Essays in the Economics of Education: School Resource Decisions and Student Achievement

Luke Sibieta

Thesis submitted to the Department of Social Science, Institute of Education, in partial fulfilment of the requirements for the degree of:

Doctor of Philosophy

University College London

2020

Declaration

I, Luke Sibieta, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature:

Date: August 19, 2020

Abstract

This thesis contains three papers on school resource decisions and the implications for student achievement. Chapter 1 provides an introduction and describes how the three papers extend the academic literature.

Chapter 2 studies the implications of an individual school offering higher teacher salaries from within a fixed budget. It exploits a natural experiment that forces some schools within a local area to pay teachers according to higher salary scales, but does not offer any extra funding. Schools largely follow this regulation and balance their budgets by making sizable reductions in other expenditures. There is no evidence of any overall effect on pupil attainment. The paper argues that the positive effect of higher teacher pay is countered by the negative effects of reductions in other expenditure.

Existing evidence shows how cash incentives can be used to improve teacher retention in hard-to-staff subjects, like maths and science. Chapter 3 extends this literature by studying the effectiveness of incentives to recruit new teachers. I evaluate the effects of an up-front cash payment worth up to £25,000 for teachers training in hard-to-staff subjects and who have high levels of college attainment. Using a triple-difference approach, I find no impact on the number of teachers or the distribution of educational attainment among teachers.

Chapter 4 uses the synthetic control approach to estimate the effects of an area-wide campaign to improve the ways in which teaching assistants are used across a large, disadvantaged area of England. The results suggest the campaign increased English scores by a modest amount of about 0.03-0.04 standard deviations, with no evidence of an improvement in maths. The impact estimates are larger than under matching and difference-in-differences, suggesting that being able to relax assumptions of parallel trends and balance in unobservables represents a major advantage of synthetic control approaches.

Impact Statement

This thesis aims to improve understanding of the effects of school resource decisions. The empirical evidence resulting from this thesis should therefore allow policymakers and individual schools to make better school resource decisions.

Chapter 2 focuses on the potential effects of an individual school increasing teacher salaries from within a fixed budget. This is important to understand as schools across the world are gaining increasing autonomy over teacher pay, but there is no evidence on how changing pay at an individual school could affect student achievement. Chapter 2 shows that the net effect of such a change is close to zero, arguing that any beneficial effects of raising teacher pay are countered by the negative effects of reducing other resources to pay for such a salary rise. This suggest that raising teacher pay at an individual school is unlikely to be a good use of resources.

There are significant challenges in recruiting and retaining enough teachers in maths and science subjects across many countries. Existing evidence has shown that targeted incentive payments can be relatively effective in improving retention of maths and science teachers. However, there is little evidence on their effectiveness to improve recruitment. In Chapter 3, I show that recruitment incentives are relatively ineffective at improving recruitment or improving the skill levels of teachers. This suggest that the high-value teacher training bursaries currently on offer in England (up to £25,000 in some cases) represent relatively poor value-for-money. They would therefore be better directed towards retention incentives.

The number of Teaching Assistants in England has grown massively over the last two decades. However, current empirical evidence suggests they are used in poor ways. In Chapter 4, I show how an area-wide campaign to improve the ways in which Teaching Assistants are used can improve student achievement, particularly in English scores. This strongly suggests that improving the deployment of Teaching Assistants across all schools could yield significant gains in terms of improved student achievement.

Chapter 4 also shows the value of using the Synthetic Control approach to study the effects of area-wide interventions. It shows that this method is relatively simple to implement and can deliver improvements relative to other non-experimental estimators, such as matching or difference-in-differences, through being able to relax assumptions of parallel trends and balance in unobservables

Acknowledgments

This thesis was made possible by various sources of funding and support from the Economic and Social Research Council (ESRC), particularly financial support as part of the UCL IoE Doctoral Training Centre. All funding and support for individual chapters is detailed below.

I am also eternally grateful to my supervisors, Professors John Jerrim and Lorraine Dearden, for their advice, guidance and encouragement throughout my PhD. This thesis would not have been possible without their support. I am also grateful to Professor Imran Rasul for supervision whilst I was a student in the Economics Department. I am particularly grateful for his advice, understanding and his high-standards. I am also grateful to many colleagues at the Institute for Fiscal Studies for their encouragement and advice.

I am also grateful to my family, my wife Elinor Landeg in particular, for their support, patience and encouragement over the years.

Chapter 2 is jointly authored with Ellen Greaves and was published in the Economics of Education Review in 2019. For this chapter, I am grateful to the ESRC for funding the original project (small grants programme: ES/J006076/1; Centre for the Microeconomic Analysis of Public Policy: RES-544-28-0001 & ES/M010147/1). I also thank Rebecca Allen, Jack Britton, Alex Bryson, Steve Gibbons, Carol Propper, Imran Rasul, Nikki Shure and seminar participants at the IFS, IoE, RES, EEA and PEUK for comments.

Chapter 3 is a sole-authored paper. For this, I am grateful to the ESRC for financial support as part of the UCL IoE Doctoral Training Centre and the ESRC Centre for the Micro economic Analysis of Public Policy at IFS (grant number RES-544-28-0001). I am also grateful to the Office for Manpower Economics for funding an earlier piece of research that led to this paper. I am also grateful to Neil Amin-Smith and Ellen Greaves for their contribution to the earlier piece of work, and to Chris Belfield, Jack Britton, Nikki Shure, Alex Bryson and RES conference participants for comments and feedback.

