

# The impact of Teach First on pupil attainment at age 16

*Rebecca Allen and Jay Allnutt*

We evaluate whether the placement of Teach First's inexperienced new teachers into secondary schools with recruitment difficulties in disadvantaged areas in England has raised or lowered pupil attainment at the age of 16. Our matched difference-in-difference panel estimation approach compares the experiences of 168 schools participating early on in the scheme to those in the same region who will go on to participate in later years. We find the programme has not been damaging to these schools who joined and most likely produced school-wide gains in GCSE results of around one grade in one of the pupils' best eight subjects. Similarly, we estimate departmental gains of over 5% of a subject grade resulting from placing a Teach First participant in a teaching team of six teachers. The estimation approach cannot assert whether these gains arise solely through the greater teaching quality of Teach First participants compared to those they displace, or whether spillovers raise standards across the department or somehow change the ethos within the school.

## Introduction

---

Teach First, a programme which has many similarities with Teach for America and other 'Teach for All' schemes across the world, has been placing graduates into schools in challenging circumstances since 2003. These schools, which are eligible by virtue of their levels of free school meals take-up, have traditionally struggled to recruit high-quality teachers and maintain low teacher turnover (Lack & Johnston, 2008; Allen et al., 2012). At the time of its inception, Teach First provided a radical challenge to the standard route of university-led initial teacher training that was taken by most new teachers. The Teach First participants commit to teach up to 80% of a standard teaching load for two years following six weeks of intensive basic training and are able to achieve fully qualified teacher status by the end of the programme, with in-school and partner university support throughout. After two years, some- where between 57% and 63% of the cohort chooses to remain in a state-funded school for at least a third year, with the rest pursuing careers in other fields (Allen et al., 2016a). These retention rates are far lower than for other teacher training routes, especially given the age profile of the participants. The equivalent retention rate for full-time PGCE students is between 64% and 70%, and it is six percentage points higher for PGCE students who share similar demographic characteristics to Teach First participants. Over the past decade the scheme has grown from fewer than 200 graduates in its first cohort to around 1500 today, has extended its reach from London into all parts of England and Wales, has expanded its recruitment to include later career participants and since 2008 has placed participants in primary schools. The quality of the teacher training and work of its participants has been praised in qualitative evaluations (e.g. Hutchings et al., 2006; Ofsted, 2011) and it claims to have a much higher ratio of applicants per place than on traditional training routes. However, given the low retention rates, it is a highly expensive teacher training route (Allen et al., 2016b).

As increasing numbers of Teach for All programmes are established across the world, it is important that success is evaluated and understood in the very wide variety of contexts under which it operates. The programme appears to challenge the traditional model of university-based teacher training and asserts that it is possible for graduates with limited intensive training to thrive in often very challenging school environments. In England and Wales, Teach First is a key programme on which successive governments have placed a great deal of emphasis, inspiring a wider move towards school-led teacher training across the country.

The similar Teach for America programme has been evaluated extensively. Evidence in elementary and middle schools is mixed, but typically shows that participants have a small, statistically significant positive impact on pupil attainment in maths and science, with inconsistent results for other subjects (key studies include Raymond et al., 2001; Decker et al., 2004; Darling-Hammond et al., 2005; Boyd et al., 2006; Kane et al., 2008; Antecol et al., 2013). The studies (Schoeneberger et al., 2009; Henry et al., 2010; Schoeneberger, 2011; Ware et al., 2011; Xu et al., 2011) which focus on high-school outcomes all show that there is a positive achievement impact of Teach for America across all subjects, which concords with the perspective that a teacher's academic qualifications are more important with older students (Goldhaber & Brewer, 2000). This US high-school evidence suggests that positive impacts should be found in England, where placement so far is overwhelmingly with older students. However, the relative success of the programme within a local context clearly depends not just on the efficacy of its implementation, but more importantly on local teacher labour market conditions and on the quality of the pre-existing recruits into new teaching positions in deprived schools.

In this paper we evaluate whether the placement of Teach First graduates has altered the educational outcomes of pupils at age 16 for the first three years of a school's participation in the scheme. We do not have matched teacher-pupil data in England so, unlike the best Teach for America evaluations (e.g. Decker et al., 2004), we are forced to identify an impact across the whole school and within the key departments where these graduates are placed. There is no random element in the sign-up of schools to the scheme, so the obvious threats to validity are that (i) headteachers who choose to join this scheme are particularly dynamic and so preside over improving schools; (ii) conversely, schools using the scheme have particularly severe teacher recruitment and retention problems that may reflect underlying difficulties at the school; and (iii) Teach First launched in the London region, which was subject to multiple interventions to improve pupil attainment running concurrently with Teach First.

To identify any school impact, we match early-participating Teach First schools to those within the same region which participated later in the scheme and then estimate impact within a difference-in-difference panel to control for any further time-invariant endogenous and unobservable variables which could otherwise bias estimates. To identify departmental impacts (and assuming no spillovers across the school), we (i) estimate triple-difference estimators that compare changes in exam outcomes across departments within the same school and (ii) estimate pupil point-in-time fixed-effect models to account directly for unobserved pupil characteristics. Whilst our approaches to estimating departmental impacts are arguably more robust than our school-level impacts, all our methods reduce the risk of biased estimates compared to the large, positive effects found using matched multi-level cross-sectional models in the only existing quantitative evaluation to date (Muijs et al., 2010).

## Data

---

Ideally, teacher-pupil matched data would be available to allow us to measure directly the success of pupils taught by a Teach First participant compared to those taught by others. Unfortunately, this is not routinely collected anywhere in the UK, so instead our analysis aggregates datasets to create school and departmental-level data to estimate overall direct and indirect impacts on the school overall. We combine two sources for this paper: Teach First's database of participant records from 2003/4 to 2012/3 and the National Pupil Database for England, which is collected and maintained by the Department for Education. Table 1 shows the number of schools taking part in the Teach First programme by year of first participation. It illustrates the gradual roll-out across regions, starting with London in 2003/04 and followed by the North West in 2006/07, the West Midlands in 2007/08 and the East Midlands and Yorkshire in 2009/10.

Our binary indicator (*TF*) of Teach First participation for the school overall and for their English, maths and science departments takes a value of one once the school/ department has participated for the first

time, regardless of whether they continue to use the scheme in future. We only measure whether Teach First participants were present in the school in the pupils' final year of compulsory schooling, because we do not yet have enough data to look at impacts over five years (this lack of lag does mean that the impact of Teach First could be understated in this paper).

