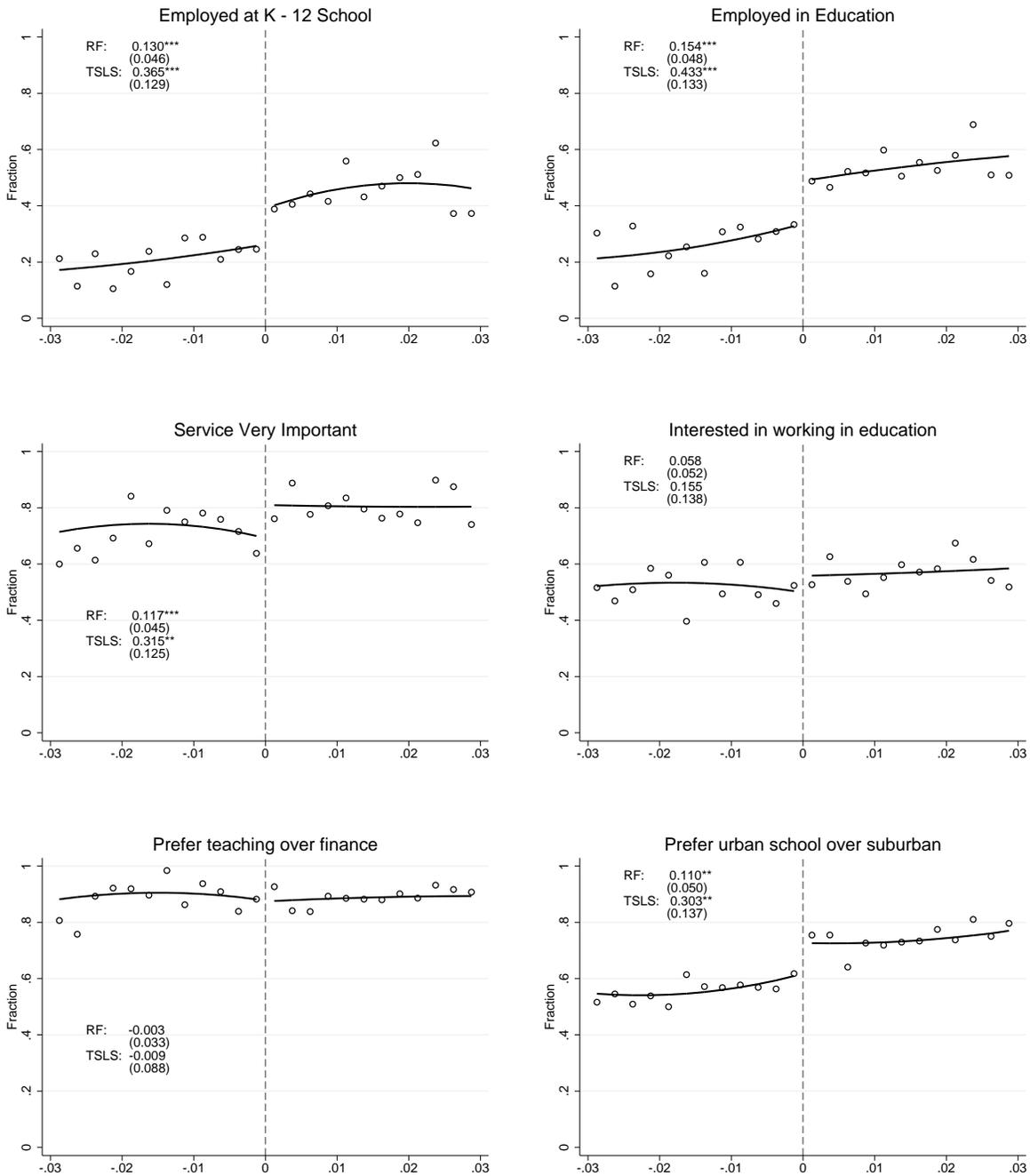
Figure 6
Faith in Education

Notes: These figures present actual and fitted values for the survey sample of 2007 TFA applicants. The actual values are plotted in bins of size 0.0025. The fitted values come from a local regression of each variable on an indicator variable for scoring above the cutoff score and a quadratic in interview score