Chapter 4 is a sole-authored paper. I am grateful to the ESRC for financial support as part of the UCL IoE Doctoral Training Centre and the Education Endowment Foundation for providing funding and support. I thank Barbara Sianesi and Guillermo Rodriguez-Guzman for their comments and feedback. This paper is partly based on an independent evaluation commissioned by the Education Endowment Foundation, which has already been published. A pre-specified Analysis plan was also published, which is largely followed. The exceptions are stated clearly, as well as the reasons for deviating from this plan.

Neither the funders nor the data owners accept responsibility for any inferences or conclusions derived in this thesis. All errors, conclusions and interpretation remain my own.

Data provision

This thesis makes use of a number of data sources. Chapters 2 and 4 make use of the National Pupil Database (NPD) for England, which was made available under license by the Department for Education. Chapter 2 also uses the School Workforce Census (SWC), which was made available by the Department for Education. The Department for Education is responsible for the collation and management of the NPD and SWC and is the Data Controller for both sets of data. Any inferences or conclusions derived from the NPD or SWC in this thesis are the responsibility of the author and not the Department for Education.

Chapter 2 also makes use of publicly available school-level data (Section 251 outturn data for school funding levels, the Local Education Authority School Information Service (LEASIS) and its later replacements, and Consistent Financial Reporting data (CFR).

Chapter 3 makes use of three sources of data provided by the Higher Education Statistics Agency (HESA). In particular, the HESA Student record 2006/07–2014/15, the HESA Destinations of Leavers from Higher Education record 2006/07–2014/15 and the HESA Longitudinal Destinations of Leavers from 2006/07, 2008/09 and 2010/11. These data are Copyright Higher Education Statistics Agency Limited. Neither the Higher Education Statistics Agency Limited nor HESA Services Limited can accept responsibility for any inferences or conclusions derived by third parties from data or other information supplied by HESA Services.

Chapter 3 also use the Quarterly Labour Force Survey (LFS), which was made available through the UK Data Service at the University of Essex. The LFS is Crown copyright material is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

All data analysis was conducted at the Institute for Fiscal Studies (IFS), with all data used under license. All ethical processes for IFS researchers were followed.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Keywords: Teacher Wages, School Resources, Student Achievement

JEL Codes: H52, I20, I21, J24, J45

1.1 Introduction

School spending represents a significant component of public spending, representing about 3% of national income, on average, across OECD countries[1]. Such spending can potentially be used as a engine of productivity growth and social mobility if it is able to generate improvements in human capital (Becker and Chiswick (1966); Card and Krueger (1992); Psacharopoulos and Patrinos (2018); Chetty and Hendren (2018)). Historically, evidence suggested that increases in school resources had minimal effects on student achievement (Hanushek (2003)). The latest evidence, however, suggests that increases in school spending can have profound effects on later life outcomes (Jackson et al. (2015); Jackson (2018)), and that cuts can lead to worse outcomes (Jackson et al. (2018)). Combined with cuts to school spending following the Great Recession, such evidence has renewed interest in how schools can use spending efficiently and mitigate the effects of cuts. Schools are also gaining increasing autonomy over budgets, resources and other decisions, with such autonomy seeming to have differential effects across countries (Hanushek et al. (2013)). This underlines the importance of understanding how schools can best use their newly gained autonomy over resource decisions to best improve student achievement.

In this thesis, I analyse the causal effects of a number key resource decisions and the mechanisms underlying them. I examine the effects of differences in teacher pay across individual schools (within fixed budgets), the effect of recruitment incentives on the quantity and quality of recruits, and whether a large scale training programme can improve the productivity of teaching assistants. The context is schools and teachers in England. I focus on England given large swings in school funding and resources over time, increasing levels of autonomy and the availability of key administrative data on pupils and their achievement.

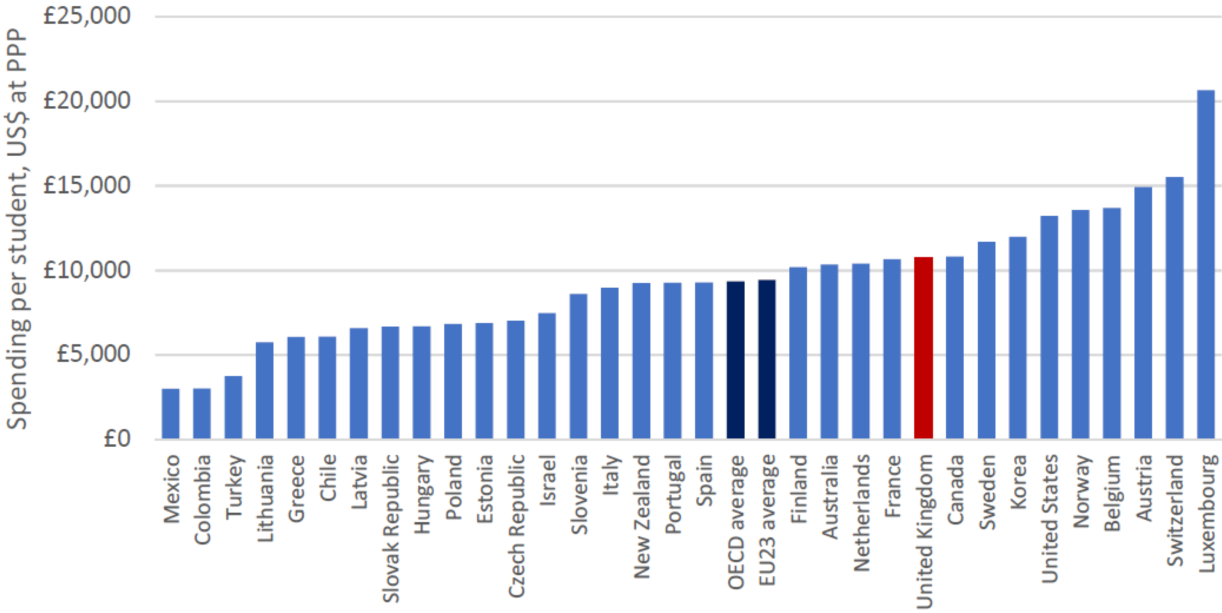
The next section describes key trends in school resources and organisation in England. Section 1.3 presents a general model of school resources and student achievement, which is used to guide a review of the existing evidence on the effects of different resource margins. Section 1.4 reviews the state of the current literature on school resources and expenditure. Section 1.5 further extends the model to consider the mechanisms driving potential effects of teacher pay and incentives, reviews the current state of empirical evidence and details how the first two chapters extend this literature. Section 1.6 augments the model to analyse the potential effects of investments in other resources, teaching assistants in particular, and how the productivity of such investments could be improved.