Our indicator deliberately does not distinguish between a school that chooses to take on dozens of Teach First participants each year and a school that takes on just one in only one year. (The average number of Teach First participants in a school is 2.9, 4.6 and 3.6 in years one, two and three, respectively. This compares to a teaching staff of 70 in a typical secondary school.) This is because such response by schools beyond their first participation decision is endogenous to their experience of the Teach First programme itself. That is, those schools with a positive experience of the participants allocated to them are more likely to continue with the programme, which would lead to an upward bias on estimates. This approach contrasts with the other major evaluation of Teach First, which identifies whether a school has participated for at least four of the following six years, with obvious resulting positive selection bias since any schools that decided to discontinue involvement due to a poor experience are dropped from the analysis. (All results using a continuous measure that identifies the number of participants in the school are available from the author.)

**Table 1: Number of schools in Teach First programme (by year of first participation)**

	London	North West	West Midlands	East Midlands	Yorkshire	East of England	South East	North East	South West	Total
<b>Cohort 2003/04</b>	<b>43</b>	0	0	0	0	0	0	0	0	<b>43</b>
<b>Cohort 2004/05</b>	<b>19</b>	0	0	0	0	0	0	0	0	<b>19</b>
<b>Cohort 2005/06</b>	<b>14</b>	0	0	0	0	0	0	0	0	<b>14</b>
<b>Cohort 2006/07</b>	<b>10</b>	<b>18</b>	0	0	0	0	0	0	0	<b>28</b>
<b>Cohort 2007/08</b>	<b>10</b>	<b>6</b>	<b>17</b>	0	0	0	0	0	0	<b>33</b>
<b>Cohort 2008/09</b>	<b>13</b>	<b>10</b>	<b>11</b>	0	0	0	0	0	0	<b>34</b>
<b>Cohort 2009/10</b>	<b>8</b>	<b>11</b>	<b>11</b>	<b>12</b>	<b>20</b>	0	0	0	0	<b>62</b>
<b>Cohort 2010/11</b>	5	10	13	9	7	3	3	0	0	<b>51</b>
<b>Cohort 2011/12</b>	21	13	14	5	16	1	4	20	0	<b>94</b>
<b>Cohort 2012/13</b>	23	10	14	8	12	2	8	9	0	<b>86</b>
<b>Never joined TF</b>	267	421	323	249	270	349	499	141	318	<b>2837</b>

We extract a pupil record from the National Pupil Database for every 16-year-old student at a state-maintained school for the years 2002 through to 2012. These data are collapsed to a school-level dataset for the majority of the analysis in order to implement a panel data approach, though we also retain pupil-level records for the pupil fixed-effects estimation described in the next section. These records provide us with a prior attainment score at age 11 in English, maths and science, indicators of the child's gender, age in months, ethnicity, free school meal status, deprivation of home neighbourhood (IDACI) and special educational needs.

We use a broad measure of age 16 pupils' exam performance across their best eight subjects in GCSE and equivalent exams, standardising to a (pupil-level) mean of zero and standard deviation of one (capped GCSE z-score). We also report a threshold measure of the proportion of students gaining five or more GCSEs at grades A\*-C, including English and maths. Core subject departmental performance is measured by taking the child's best grade in the subject, scored on a scale of 0 (=U or no entry) to 8 (=A\*).

## Method

---

The group of schools and departments in schools that take on Teach First participants are clearly unusual, so the matching of participating schools to those schools which have not yet taken up the programme is central to dealing with quite serious potential endogeneity bias. The matching is combined with a school-level difference-in-difference regression, with school fixed effects soaking up unobserved time-invariant characteristics and background control variables that are intended to account for observable time-variant population changes at the school.

We introduce an education production function to illustrate the diverse means by which the introduction of Teach First participants might affect the attainment,  $y$ , of pupil  $i$  in subject  $h$  in school  $j$  at time  $t$ . Suppose we can separate the impact of the pupil's own (time-varying or otherwise) characteristics  $X_{ijt}$  and prior attainment  $y_{hijt-1}$  from the impact of the school they attend,  $S_{hjt}$ :

$$y_{hijt} = f(y_{hijt-1}, X_{ijt}, S_{hjt}(T_{hjt}, D_{hjt}, R_{jt}, L_{jt}))$$

The impact of the school on the pupil's attainment in subject  $h$  in turn depends on their subject teacher's effectiveness,  $T_{hjt}$ , departmental ethos and decision-making regarding curriculum and exam entry,  $D_{hjt}$ , the non-teaching resources in the school,  $R_{jt}$ , and the quality of school leadership, ethos and whole-school activities,  $L_{jt}$ . Of course, all these aspects of the school experience are inter-linked, especially in the long run. For example, strong school leadership might influence teacher quality through attracting good teachers, investing in effective training and motivating staff to work hard.

A school's participation in Teach First might influence pupil  $i$ 's attainment in subject  $h$  in a number of ways. First, pupil  $i$  may be taught by a Teach First participant who is more or less effective than the teacher they would otherwise have been allocated had the school not joined the scheme (i.e. directly through  $T_{hjt}$ ). Second, even if not directly taught by one, the presence of a Teach First participant may raise or lower the general standard of teaching in the department, through raising expectations or the creation of new shared resources or negatively through other teachers' need to provide mentoring time and support to the inexperienced teacher (i.e. indirectly through  $D_{hjt}$ ). An Ofsted report (2011) on Teach First suggested that the programme was having an impact 'on the professional development of other staff as well as on their students' (p. 5). We cannot know whether or not they are displacing the presence of other trainee teachers. Finally, Teach First participants may engage in activities that impact on the ethos of the school more widely (i.e. indirectly through  $L_{jt}$ ). Unfortunately, without linked teacher-pupil data it is difficult to distinguish between the first two of these three mechanisms, so instead we estimate the impact of Teach First on departments and on the school more generally.

### Propensity score matching

We use matching to identify an untreated set of schools, dealing with a number of potential sources of selection bias without the imposition of functional form assumptions or risk of insufficient common support. Participating in Teach First may reflect difficulty in recruiting teachers or high turnover, which in turn suggests higher levels of deprivation (cf Allen et al., 2012) or local reputational difficulties. Conversely, headteachers who are attracted to the programme may be different from others (e.g. less conservative) and this might be correlated with improvements in effectiveness at the school. Either way, it suggests that Teach First-participating schools would best be matched to others that choose to join the programme at some point in the future.