interacted with an indicator variable for scoring above the cutoff score. The reduced form and two stage least squares estimates along with their standard errors are reported on each plot. Standard errors are clustered at the interview score level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.
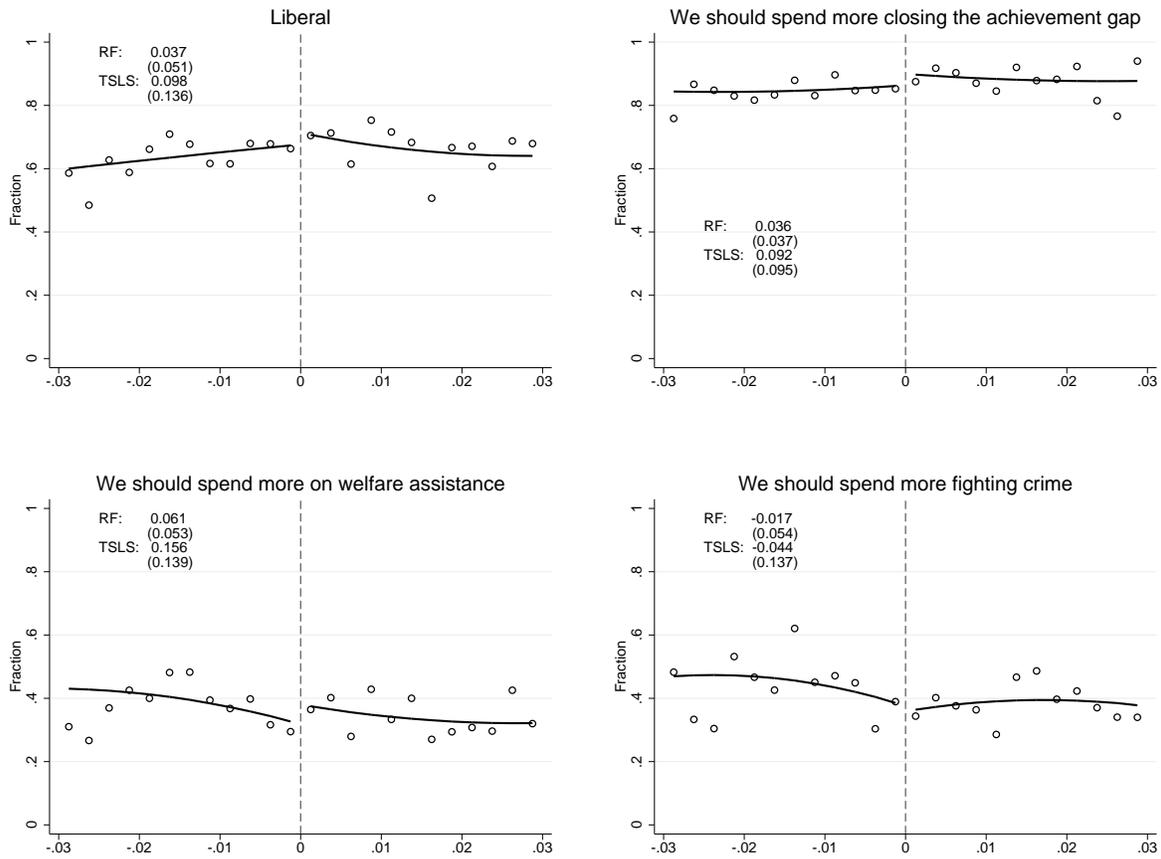
Figure 7
Involvement in Education