1.2 Policy Context: School Resources in England

Spending per pupil in the UK is slightly above average for OECD countries. As shown in Figure 1.1, spending per pupil in the UK as a whole was about \$10,700 in 2016 (in purchasing power parity terms), slightly above the OECD and EU averages of just under \$9,500. This places spending per pupil in the UK at a similar level as France (\$10,600) and Canada (\$10,700), below the USA (\$13,200), but above that in other large European countries such as Italy (\$9,000) and Spain (\$9,300)[2]. Whilst these figures relate to the UK as a whole, evidence suggests that spending per pupil in England is close to the UK average, which is unsurprising given that England makes up about 83% of the UK population (Britton et al. (2019)). Spending per pupil within the UK is highest in Scotland and lowest in Northern Ireland.

[1] https://www.oecd.org/education/education-at-a-glance/educationataglance2019-dataandmethodology.htm
[2] Data is not available for some countries, such as Germany and Japan

Figure 1.1: School spending per student across OECD countries, 2016 US$ at PPP



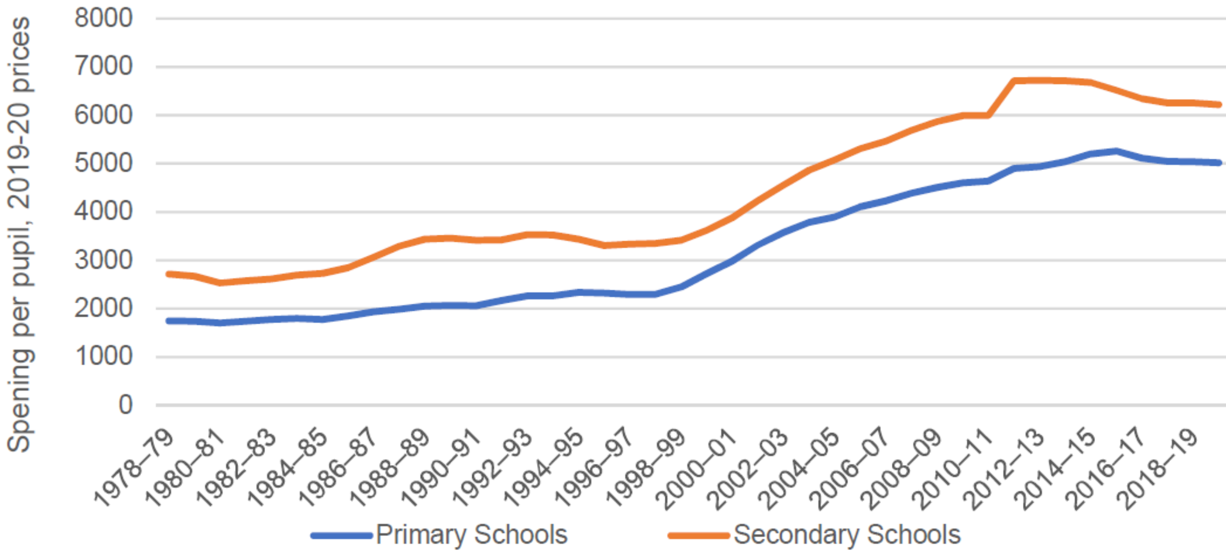
Notes and sources: OECD Education at a Glance 2019, Table C1.5 (https://www.oecd.org/education/education-at-a-glance/educationataglance2019-dataandmethodology.htm). Figures relates to public spending on primary, secondary and post-secondary non-tertiary education per full-time equivalent pupil. Figures expressed in 2016 US$ in purchasing power parity terms.

Focusing on England, spending per pupil has seen large fluctuations over time, as is shown in Figure 1.2 for primary and secondary schools. Following a period of stagnation during the 1990s, spending per pupil grew by about 5% per year in real-terms during the 2000s. This led to a real-terms increase in spending per pupil of over 65% across both primary and secondary schools between financial years 1999-00 and 2009-10.

In order to look at school funding for more recent years, one needs to look at a measure of school spending that incorporates spending by individual schools and spending by local authorities, such as with respect to services for pupils with special education needs and a number of key central functions (e.g. admissions or some support services). Trends in such a comprehensive measure of school spending are shown in Figure 1.3, which accounts for the fact that the responsibility for some services and spending has moved from local authorities to schools over time. Once one accounts for these transfers, total school spending per pupil has fallen by about 8% in real-terms since 2009-10. This represents a significant squeeze on spending and generates considerable interest in how schools can spend their budgets more efficiently.