Within the group of future participating schools, we may be concerned that schools joining at a later stage are somehow less committed or suited to the programme, in which case the best match would be to those future participating schools who were geographically blocked from taking part in the early years because the programme did not yet operate in their area. However, on testing this type of match we

encounter two problems: first, the match across regions is poorly balanced on ethnicity and English as an additional language characteristics; second, we know that different regions of England were subject to different policy regimes over this time period. Specifically, 60% of our treatment schools are in London, a city where exam results were rapidly improving over this period of time. Whilst part of this phenomenon could indeed have reflected Teach First’s penetration in the city, funding and new support for schools under the London Challenge project almost certainly contributed to school improvement (Ofsted, 2010) and the city also experienced significant demographic change during this time (Blanden et al., 2015).

We therefore decide to restrict our potential control schools to any future Teach First schools in the same region which join the programme at least three years after the treatment cohort in question (results using alternative matching strategies are available from the authors). The bottom section of Table 2 shows the number of potential control schools for each cohort of new Teach First schools. For example, 40 schools participated in Teach First for the first time in 2003/4, all in London. There are 187 schools participating for the first time in 2006/7 or later who could act as potential control schools, but only 90 of these are in London and we restrict our match to these schools.

**Table 2: Potential and actual matched control schools (by year of participation)**

	Treatment cohort year						
	2004	2005	2006	2007	2008	2009	2010
<b>Year of first participation</b>							
<b>Cohort 2003/04</b>	<b>40</b>	0	0	0	0	0	0
<b>Cohort 2004/05</b>	0	<b>11</b>	0	0	0	0	0
<b>Cohort 2005/06</b>	0	0	<b>11</b>	0	0	0	0
<b>Cohort 2006/07</b>	4	0	0	<b>25</b>	0	0	0
<b>Cohort 2007/08</b>	2	3	0	0	<b>26</b>	0	0
<b>Cohort 2008/09</b>	8	2	1	0	0	<b>31</b>	0
<b>Cohort 2009/10</b>	5	1	1	7	0	0	<b>24</b>
<b>Cohort 2010/11</b>	1	0	0	4	7	0	0
<b>Cohort 2011/12</b>	7	2	5	9	8	21	0
<b>Cohort 2012/13</b>	13	3	4	5	11	10	24
<b>No. treatment schools with no match</b>	3	8	3	3	7	3	38
<b>No. potential controls available</b>	387	359	326	292	230	180	86
<i>Of which:</i>							
<b>London</b>	90	80	70	57	49	44	23
<b>North West</b>	-	-	-	44	33	23	10
<b>West Midlands</b>	-	-	-	-	41	28	14
<b>East Midlands and Yorkshire</b>	-	-	-	-	-	-	20

We implement propensity score matching in Stata using `psmatch2` (Leuven & Sianesi, 2003) to deal with the dimensionality of matching on multiple variables, instead matching on a single propensity score which represents the likelihood of a school having been included in the treatment group conditional upon its being selected for treatment (Rosenbaum & Rubin, 1983). Sixteen matches with replacement are actually performed—one probit regression for every cohort–region set of participating schools, since the group of potential control schools changes each time. We apply the nearest- neighbour method with a calliper of 0.2 and the imposition of common support to avoid very poor matches contributing to the calculation of the average treatment effect.

Conditional independence requires the propensity score to capture all variables that correlate with the outcome and programme participation. The following 2003 characteristics of the schools are chosen on the basis that they either (1) formally determine participation eligibility in Teach First (these were free school meals proportion greater than 30% and percentage achieving five good GCSEs less than 25%) or (2) are statistically important in determining both participation and attainment:

- school average prior attainment of pupils at age 11 (i.e. mean Key Stage 2 score);

- average deprivation of pupils' small area neighbourhood (i.e. IDACI);
- proportion eligible for free school meals;
- proportion of white British ethnicity;
- proportion achieving five or more good GCSEs including English and maths in 2003; and
- change in GCSE results between 2000 and 2003.

The last matching covariate is particularly important, since it aims to capture any underlying changes that are taking place at the school during the period of adoption of Teach First. For example, given that headteachers choose to participate, the treatment may be correlated with improvements in performance at the school prior to the programme. Alternatively, in the spirit of an Ashenfelter dip (Ashenfelter, 1978), adoption of the programme may reflect increasing difficulties in recruiting good teachers due to (and exacerbating) declining exam performance. Thus, the matching strategy does deal with changes based on unobservable factors, but only if these factors were already present and captured in the change in exam score variable prior to the treatment.

We report balancing tests in Table 3 on a wide variety of covariates at  $t-1$ . The match is very strong—none of the differences are statistically significant at the 5% level; the one-year change in the best eight subjects' GCSE z-score is significantly different at the 10% level. Matching within region is critical to achieving this strong match, particularly on ethnicity characteristics; restricting the match to only future Teach First schools seems less important in terms of balancing background characteristics, but is important for other reasons as discussed earlier.

**Table 3: Balancing tests (year before treatment)**

	Number of schools	Average capped GCSE z-score	1 Year prior change in GCSE score	Average KS2 score	1 Year prior change in KS2 score	Average IDACI deprivation score	% Free school meals	% White British ethnicity
<b>Treatment group</b>	168	-0.297	0.048	-0.319	0.015	0.398	0.329	0.472
<b>Control (future TF schools)</b>	168	-0.260	0.024	-0.309	-0.008	0.400	0.316	0.495
<b>Difference</b>		-0.037	0.023	-0.009	0.024	-0.001	0.013	-0.023
<b>(Standard error)</b>		(0.032)	(0.016)	(0.028)	(0.017)	(0.011)	(0.016)	(0.035)

Notes: The 168 control schools include multiple counts of schools drawn more than once in the propensity score matching (16 drawn twice; 9 drawn three times; 1 drawn four times; 2 drawn six times).

It is convenient for us to note that, whilst Teach First schools are markedly different from typical English secondary schools, their characteristics have neither improved nor declined as the programme has expanded. This is perhaps not surprising given eligibility restrictions and regional roll-out, but it aids any generalisations we might want to make from estimates using the first seven cohorts.

### Difference-in-difference

We exploit our longitudinal data to combine matching methods with a difference-in-difference estimator (see Heckman et al., 1997). This estimator removes any variation in unobserved time-invariant characteristics between treatment and control observations and, as such, provides a more reliable estimate of the effect of Teach First under the following circumstances: there are common time effects between the treatment and control groups (captured by  $time_i$  in the equation below); the outcome is independent of assignment to treatment; and there are no unmeasured composition changes that occur over time in either group (measured composition changes are captured by  $X_{it}$ ).