Notes: These figures present actual and fitted values for the survey sample of 2007 TFA applicants. The actual values are plotted in bins of size 0.0025. The fitted values come from a local regression of each variable on an indicator variable for scoring above the cutoff score and a quadratic in interview score interacted with an indicator variable for scoring above the cutoff score. The reduced form and two stage least squares estimates along with their standard errors are reported on each plot. Standard errors are clustered at the interview score level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.
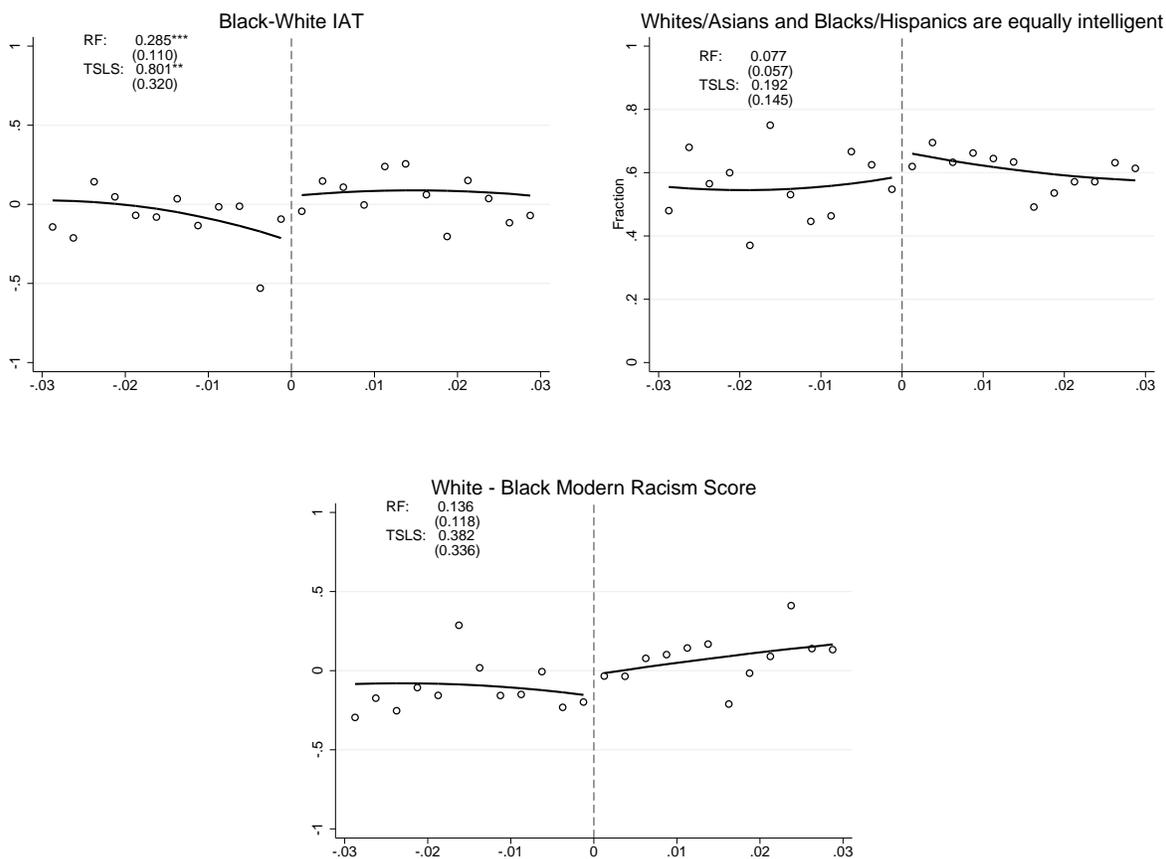
32