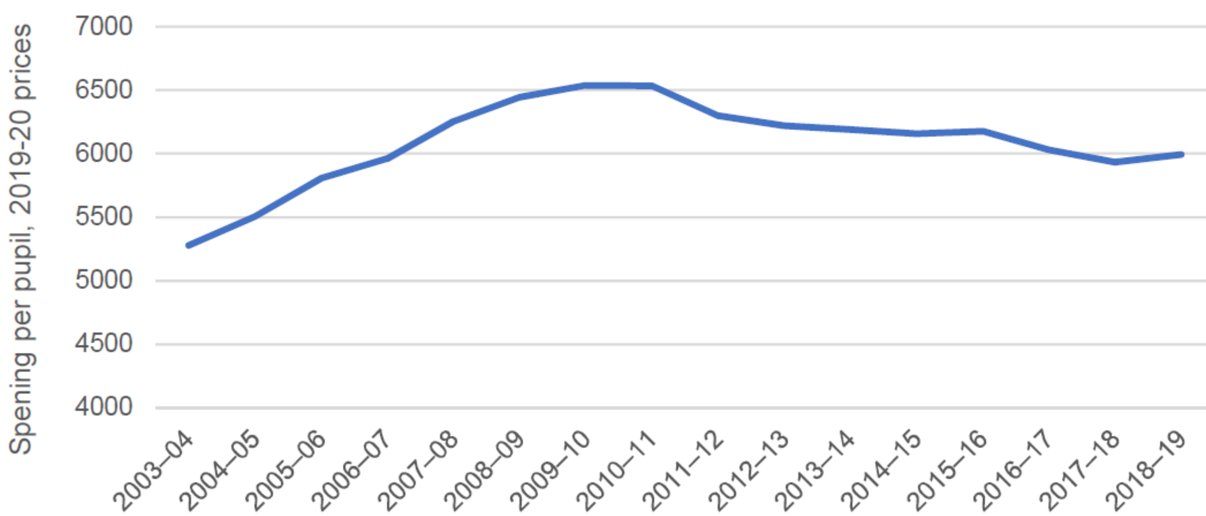
One of the biggest challenges facing schools in England is recruiting sufficient numbers of teachers to meet a growing pupil population, which has risen by around 10% or 850,000 over the last decade. Unfortunately, teacher recruitment targets have been persistently missed, particularly in maths and science subjects where graduates can earn significantly more outside of teaching (Sibieta (2018)). As a result, there is active policy interest in how incentives can be best used to attract and retain sufficient numbers of teachers, particularly in shortage subjects.

Figure 1.2: Primary and secondary school spending per pupil over time



Notes and sources: Britton et al. (2019). Relates to spending by individual schools over time. It excludes any spending by local authorities, including a big transfer of funding from local authorities to schools that took place in 2011.

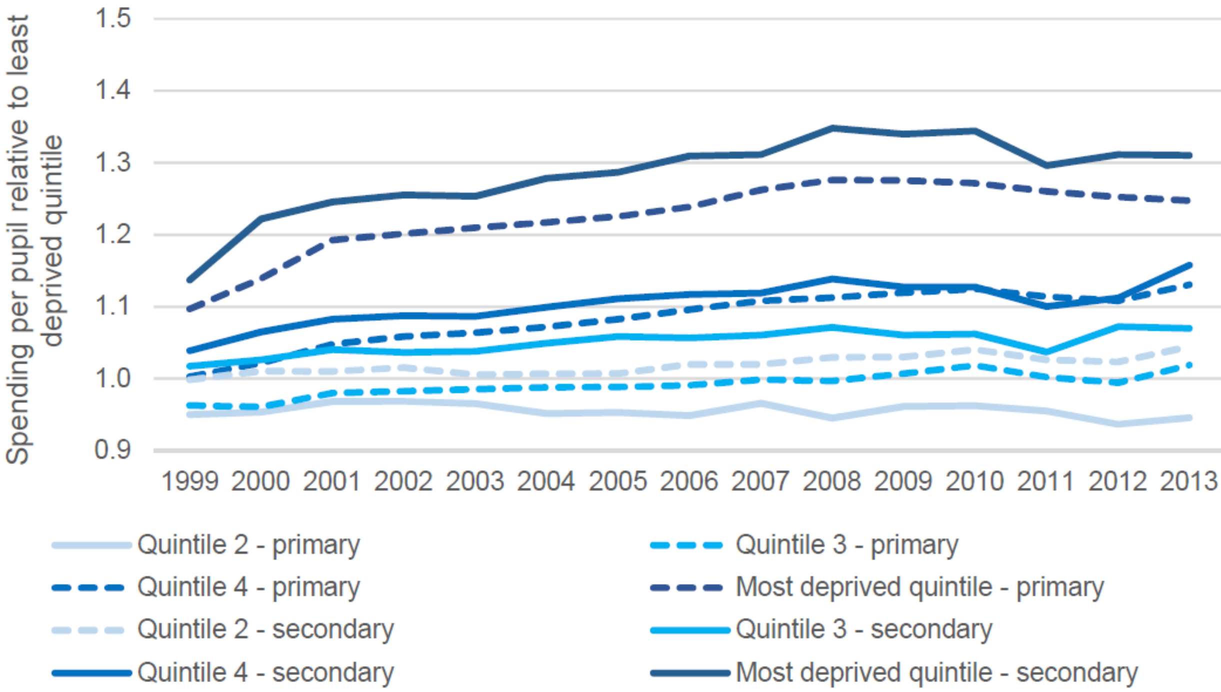
Figure 1.3: Total school spending per pupil over time



Notes and sources: Britton et al. (2019). Relates to spending by individual schools and local authorities over time on pupils aged 3-19.