We assemble a balanced panel of five observations per treatment and matched control school with two observations prior to first take-up of Teach First and three observations following take-up. The school fixed-effects regression equation for such panel analysis is given as:

$$Y_{jt} = \beta_1 TF1_j + \beta_2 TF2_j + \beta_3 TF3_j + \mathbf{X}_{jt} \cdot \beta_4 + \mathbf{time}_t \cdot \beta_5 + u_j + \varepsilon_{jt}$$

Thus,  $\beta_1$  is the mean outcome for treated schools in the first year of TF participation;  $\beta_2$  is the effect of second-year TF participation; and  $\beta_3$  is the effect of third-year TF participation. We do not extend the post-treatment period further, because: (1) data availability would severely restrict sample size; (2) the nature of the school's involvement in Teach First becomes less clear, with just under half of participants leaving their placement school after two years, and more after 3 years; and (3) as the Teach for America literature points out (see below), it would be hard theoretically to justify a teacher's impact after three years being solely or mainly a result of their having been recruited through Teach First and not a product of internal school training and other professional development.

### Estimating departmental impacts

We can apply identical methods to those described above to estimate the impact of Teach First on maths, science and English attainment, replacing the school-wide participation measure with an indicator for the first participation of the department. As above, these estimates are only valid if there are no time-varying unobservable characteristics associated with the decision to join the Teach First programme.

Going further, if we assume that a department's Teach First participation does not spill over into improvements elsewhere in the school, we can implement a triple-difference estimation approach, using changes in other departments as an additional control, thus holding constant pupil characteristics and school-wide policies:

$$Y_{hjt} = \beta_1 TF1_{hjt} + \beta_2 TF2_{hjt} + \beta_3 TF3_{hjt} + \mathbf{X}_{jt} * \mathbf{subject}_h * \mathbf{time}_t \cdot \beta_4 + \mathbf{time}_t * \mathbf{subject}_h \cdot \beta_5 + \mathbf{Subject}_h \cdot \beta_6 + u_j + \varepsilon_{jt}$$

Here,  $\beta_1$ ,  $\beta_2$  and  $\beta_3$  represent the impact on subject-specific test scores in years 1, 2 and 3, respectively;  $\beta_4$ ,  $\beta_5$  and  $\beta_6$  reflect subject–time trends in average performance and subject–time specific impacts of observed background characteristics. This approach effectively deals with any non-random assignment of the Teach First programme to schools, though clearly does not account for non-random assignment to departments within schools.

Alternatively, we can use annual cross-sections of pupil-level data to estimate point-in-time pupil fixed-effect models that associate a pupil's attainment in a subject with the department's Teach First participation, applying a pupil fixed effect ( $u_i$ ) to account for the pupil's attainment in other subjects.  $\beta_3$ ,  $\beta_4$  and  $\beta_5$  reflect subject–cohort differences in average performance and subject–cohort specific impacts of observed background characteristics:

$$Y_{hij} = \beta_1 TF_{hj} + \beta_2 \mathbf{prior} Y_{hij} + \mathbf{X}_{jt} * \mathbf{subject}_h \cdot \beta_3 + \mathbf{cohort}_{ij} * \mathbf{subject}_h \cdot \beta_4 + \mathbf{subject}_h \cdot \beta_5 + v_{ij} + \varepsilon_{hij}$$

Table 4 shows the variation between departments that is exploited to estimate the impact of Teach First placement within a department. We restrict our analysis here to the core departments of English, maths and science, as these are the only subject areas with participant sample sizes large enough to offer meaningful evaluation. The data show that if a school chooses to participate in Teach First, then each of the three core departments usually do so at some point in the future. However, it is unusual for all three departments to participate together in the first year that a school takes Teach First participants (less than half do) and there are instances of core departments not yet having participated.

We have no reason to necessarily expect the estimates of impact from these three departmental estimation approaches to be the same. Any differences in coefficient sizes might tell us something about

the relative importance of: (1) sorting into schools and into departments on unobservables; (2) the size of spillovers from departments across schools; and (3) variation in effect sizes across different subjects.

**Table 4: Timing of first participation across departments within schools**

	English				Maths				Science			
<b>Dept participates with:</b>												
<b>0-year lag</b>				81				76				76
<b>1-year lag</b>				31				27				32
<b>2-year lag</b>				13				20				15
<b>3-year lag</b>				9				5				10
<b>4-year lag</b>				1				3				0
<b>5-year lag</b>				3				1				2
<b>6-year lag</b>				2				2				0
<b>7-year lag</b>				0				0				2
<b>8-year lag</b>				1				2				0
<b>9-year lag</b>				1				0				0
<b>Not TF yet</b>				26				32				31
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
<b>No. TF participants:</b>												
<b>Year 1</b>	0.6	0.7	0	3	0.6	0.7	0	3	0.6	0.7	0	3
<b>Year 2</b>	1.0	1.0	0	5	0.8	0.8	0	3	1.0	0.9	0	3
<b>Year 3</b>	0.9	1.1	0	5	0.6	0.7	0	3	0.8	0.9	0	3

## Results

We first present the results from the estimation of the impact of Teach First participation on whole-school achievement, before moving on to departmental estimates. The impact of whole-school achievement is likely to reflect a combination of the relative effectiveness of the Teach First participant themselves, the impact of the participant on the teaching quality of others in their department and elsewhere in the school, and any wider contribution the Teach First participants make to the life of the school.

In Table 5 we present results from six regression equations. These are all difference-in-difference regressions estimated in a balanced panel of five observations per school (two prior and three following first Teach First participation). We show estimates from the full sample of schools, followed by the sample restricted to the matched control group of future Teach First schools located in the same region. The estimates using the full sample of schools are somewhat larger, which is logical since the type of more disadvantaged schools participating in Teach First were rapidly improving over the period of study. We report the results on the matched sample both with and without time-varying control variables. These control variables do not change the substantive estimates, which is correct since we have no reason to believe that the time-varying controls are correlated with Teach First participation. However, they may marginally improve precision on estimates and so we do include them for all other tables in this paper.