Figure 8
Political Beliefs

Notes: These figures present actual and fitted values for the survey sample of 2007 TFA applicants. The actual values are plotted in bins of size 0.0025. The fitted values come from a local regression of each variable on an indicator variable for scoring above the cutoff score and a quadratic in interview score interacted with an indicator variable for scoring above the cutoff score. The reduced form and two stage least squares estimates along with their standard errors are reported on each plot. Standard errors are clustered at the interview score level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.
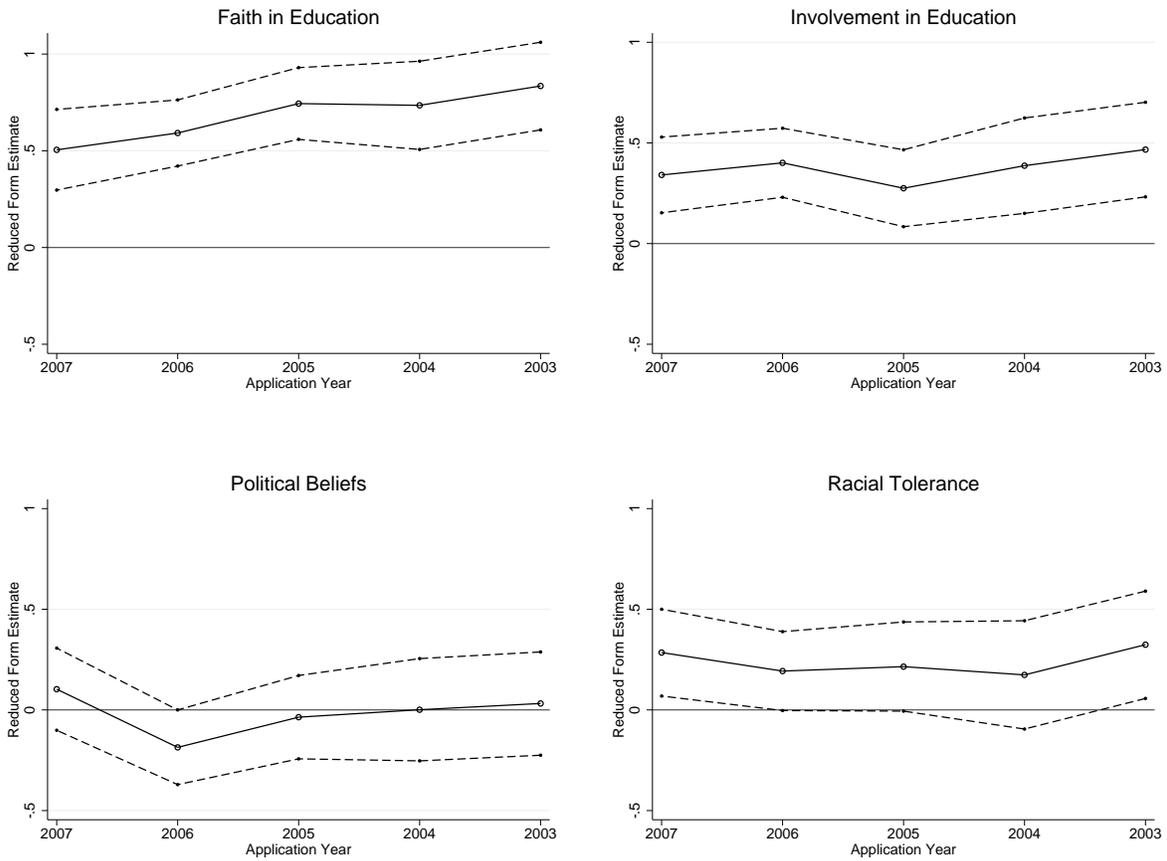
Figure 9
Racial Tolerance

Notes: These figures present actual and fitted values for the survey sample of 2007 TFA applicants. The actual values are plotted in bins of size 0.0025. The fitted values come from a local regression of each variable on an indicator variable for scoring above the cutoff score and a quadratic in interview score interacted with an indicator variable for scoring above the cutoff score. The reduced form and two stage least squares estimates along with their standard errors are reported on each plot. Standard errors are clustered at the interview score level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Figure 10
Main Results by Cohort

Notes: This figure presents reduced form estimates by cohort. The sample includes all 2003 - 2007 applicants who answered at least one survey question. Results for the 2007 cohort are estimated using a regression discontinuity design, controlling for a local quadratic in interview score interview score interacted with scoring above the cutoff score. Results for the 2003 - 2006 cohorts are estimated using the interaction between interview subscores that determines TFA selection, controlling for the impact of each interview subscore. We report the point estimate for serving in TFA. See text for details.