There have also been big changes in the distribution of funding across different schools over time. During the 2000s, funding per pupil increased faster in deprived areas than in less deprived areas. This was a direct result of the introduction of a range of grants specifically targeted at more deprived schools as part of government efforts to reduce the achievement gap between rich and poor pupils. This continued with the introduction of the pupil premium in 2010 (a fixed extra amount of funding for each deprived pupil). As shown in Figure 1.4, the difference in spending per pupil between the most deprived quintile of secondary schools and the least deprived quintile was about 30% in 2013–14; the equivalent difference for primary schools was 25%. These figures compare with differences of just over 10% in the late 1990s.

Figure 1.4: Spending per pupil across schools by quintile of percentage of pupils eligible for free school meals (relative to least deprived quintile)



Notes and sources: Belfield and Sibieta (2016). Years refer to financial years (e.g. 1999 = 1999–2000). Adjusted for inflation using GDP deflator. Quintiles are defined according to the proportion of pupils eligible for FSM (weighted by pupil numbers).

Empirical evidence suggests that a large part of the increase in overall funding per pupil during the 2000s was used to employ extra teaching assistants in primary schools and other non-teaching staff in secondary schools, particularly in more deprived schools (Sibieta (2015a)). This change in the staffing mix is likely to have occurred as a result of a desire to provide more one-to-one support to individual pupils, a targeting of wider outcomes beyond student achievement and because such staff can be employed on relatively flexible terms and conditions. Unfortunately, there is relatively little evidence on how such staff can be used effectively to improve overall achievement or narrow the achievement gap.

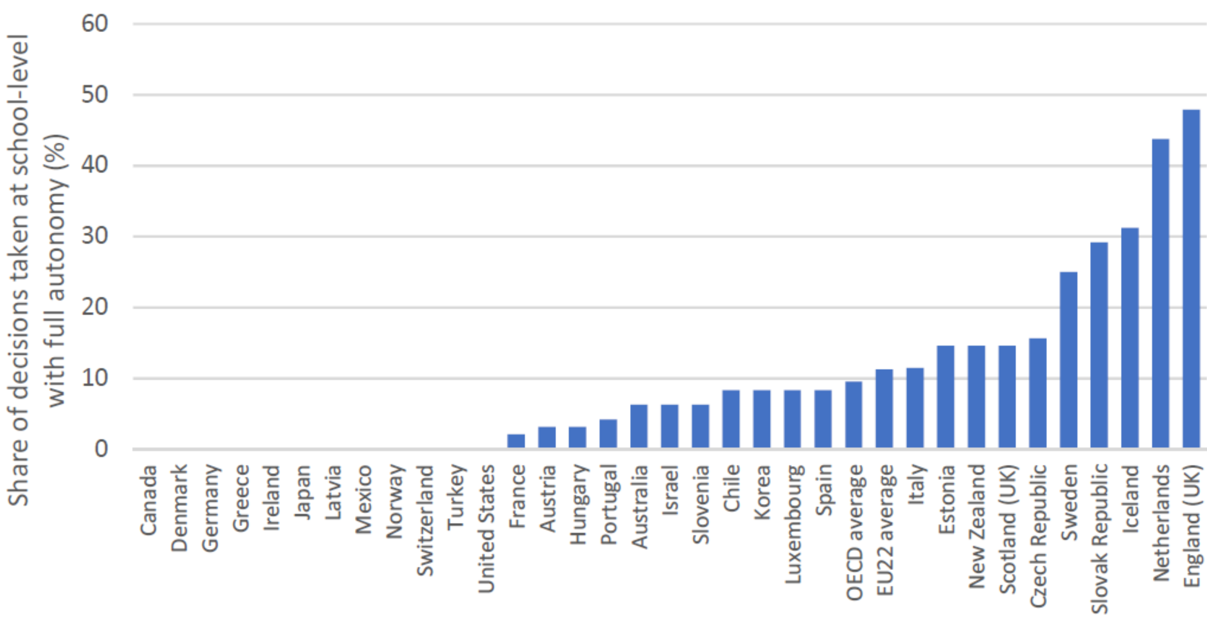
As with many other countries, school autonomy has been increasing over time in England. Indeed, as is shown in Figure 1.5, a greater share of decisions are taken at the school-level in England than in most other OECD countries. This includes expenditure decisions[3], hiring and firing decisions for all staff and

[3]Devolved to all schools in England under Local Management of Schools in 1991

considerable freedom on how much to pay teachers and other staff[4]. In addition, many schools have become self-governing and highly autonomous Academies, which are similar to Free Schools in Sweden and charter schools in the USA. Academies have further freedoms over the curriculum and many other aspects of school life. Between 2002 and 2010, about 200 Academies were opened in England, sometimes brand new schools but often replacements for previously under-performing schools. However, they remained a relatively small feature of the school system, comprising just over 5% of all secondary schools. Since 2010, schools were given the option of converting to Academy status, which led to a rapid transformation of the school system, with over 70% of secondary schools now Academies and over 25% of primary schools[5].

Empirical evidence shows that this massive restructuring of school governance in England is unlikely to have led to any significant improvements in student achievement (Eyles et al. (2018)). However, there is international evidence that increasing school autonomy can lead to improvements in student achievement in already high-performing systems (Hanushek et al. (2013)). My contribution is to better understand the ways in which schools can (and cannot) use newly found autonomy of specific resource decisions in England to improve student achievement. This is relevant to many other jurisdictions across the world where school autonomy has been increasing.