**Table 5: Difference-in-difference regression results**

	Full sample of schools			Matched sample of schools			Matched sample of schools		
	Beta	SE	Sig	Beta	SE	Sig	Beta	SE	Sig
<b>Best 8 GCSE z-score:</b>									
Treatment year 1	0.128	(0.009)	***	0.027	(0.019)	n.s.	0.019	(0.019)	n.s.
Treatment year 2	0.160	(0.010)	***	0.058	(0.019)	***	0.048	(0.019)	***
Treatment year 3	0.195	(0.195)	***	0.082	(0.019)	***	0.081	(0.019)	***
Rho (fraction var. due to $u_i$ )		0.765			0.758			0.706	
R-sq (overall)		0.759			0.018			0.294	
<b>5+A*-C incl Eng and maths:</b>									
Treatment year 1	0.039	(0.003)	***	-0.001	(0.008)	n.s.	-0.004	(0.008)	n.s.
Treatment year 2	0.060	(0.003)	***	0.019	(0.008)	**	0.015	(0.008)	**
Treatment year 3	0.074	(0.004)	***	0.021	(0.008)	***	0.020	(0.008)	***
Rho (fraction var. due to $u_i$ )		0.771			0.788			0.746	
R-sq (overall)		0.776			0.129			0.341	
<b>Controls</b>		<b>Yes</b>			<b>No</b>			<b>Yes</b>	
<b>N (overall)</b>		<b>33,720</b>			<b>1680</b>			<b>1680</b>	
<b>N (schools)</b>		<b>3,300</b>			<b>336</b>			<b>336</b>	

Notes: n.s. = not statistically significant, \*\*\*sig at 1% level, \*\*sig at 5% level, \*sig at 10% level

School and time dummy variables included, time-variant controls are mean KS2 prior attainment and proportions of free school meals, English as an additional language, white British ethnicity pupils.

Difference-in-difference on full school samples also includes dummies for each post-treatment (up to 9).

The results show that Teach First participation has no impact on a school’s exam performance in year 1, as measured by pupils’ best eight subject grades (i.e. capped GCSE z-score). The impact in years 2 and 3 is positive and statistically significant at around 5% and 8% of a standard deviation. This pattern of no effect in year 1 and positive and increasing effects in years 2 and 3 is found throughout our results section.

An effect size of 5% of a pupil standard deviation is equivalent to the school moving up 10% of a standard deviation across the distribution of school average capped GCSE z-scores, or a little less than one grade in one of the child’s best eight subjects. This may sound relatively small (Hattie, 2003), but it does translate into 1.5 to 2 months of extra learning, which is substantial given that Teach First simply places a small number of inexperienced teachers in a large secondary school. Since there are multiple and complex possible pathways to impact for Teach First participants, we will reserve more detail on the magnitude of likely mechanisms for our estimation of departmental impacts.

The finding of no effect in the first year may occur for several reasons. These year- 11 pupils would have had little direct exposure to the Teach First participants, who would only have been in school for one of the pupils’ five years. They would have lower chances of having been allocated a new Teach First teacher at the start of year 11, given standard practice in England of retaining the same teacher across years 10 and 11 where possible. Also, clearly as novice teachers in the first year, the Teach First participants will be less effective than in their second year. It is not possible to distinguish between these potential mechanisms in the data.

The impact of Teach First is less precise on the threshold outcome measure of the proportion of pupils attaining five or more good GCSEs, including English and maths. In the second and third years, where point estimates are positive, they are in the region of just two percentage points (on a metric with an average of around 30%). This is a far smaller effect size than that reported in Muijs et al. (2010).

In Table 6 we explore whether the impact of Teach First varies across time and across regions. We are relieved to find that the impact of the programme is not bigger within London than it is outside London, because this was a period of considerable change within the capital city and so there was a serious risk that estimates were inflated by Teach First schools receiving other interventions at the same time. The

impact of the programme for schools first joining in the later years of 2008/09 to 2010/11 has not shrunk compared to earlier years; indeed, estimates are slightly larger. We had expected to find that the impact of the programme shrinks as it expands, because the quality of the marginal participant should decline and the programme itself may experience scale diseconomies. The finding that this is not the case does not mean it will not face expansion difficulties in the future, since the largest cohort we estimate impacts on in this paper is just over 500, compared to about 1000 in 2012/13. The finding of smaller effect sizes in earlier years (the same years for which Muijs et al. find large effects) is not due to lower intensity of participation, nor can it be explained by particularly different school characteristics.

**Table 6: Testing for heterogeneity across regions and over time**

	Main specification			London only		
	Beta	SE	Sig	Beta	SE	Sig
Treatment year 1	0.019	(0.019)	n.s.	0.025	(0.023)	n.s.
Treatment year 2	0.048	(0.019)	***	0.050	(0.023)	**
Treatment year 3	0.081	(0.019)	***	0.087	(0.023)	***
Rho (% var due to $u_i$ )		0.706			0.739	
R-sq (overall)		0.294			0.441	
Controls		Yes			Yes	
N (overall)		1680			950	
N (schools)		336			190	
<i>Average characteristics of sample in year prior to treatment:</i>						
Capped GCSE score		0.279			-0.261	
IDACI depr- score		0.399			0.398	
% White British ethnicity		0.484			0.323	
	2007 and earlier			2008 onwards		
	Beta	SE	Sig	Beta	SE	Sig
Treatment year 1	0.005	(0.021)	n.s.	0.032	(0.032)	n.s.
Treatment year 2	0.005	(0.021)	n.s.	0.095	(0.023)	***
Treatment year 3	0.049	(0.021)	**	0.115	(0.032)	***
Rho (% var due to $u_i$ )		0.792			0.629	
R-sq (overall)		0.516			0.091	
Controls		Yes			Yes	
N (overall)		870			810	
N (schools)		174			162	
<i>Average characteristics of sample in year prior to treatment:</i>						
Capped GCSE score		-0.288			-0.269	
IDACI depr- score		0.395			0.404	
% White British ethnicity		0.242			0.547	

Notes: n.s. = not statistically significant, \*\*\*sig at 1% level, \*\*sig at 5% level, \*sig at 10% level.

School and time dummy variables included, time-variant controls are mean KS2 prior attainment, free school meal proportion, English as an additional language proportion, white British ethnicity proportion.