Table 1
Summary Statistics

	TFA			Not TFA		
	Mean	SD	N	Mean	SD	N
Background Variables	(1)	(2)	(3)	(4)	(5)	(6)
White	0.808	0.394	1,023	0.791	0.407	1,277
Asian	0.061	0.239	1,023	0.067	0.249	1,277
Black	0.063	0.242	1,023	0.073	0.260	1,277
Hispanic	0.050	0.218	1,023	0.050	0.218	1,277
College GPA	3.578	0.283	1,023	3.480	0.351	1,277
Received Pell Grant	0.198	0.399	1,023	0.220	0.414	1,277
Math or Science Major	0.168	0.374	1,023	0.186	0.390	1,277
Married	0.110	0.314	1,023	0.142	0.349	1,277
Mother has BA	0.327	0.469	995	0.413	0.493	1,232
Mother has more than BA	0.404	0.491	995	0.324	0.468	1,232
Father has BA	0.279	0.449	994	0.283	0.451	1,229
Father has more than BA	0.467	0.499	994	0.411	0.492	1,229
Faith in Education						
Poor children can compete with more advantaged children	0.803	0.398	917	0.546	0.498	1,115
The achievement gap is solvable	0.599	0.490	917	0.409	0.492	1,115
Fraction of minorities that should graduate college	0.679	0.250	781	0.537	0.270	894
Teachers are most important determinant of student success	0.738	0.440	896	0.382	0.486	1,070
Schools can close the achievement gap	0.772	0.420	916	0.532	0.499	1,117
Teachers can ensure most students achieve	0.802	0.399	917	0.534	0.499	1,117
Involvement in Education						
Employed at K - 12 School	0.517	0.500	1,023	0.193	0.395	1,277
Employed in Education	0.622	0.485	1,023	0.243	0.429	1,277
Service Very Important	0.822	0.383	955	0.718	0.450	1,162
Prefer teaching over finance	0.894	0.308	946	0.882	0.322	1,138
Prefer urban school over suburban	0.803	0.398	947	0.550	0.498	1,141
Interested in working in education	0.604	0.489	954	0.503	0.500	1,168
Political Beliefs						
Liberal	0.664	0.472	909	0.643	0.479	1,106
We should spend more closing the achievement gap	0.890	0.313	876	0.851	0.356	1,040
We should spend more on welfare assistance	0.307	0.462	876	0.410	0.492	1,040
We should spend more fighting crime	0.377	0.485	876	0.434	0.496	1,040
Racial Tolerance						
IAT White-Black	0.089	1.036	841	-0.074	0.963	1,015
Whites/Asians and Blacks/Hispanics are equally intelligent	0.601	0.490	764	0.578	0.494	924
White - Black Modern Racism Score	0.106	0.906	794	-0.089	1.065	946

This table reports summary statistics for the 2007 TFA application cohort. The sample includes all applicants who answered at least one survey question.

Table 2
Subsample Results

	Male	Female	p-value	Holm p-value	Asian/ White	Black/ Hispanic	p-value	Holm p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Faith in Education	1.470***	1.259***	0.172	1.000	1.289***	0.989***	0.087	0.868
	(0.317)	(0.266)			(0.280)	(0.328)		
	569	1,460			1,764	232		
Involvement in Education	1.100***	0.863***	0.092	0.868	1.004***	0.481	0.002	0.021
	(0.302)	(0.265)			(0.281)	(0.324)		
	641	1,649			1,979	271		
Political Beliefs	0.272	0.230	0.793	1.000	0.283	0.379	0.609	1.000
	(0.322)	(0.267)			(0.285)	(0.340)		
	568	1,451			1,756	230		
Racial Tolerance	0.709**	0.807**	0.562	1.000	0.738**	0.816**	0.718	1.000
	(0.359)	(0.314)			(0.324)	(0.404)		
	519	1,329			1,601	216		

37

Table 2 Continued
Subsample Results

	Pell Grant	No Pell Grant	p-value	Holm p-value
	(9)	(10)	(11)	(12)
Faith in Education	0.930***	1.398***	0.003	0.029
	(0.302)	(0.273)		
	417	1,615		
Involvement in Education	0.985***	0.969***	0.917	1.000
	(0.305)	(0.272)		
	484	1,810		
Political Beliefs	0.221	0.301	0.628	1.000
	(0.306)	(0.279)		
	413	1,609		
Racial Tolerance	0.749**	0.816**	0.701	1.000
	(0.359)	(0.318)		
	391	1,461		

This table reports two-stage least squares estimates interacted by subsample. The sample is all 2007 applicants who answered at least one question included in the composite index. All regressions control for a common local quadratic trend in the interview score interacted with an indicator variable for scoring above the cutoff score. Columns (3), (7), and (11) report the unadjusted p-value from the test that the coefficients reported in the preceding two columns are equivalent. Columns (4), (8), and (12) report p-values for the same tests as columns (3), (7), and (11), respectively, but control for the Familywise Error Rate, the probability of at least one false rejection. Specifically, columns (4), (8), and (12) use the Holm stepdown method described in Romano, Shaikh, and Wolf (2010). Standard errors are clustered at the interview score level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.