Figure 1.5: Share of decisions taken at school-level across OECD countries



Notes and sources: OECD Education at a Glance 2018, Table D6.3 (https://www.oecd-ilibrary.org/education/education-at-a-glance-2018_eag-2018-en).

[4]Schools have always possessed large degrees of autonomy over the pay and conditions of non-teaching staff, but gained significant autonomy over teacher pay levels and structures in 2013

[5]https://www.gov.uk/government/statistics/schools-pupils-and-their-characteristics-january-2019

1.3 School Resources and Student Achievement

This section presents a theoretical model of student achievement in relation to school resources. This incorporates a variety of mechanisms and different resource margins in order to illustrate the state of the current literature and how this thesis extends understanding.

Consider the level of student achievement at an individual school within a large geographic area such as a state or city-region. Schools are indexed by s and there are many schools within the area. Taking a standard school production function approach (Todd and Wolpin (2003); Hanushek (2006)), equation 1.1 assumes average student achievement (Y_s) at school s is a function of the characteristics of pupils at the school (X_s^P), the pupil:teacher ratio at the school (Q_s^T), the quantity of other resources per pupil (Q_s^R), the average quality of teachers at the school ($\bar{T}_s$) and the average quality of other resources per pupil ($\bar{R}_s$). I assume that schools are provided with a fixed level of funding per pupil from government (B_s). Schools maximise Y_s subject to a budget constraint, such that spending on teachers and spending on other goods is less than B_s (equation 1.2).

Average teacher quality ($\bar{T}_s$) is assumed to be a function of teacher wages at the school (W_s^T), the average wage at other schools in the area (W_{-s}^T), the outside wage for teachers (W^O) and pupil characteristics (X_s^P) (equation 1.3). This incorporates potential effects of teacher wages through occupational choices, efficiency wages and sorting across schools, as well as incentive effects if we allow W_s^T to represent the distribution of teacher wages.

$$max\, Y_s = f\left(X_s^P,\, Q_s^T\, Q_s^R,\, \bar{T}_s, \bar{R}_s\right)\ s.t. \tag{1.1}$$

$$B_s = Q_s^T W_s^T + Q_s^R P^R \tag{1.2}$$

$$\bar{T}_s = g\left(W_s^T,\, W_{-s}^T,\, W^O,\, X_s^P\right). \tag{1.3}$$

$$X_s^P, P^R, B_s, W_{-s}^T, W^O, \bar{R}_s\ given$$

Schools are assumed to be able to set their own teacher wage rate, reflecting evidence that suggests they possess significant monopsonistic power over wages (Ransom and Sims (2010)) and the increasing autonomy over teacher salaries possessed by schools in England, the US and other countries. Schools are assumed to be price takers for other inputs and that non-wage benefits (such as conditions of employment) are taken as fixed due to regulation[6]. Budgets are assumed to be provided by government and taken as fixed by schools when making resource decisions. The quality of other resources is taken as given ($\bar{R}_s$), an assumption which is relaxed in section 1.6 below.

[6] In England, this will reflect national conditions of service. In the US, conditions of service are generally constant within school districts.

1.4 Review of evidence on overall school resources

Since the publication of the Coleman Report in the 1960s, there have been a plethora of studies seeking to estimate the effect of school spending on student outcomes. In the context of the above model, this can be interpreted as an increase in schools budgets, B_s, holding everything else fixed. This can then affect student outcomes through increasing the quantity of teaching resources (Q_S^T) and other resources (Q_S^R). Higher spending can also improve teacher and other resource quality if wages/prices are increased as part of a spending rise and this induces a rise in quality.

Reviewing past studies, Hanushek (2003) concluded that most failed to adequately account for the endogeneity of school resources and, among those that did, only a small fraction found positive and statistically significant effects. The endogeneity results from the fact that resources are often determined by factors that also determine student outcomes, e.g. pupil characteristics (X_s^P). There is potential for either a positive or negative bias depending on the nature of the school funding system. In the US, the link between house prices and local tax revenues can create a positive bias, whilst a negative bias can result in systems where resources are highly targeted towards social deprivation, as is the case in England.

The literature has sought to address this endogeneity by controlling for a range of characteristics that affect student outcomes, making use of experiments and finding instruments for school resources. A range of recent papers have developed highly sophisticated identification strategies. Jackson et al. (2015) use court-ordered school finance reforms to estimate the effects of funding per pupil on later life outcomes. They find that a 10% increase in funding per pupil throughout childhood increases later life wages by 7%, with larger effects for children from disadvantaged backgrounds. The main mechanisms driving these effects relate to a greater quantity and quality of teaching resources (lower pupil:teacher ratios, longer instructional hours and higher teacher wages).

This pattern of findings is confirmed across a range of research designs and countries. Gibbons et al. (2018) make use of historical school funding anomalies in England to estimate that a 30% increase in school spending (relative to the mean) increases pupil attainment by about 30% of a standard deviation, with larger effects at more disadvantaged schools. Hyman (2017) also finds large effects of increases in school spending (a 10% increase in funding increases the share of students achieving a post-secondary degree by 7%), with larger effects in more disadvantaged areas. Baker (2019) finds that a 10% increase in school expenditure during childhood increases later life wages by about 7%. Lafortune et al. (2018) use school finance reforms during the 1990s to estimate that a \$1 increase in spending leads to a \$1 increase in discounted lifetime earnings. Whilst most papers examine the impact of increases in school funding, Jackson et al. (2018) show that reductions in school spending lead to large reductions in test scores too, with larger reductions for disadvantaged children.