In Table 7 we turn to estimates of the departmental impact of Teach First participants. There are three estimation strategies here, so the findings are rather complex and make different assumptions regarding likely pathways to impact. The first three columns of estimates are from difference-in-difference regressions of changes in the effectiveness of English, maths and science departments in a school separately, without holding constant any changes taking place in other departments in the school. The advantage of these regressions is that they allow for cross-departmental spillovers in impact, with the corresponding disadvantage that any unobserved changes in overall school processes cannot be accounted for. The estimates here are not consistent across subjects: the impact is strongest and most precisely estimated in English, a subject where Teach First themselves will claim they find it easiest to recruit high-quality participants (by contrast, some Teach First participants in maths and science do not have a degree in these subjects). No impact is ever found in maths, though the point estimates are positive.

**Table 7: Departmental impact of Teach First participants**

	Difference-in-difference								
	English grade			Maths grade			Science grade		
	Beta	SE	Sig	Beta	SE	Sig	Beta	SE	Sig
<b>Pre-treatment</b>									
<b>Year 1</b>	-0.014	(0.410)	n.s.	-0.003	(0.038)	n.s.	-0.013	(0.085)	n.s.
<b>Year 2</b>	0.057	(0.038)	n.s.	0.047	(0.035)	n.s.	0.093	(0.077)	n.s.
<b>Year 3</b>	0.128	(0.037)	***	0.045	(0.033)	n.s.	0.138	(0.075)	*
<b>Rho</b>		0.736			0.757			0.596	
<b>R-sq</b>		0.295			0.336			0.274	
<b>Controls</b>		Yes			Yes			Yes	
<b>N (overall)</b>		1680			1680			1680	
<b>N (schools)</b>		336			336			336	
	Triple difference			Pupil fixed effects					
	GCSE grade			GCSE score					
	Beta	SE	Sig	Beta	SE	Sig	Rho		
<b>Pre-treatment</b>				-0.086	(0.007)	***	0.765		
<b>Year 1</b>	-0.008	(0.034)	n.s.	-0.073	(0.008)	***	0.724		
<b>Year 2</b>	0.077	(0.031)	**	0.159	(0.009)	***	0.698		
<b>Year 3</b>	0.110	(0.030)	***	0.146	(0.011)	***	0.682		
<b>Rho</b>		0.540							
<b>R-sq</b>		0.410							
<b>Controls</b>		Yes			Yes				
<b>N (overall)</b>		5040			186,670				
<b>N (schools)</b>		336			336				

Notes: n.s. = not statistically significant, \*\*\*sig. at 1% level, \*\*sig. at 5% level, \*sig. at 10% level.

a School and time dummy variables included, time-variant controls are mean KS2 prior attainment, % free school meal, % English as an additional language, % white British ethnicity.

b School, time, subject and time-subject dummy variables included, within-school subject differences at t-2 included, time-variant controls are mean KS2 prior attainment, % free school meal, % English as an additional language, % white British ethnicity.

c Pupil, subject and subject-year dummy variables included, subject-variant controls are pupil prior attainment in subjects, sex and English as an additional language status. Here we only report the year 2 impact of intensity in the pupil fixed-effects model.

The triple-difference estimates in column four measure changes in departmental effectiveness, holding constant changes taking place in core departments which have not yet taken on a Teach First participant. It imposes an assumption of equal potential impact across subjects and will be seriously biased downwards if, for example, Teach First participants in English departments are able to positively influence a child's maths GCSE grade. The impact estimates are zero in year 1 and positive in years 2 and 3, in the order of about 8% and 11% of a grade, respectively.

The pupil point-in-time fixed-effects models take a cross-section of data separately for one year prior and one, two and three years following the school's first Teach First participation and estimate the impact of departmental participation in pupil-by-subject data. The pre-treatment estimate shows relative effectiveness in the year before participation for departments which take on Teach First participants in the first year that the school first participates, holding constant the effectiveness of those departments which do not take on a participant in the first year. It shows that the departments who participate early on are significantly less effective before the arrival of the Teach First participants than those who do not immediately participate. They may have staff recruitment difficulties or higher teacher turnover, which presents vacancy opportunities, or they might be viewed as struggling by the headteacher who therefore encourages them to try Teach First. These early-participating departments are also significantly less effective in year 1, but more effective in years 2 and 3, to the tune of 15% of a subject grade. Clearly, the year 3 estimates only reflect differences between the early-participating departments and others in

schools where one or more core departments do not participate by year 3; in 22% of treatment schools, all three core departments have participated at least once by year 3.

We run interactions (not reported here) with pupil demographic variables to see whether some pupils benefit more than others from the treatment. These show some- what larger impacts (e.g. 5% of a standard deviation in year 2) for those eligible for free school meals, compared to others. Those with the lowest prior attainment also experience larger gains; the impact on high-prior-attainment pupils is no different from those in the middle of the distribution. Of course, without knowing which ability sets the Teach First participants were allocated to, it is hard to interpret the significance of these demographic interactions.

Within a core subject department in a participating school, about one in six teachers will be a Teach First participant in each of the first three years. Our estimate of impact of the order of at least 5% of a subject grade could be as high as 30% of a grade if we assume no spillovers of participation to other teachers in the same department. This implies that Teach First participants are highly effective, on average, compared to those they have displaced. Using estimates from Slater et al. (2012), 30% of a subject grade is equivalent to one standard deviation in higher teacher effectiveness. It is possible that Teach First selection and training processes are this effective, but more likely there are some spillovers to other teachers in the department.

## Discussion

---

In this paper we provide convincing evidence that placing carefully selected, yet inexperienced, graduates into English secondary schools has not been damaging to pupils and most likely produced school-wide gains in GCSE results of the order of 5% of a pupil standard deviation, or around one grade in one of the pupil's best eight subjects. This is a consistently estimated positive effect, though clearly not as large as the impact of other interventions to improve teaching standards—such as training to improve the quality of pupil feedback (Hattie, 2003). It is also substantially smaller than the estimate of one-third of a grade per subject found by Muijs et al. (2010). The claims we make about the school-wide impact of Teach First assume its adoption is not confounded by the arrival of dynamic management and structural changes at the school, and although the matched difference-in-difference approach reduces the chances that this biases estimates, we cannot dismiss this endogeneity problem entirely.

Within core departments, our estimates are not subject to the same endogeneity problems and suggest a gain of over 5% of a subject grade, which could translate into as high as 30% of a grade in the Teach First participant's classroom if we assume no spillover to other teachers. If there were no spillovers of Teach First participation to other teachers, this would suggest that the Teach First selection process succeeds in attracting and selecting good teachers who are, on average, one standard deviation more effective than those they displace. This is perfectly possible, not least because they likely displace many temporary or non-subject-specialist teachers, but it seems more likely that Teach First presence also raises the teaching standards of those who teach alongside them in the same department. This echoes the findings of Jackson and Bruegmann (2009), who identify the importance of teacher peers and the work of Hurd (2008), who describes how the presence of trainees in schools can promote professional learning in other teachers.