In addition to this literature on the general impact of school expenditure, there is also a growing literature on the dynamics and efficient timing of investments. Cunha et al. (2010) argue that earlier investments in skills are likely to be more productive than later investments, and that there is likely to be dynamic complementarity between investments at different stages. This is confirmed by Johnson and Jackson (2019) who show that the effect of higher school spending is larger when preceded by high-quality pre-school. Nicoletti and Rabe (2018) have also shown that the effects of investments in secondary school are higher for those with greater levels of skills gained at primary school.

The effects of school investments in the education production function thus appear to be large (with implied elasticities of 0.7-1) and can, at least partially, compensate for lower parental investments among disadvantaged children. There is also an emerging consensus that earlier investments can have larger effects and increase the productivity of later investments.

Whilst a clear consensus exists on the effects of changes in school budgets, B_s, there is much less evidence on the effects of specific resource margins, such as salaries, incentives and the deployment of different types of staff, holding budgets fixed. This is an important evidence gap for at least two reasons. First, schools across the world are gaining increasing autonomy over resource decisions. In almost all cases, they must make these decisions within fixed budgets provided by government. Second, reductions in public spending and school budgets significantly increases interest in how existing school resources can be used more efficiently. How does the mix of spending affect student achievement? Can incentives and remuneration packages be used to maximise teacher quality, recruitment and retention in the most challenging areas? Can we use non-teaching resources more efficiently. In what follows, I evaluate the current evidence on teacher pay, recruitment and retention incentives, and the effects of how other resources are deployed.

1.5 Teacher pay and incentives

Expenditure on teachers comprises about 50% of school spending in England. A key resource choice is therefore how much to pay individual teachers and how to structure remuneration. This section extends the model in section 1.3 to show the various mechanisms by which teacher pay levels and incentives can influence student achievement. It then reviews the state of the evidence on teacher pay levels and incentives, and how chapter 2 extends this literature.

1.5.1 Relationship between teacher pay and quality

The model in section 1.3 treats the relationship between teacher pay and teacher quality as a black-box. To better understand the specific mechanisms that could drive this relationship, let us start by assuming in equation (1.4) that the quality of teacher j (T_j) includes two key components: an individual-specific component (η_j) and a component related to effort levels (e_j). One can think of the individual-specific component as including some idiosyncratic element that is fixed over time and other elements that reflect teachers' characteristics, which may or may not change over time. Most evidence suggests that teacher quality is relatively hard to predict based on observable characteristics (Ehrenberg and Brewer (1994); Wayne and Youngs (2003); Kane et al. (2008)), implying a high role for the idiosyncratic component. The two main exceptions are detailed measures of educational attainment, which have been shown to have a modest and positive relationship with teacher quality. Teacher quality also improves with experience, particularly during teachers' first few years of teaching (Wiswall (2013); Papay and Kraft (2015)). One can think of teacher effort as comprising hours of work, as well as work intensity. The importance of teacher effort is harder to evaluate empirically, given a lack of verifiable data on effort levels. However, there is work linking teacher attendance levels and student achievement (Herrmann and Rockoff (2012)).

$$T_j = q\,(\eta_j,\, e_j) \tag{1.4}$$

In order to draw a link with pay and incentive levels, let us further assume that the utility level of teacher j is given by equation (1.5) if they chose to be a teacher and by equation (1.6) if they choose their best available alternative career, as per a standard Roy model (Roy (1951); Bacolod (2007))

Expected utility as a teacher is given by a combination of mean log teacher wages (w_t), the extra wages that individual j can command as a teacher, such as through an incentive scheme (κ_j), minus the expected cost of teaching effort (e_j). Similarly, expected utility in the non-teaching sector includes average log wage in the non-teaching sector (w_o) and the extra wages that individual j can command outside of teaching (ε_j). This may reflect differences in underlying ability levels or differences in accumulated human capital, e.g. those with degrees in maths and science subjects can generally command higher outside wages (Kirkeboen et al. (2016); Britton et al. (2016a)). For illustrative purposes, expected effort outside of teaching is included within non-teaching wages.

$$u_{t,j} = w_t + \kappa_j - c.e_j,\ where\, \kappa \sim N\left(0, \sigma_\kappa^2\right) \tag{1.5}$$

$$u_{o,j} = w_o + \varepsilon_j,\ where\, \varepsilon \sim N\left(0, \sigma_\varepsilon^2\right) \tag{1.6}$$

The probability that an individual will become a teacher is given by:

$$P \equiv Pr\left(u_{t,j} - u_{o,j} > 0\right) = Pr\left(v_j > (w_o - w_t)\right) = 1 - \Phi\left(z_j\right) \tag{1.7}$$

where $v_j = (\kappa_j - c.e_j - \varepsilon_j)$and $z_j = \frac{(w_t - w_o)}{\sigma_{v_j}}$. This simply says that an individual will chose to be a teacher if the net idiosyncratic benefits are larger than the gap in average wages between the non-teaching and teaching sector. As one would expect, the probability of becoming a teacher is clearly increasing in teacher wages (w_t) and decreasing in outside wage opportunities (w_o).

Schools and policymakers can then seek to adjust both average pay levels (w_t) and individual specific components or incentives (κ_j) in ways that could improve teacher quality through various mechanisms, which one can broadly separate out into extensive and intensive margins.