Our findings are best interpreted alongside some understanding of who participating schools might have recruited in the absence of Teach First. An examination of the first available years of the School Workforce Census (2010/11 and 2011/12) shows that a school's participation in Teach First increases the number of young teachers (under age 30) in the school (from around 22% of all staff to around 27% of all staff), suggesting that the scheme is not used solely to replace other newly qualified graduates. We can also observe that schools participating in Teach First have greater numbers of teachers who have taught

at the school for less than 5 years (as many as 60% with tenure less than 5 years, versus 50% for comparison schools), though not larger numbers of new arrivals once the scheme is established in their school. This School Workforce Census data suggest that schools are displacing slightly older and more experienced teachers who might have spent longer working at the school, had they been recruited.

If it is true that Teach First participants displace more experienced teachers, it is not surprising that many believed the programme could be damaging. Participants have excellent academic credentials, but there is little evidence that this will make them better teachers (e.g. Hanushek et al., 2005; Aaronson et al., 2007). In general, it is true that teachers are less effective at the start of their career (e.g. Rockoff, 2004 estimates that teachers improve by around 10% of a standard deviation in the first two years in maths). However, recent academic papers from the USA explain that getting this initial selection of teachers correct is critical because, whilst a first-year novice teacher is less effective than they will be in their second year, the improvement in teaching quality gained through experience is actually relatively modest compared to the very wide variation in teacher quality at the outset. Furthermore, those who are weak teachers in the first year improve their practice at a slower rate than others, thus widening gaps in effectiveness in the second, third and fourth years (TNTP, 2013; Atteberry et al., 2015). That said, we must recognise the disruption and recruitment costs to schools of dealing with the higher teacher turnover that the Teach First programme necessarily produces. This turnover also places limits on how large the scheme should become for any individual school.

The research design used here makes no claims about the relative cost-effectiveness of Teach First participants versus those trained via other routes, nor can it assess the quality of the short Teach First summer training programme which is undertaken alongside ongoing university and in-school support for participants. Indeed, even in North America there is no clear evidence that certain approaches in teacher education are more effective than others (McConney et al., 2012).

We also do not examine what happens to the long-run effectiveness of those schools which participated in the early years of the programme. Teacher effects are often not sustained (Jacob et al., 2010) and future research could now find out whether these early-participating schools continue to do well, either through recruitment of further Teach First participants, through retention of existing ones or through spillover effects to other staff.

Our estimates relate to a period of time when the Teach First programme was a fraction of the size it is now, or aspires to be in the future. We can say nothing about whether its effectiveness will fall as it expands in the type of graduates it recruits and the type of schools within which it places. With well over 1000 participants a year from a much more diverse range of undergraduate universities, it is likely that they are now recruiting many participants who would otherwise have joined the teaching profession through the traditional university-led graduate route. Whether or not this reduces the value of Teach First depends on whether its impact arises more from the recruitment of talented graduates than from the efficacy of their very short, intensive training programme and subsequent in-school and partner university support.

Overall, this study lends strong support to studies from the USA regarding the effectiveness of these types of teacher recruitment programme, particularly where graduates are placed in classrooms with older students. The growth of similar programmes in a number of other countries affiliated to the international 'Teach for All' umbrella organisation, created in cooperation between Teach First and Teach for America in 2007, means that our conclusions are relevant beyond Teach for America and Teach First. However, the lack of matched teacher-pupil data means we can say little about individual participant effectiveness or about how participants influence the teaching experiences of others in the school. Understanding precise mechanisms of impact is important, because Teach First itself cannot expand indefinitely. If the founder's claim that Teach First has 'detoxified teaching' for high-attaining graduates (Wigdortz, 2012, p. 230) is true, then maintenance of its position as a premium brand is somewhat

contingent on retaining exclusivity as a minor training route. However, understanding exactly what Teach First does that makes it an effective programme may enable us to replicate small parts of the behaviour of the participants across the education system.

## References

---

- Aaronson, D., Barrow, L. & Sander, W. (2007) Teachers and student achievement in the Chicago public high schools, *Journal of Labour Economics*, 25(1), 95–135.
- Allen, R., Burgess, S. & Mayo, J. (2012) *The teacher labour market, teacher turnover and disadvantaged schools: New evidence for England*. CMPO Working Paper No. 12/294 and DoQSS Working Paper No. 12/09.
- Allen, R., Parameshwaran, M. & Nye, P. (2016a) *The careers of Teach First ambassadors who remain in teaching: Job choices, promotion and school quality*. Report to Teach First, available online at: [educationdatalab.org.uk/wp-content/uploads/2016/07/The-careers-of-Teach-First-Ambassadors-who-remain-in-teaching-FINAL.pdf](http://educationdatalab.org.uk/wp-content/uploads/2016/07/The-careers-of-Teach-First-Ambassadors-who-remain-in-teaching-FINAL.pdf).
- Allen, R., Belfield, C., Greaves, E., Sharp, C. & Walker, M. (2016b) *The longer-term costs and benefits of different initial teacher training routes*. IFS report, available online at: [www.ifs.org.uk/uploads/publications/comms/R118.pdf](http://www.ifs.org.uk/uploads/publications/comms/R118.pdf).
- Antecol, H., Eren, O. & Ozbeklik, S. (2013) The effect of Teach for America on the distribution of student achievement in primary school: Evidence from a randomized experiment, *Economics of Education Review*, 37(Dec), 113–125.
- Ashenfelter, O. (1978) Estimating the effects of training programmes on earnings, *Review of Economics and Statistics*, 60(1), 47–57.
- Atteberry, A., Loeb, S. & Wyckoff, J. (2015) Do first impressions matter? *Predicting early career teacher effectiveness*. AERA Open, doi: 10.1177/2332858415607834.
- Blanden, J., Greaves, E., Gregg, P., Macmillan, L. & Sibieta, L. (2015) *Understanding the improved performance of disadvantaged pupils in London*. CASE – Social Policy in a Cold Climate Working Paper 21 (London, LSE).
- Boyd, D.J., Grossman, P., Lankford, H., Loeb, S., Michelli, N.M. & Wyckoff, J. (2006) How changes in entry requirements alter the teacher workforce and affect student achievement, *Education Finance and Policy*, 1(2), 176–216.
- Darling-Hammond, L., Holtzman, D.J., Gatlin, S.J. & Heilig, J.V. (2005) Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness, *Education Policy Analysis Archives*, 13(42), 1068–2341.
- Decker, P.T., Mayer, D.P. & Glazerman, S. (2004) The effects of Teach for America on students: Findings from a national evaluation, *Journal of Policy Analysis and Management*, 25 (1), 75–96.
- Goldhaber, D. & Brewer, D. (2000) Does teacher certification matter? High school teacher certification status and student achievement, *Educational Evaluation and Policy Analysis*, 22(2), 129–145.
- Hanushek, E. A., Kain, J. F., O'Brien, D. M. & Rivkin, S. G. (2005) *The market for teacher quality*. NBER Working Paper No. 11154.
- Hattie, J. (2003) *Teachers make a difference: What is the research evidence?* Melbourne: Australian Council for Educational Research.
- Heckman, J.J., Ichimura, H. & Todd, P. (1997) Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme, *Review of Economic Studies*, 64(4), 605–654.

- Henry, G. T., Thompson, T. L., Fortner, C. K., Zulli, R. A. & Kershaw, D. C. (2010) *The impact of teacher preparation on student learning in North Carolina public school*. Carolina Institute of Public Policy, University of North Carolina. Available online at: [www.publicpolicy.unc.edu](http://www.publicpolicy.unc.edu) (accessed 24 June 2013).
- Hurd, S. (2008) Does school-based initial teacher training affect secondary school performance?, *British Educational Research Journal*, 34(1), 19–36.
- Hutchings, M., Maylor, U., Mendick, H., Menter, I. & Smart, S. (2006) *An evaluation of innovative approaches to teacher training on the Teach First programme*. Final report to the Training and Development Agency for Schools (London, Teaching and Development Agency).
- Jackson, C.K. & Bruegmann, E. (2009) Teaching students and teaching each other: The importance of peer learning for teachers, *American Economic Journal: Applied Economics*, 1(4), 85–108.
- Jacob, B.A., Lefgren, L. & Sims, D.P. (2010) The persistence of teacher-induced learning, *Journal of Human Resources*, 45(4), 915–943.
- Kane, T.J., Rockoff, J.E. & Staiger, D.O. (2008) What does certification tell us about teacher effectiveness?, Evidence from New York City, *Economics of Education Review*, 27(6), 615–631.
- Lack, A. & Johnston, A. (2008) *Schools facing challenging circumstances: Quantitative findings* (London, Teaching and Development Agency). Available online at: [www.shu.ac.uk/\\_assets/pdf/cer-TDA-NQT-Part2.pdf](http://www.shu.ac.uk/_assets/pdf/cer-TDA-NQT-Part2.pdf) (accessed 24 June 2013).
- Leuven, E. & Sianesi, B. (2003) *PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing*. Available online at: [ideas.repec.org/c/boc/bocode/s432001.html](http://ideas.repec.org/c/boc/bocode/s432001.html) (version 3.1.5, May 2009).
- McConney, A., Woods-McConney, A. & Price, A. (2012) *Fast track teacher education: A review of research literature on Teach for All schemes* (Perth, Murdoch University, Centre for Learning, Change and Development).
- Muijs, D., Chapman, C., Collins, A. & Armstrong, P. (2010) *Maximum impact evaluation. The impact of Teach First teachers in schools: An evaluation* (London, Teach First). Available online at: [www.teachforall.org/articles/max\\_impact.pdf](http://www.teachforall.org/articles/max_impact.pdf) (accessed 24 June 2013).
- Ofsted (2010) *London Challenge* (London, Ofsted).
- Ofsted (2011) *Teach First: Initial teacher education inspection report* (London, Ofsted).
- Raymond, M., Fletcher, S. H. & Luque, J. (2001) *Teach for America: An evaluation of teacher differences and student outcomes in Houston, Texas* (Palo Alto, CA, Center for Research on Education Outcomes).
- Rockoff, J.E. (2004) The impact of individual teachers on student achievement: Evidence from panel data, *American Economic Review*, 94(2), 247–252.
- Rosenbaum, P.R. & Rubin, D.B. (1983) The central role of the propensity score in observational studies for causal effects, *Biometrika*, 70(1), 41–55.
- Schoeneberger, J. A. (2011) *Evaluation of Teach for America in Charlotte-Mecklenburg Schools* (Center for Research and Evaluation, Office of Accountability, Charlotte-Mecklenburg Schools). Available online at: [www.cms.k12.nc.us/cmsdepartments/accountability/cfre/Documents/TeachForAmerica\\_Evaluation\\_Report\\_2011.pdf](http://www.cms.k12.nc.us/cmsdepartments/accountability/cfre/Documents/TeachForAmerica_Evaluation_Report_2011.pdf) (accessed 24 June 2013).
- Schoeneberger, J. A., Dever, K. A. & Tingle, L. (2009) *Evaluation of Teach for America in Charlotte-Mecklenburg Schools* (Center for Research and Evaluation, Office of Accountability, Charlotte-Mecklenburg Schools). Available online at: [www.cms.k12.nc.us/cmsdepartments/accountability/cfre/Documents/TFA\\_Evaluation\\_Report.pdf](http://www.cms.k12.nc.us/cmsdepartments/accountability/cfre/Documents/TFA_Evaluation_Report.pdf) (accessed 24 June 2013).
- Slater, H., Davies, N.M. & Burgess, S. (2012) Do teachers matter? Measuring the variation in teacher effectiveness in England, *Oxford Bulletin of Economics and Statistics*, 74(5), 629–645.

TNTP (2013) *Leap year: Assessing and supporting effective first-year teachers* (Brooklyn, NY, TNTP). Available online at: [tntp.org/assets/documents/TNTP\\_LeapYear\\_2013.pdf](http://tntp.org/assets/documents/TNTP_LeapYear_2013.pdf) (accessed 24 June 2013).

Ware, A., LaTurner, R. J., Parsons, J., Okulicz-Kozaryn, A., Garland, M. & Klopfenstein, K. (2011) *Teacher preparation programs and Teach for America research study* (Dallas, TX, University of Texas at Dallas, Education Research Center). Available online at: [www.teachforamerica.org/our-organization/research](http://www.teachforamerica.org/our-organization/research) (accessed 24 June 2013).

Wigdortz, B. (2012) *Success against the odds. Five lessons in how to achieve the impossible: The story of Teach First* (London, Short Books).

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