1.5.1.1 Extensive margin

Higher average teacher salaries are likely to attract individuals who can command higher salaries outside of teaching. If such individuals have the propensity to be higher quality teachers than the existing stock, then teacher quality will rise with teacher salaries. In terms of the above model, this will occur if there is a strong link between the idiosyncratic component of outside wages (ε_j) and individual specific teacher quality (η_j). The fact that detailed measures of educational attainment are modest predictors of teacher quality makes such a relationship highly plausible (Ehrenberg and Brewer (1994); Wayne and Youngs (2003); Kane et al. (2008)).

The strength of such a relationship could vary, however, at different points in teachers careers. When individuals are making initial occupational choices early on in their career, there are likely to be a wide range of individuals choosing among a wide variety of career options with different expectations for lifetime earnings and ability. However, later on in teacher's careers they are likely to have built up specific occupational human capital as a teacher and a better idea of the effort costs of teaching. As a result, the effect of teacher pay levels are likely to affect teacher recruitment and retention differently.

Incentive payments (κ_j) could also have extensive margin effects depending on how they are structured. Incentives targeted specifically at high ability or high quality teachers (with κ_j directly linked to η_j) could lead to greater recruitment and retention of high-quality teachers. Similarly, one could structure payments to target teachers with potentially high-value options, such as maths and science graduates.

Most of these mechanisms relate to the teaching workforce as a whole. As implied in the model in section 1.3, one could also structure teacher pay and incentives to affect the sorting of teachers across schools. Existing evidence suggests that the sorting of teachers to schools is highly sensitive to school and pupil characteristics, with teachers less likely to be attracted to relatively deprived schools (Hanushek et al. (2004); Bonesronning et al. (2005)). One could interpret this as representing a higher expected effort cost and therefore structure incentive payments to offer some kind of compensating variation. More generally, schools choosing higher teacher salaries or higher incentive payments are likely to attract more applicants, all else being equal. As long as schools can observe potential teacher quality and high quality teacher prefer higher pay levels, this should allow higher pay schools to employ higher quality teachers (Delfgaauw and Dur (2007); Bó et al. (2013))

In summary, teacher pay and incentive payments could improve teacher quality through the extensive margin, though the effects are likely to be different for recruitment and retention given that the marginal teacher is likely to be different in each case. Pay and incentive payments could also be used by schools to affect the sorting of teachers across schools.

1.5.1.2 Intensive margin

Teacher pay and incentive payments could affect teacher quality along the intensive margin. This could be through directly rewarding teachers based on the value-added of pupils they teach. Assuming a link between teacher effort and student achievement, this effectively creates a reward for higher effort levels (Burgess and Ratto (2003); Dixit (2002)). However, there are some well known concerns regarding performance-related pay, such as reallocating effort away from outcomes that are not incentivised, dilution of incentive effects in teams and potential crowding out of intrinsic motivation (Dixit (2002); Burgess and Ratto (2003); Tirole and Bénabou (2006)).

Teacher wages could affect teacher effort levels through efficiency wage concerns (Shapiro and Stiglitz (1984)). Higher teacher salaries could, for instance, reduce teacher absence, which has been shown to reduce student achievement (Herrmann and Rockoff (2012)).

1.5.2 Teacher pay levels

The above model reveals that teacher pay levels can affect student achievement through various mechanisms, including occupational choices, efficiency wage effects and the sorting of teachers across schools. To date, the literature has mainly focused on how teacher wages affect student achievement through the first two mechanisms (occupational choices and efficiency wages), finding broadly positive effects. Gilpin (2012), Leigh (2012), Loeb and Page (2000), Britton and Propper (2016a) and Tran (2017) focus on differences in teacher wages across large geographical areas and are thus likely to exclude teacher sorting. Hendricks (2014) focuses exclusively on the occupational choices mechanism. de Ree et al. (2018) isolate the effect of higher teacher salaries on pupil attainment for existing teachers, and so through efficiency wages only.

There is much less work on the effects of increasing teacher salaries at individual schools. In contrast to the existing literature on teacher pay, this is likely to operate through a combination of efficiency wages and sorting across schools, rather than through the occupational choice mechanism. The overall effect on student achievement will also include potential negative effects of reduced spending on other resources (Q_s^R), given that the budget must balance. The existing work on teacher pay assumes, either explicitly or implicitly, that school budgets co-vary with changes in teacher pay (e.g. Loeb and Page (2000) state that pay higher salaries are also likely to provide higher levels of school funding).

In chapter 2, I therefore study a natural experiment that effectively forced schools to increase teacher salaries by about 5% within fixed budgets. This allows one to study the overall effects on student achievement of an increase in teacher salaries at individual schools within fixed budgets and how schools adjust other resources to pay for the higher salaries. This is the policy effect most relevant to an individual school.

The results suggest no evidence of any net effect of offering higher salaries on pupil attainment. Schools offer higher salaries, despite not receiving any commensurate increase in funding. In order to pay for these higher salaries, they reduce expenditure on non-teaching and non-staff costs. This paper argues that the zero net effect on pupil attainment reflects offsetting positive effects of higher teacher salaries, with evidence that this reduces teacher absence and sickness, and negative effects of reduced expenditures on other resources.

1.5.3 Incentives

Given squeezes in overall school budgets, there is naturally significant interest in the extent to which targeted incentives and payments can be used to improve teacher quality. This is particularly important given evidence suggesting that it is harder to recruit high quality teachers in some subjects, principally maths and science subjects, and in more disadvantaged areas Sibieta (2018). In line with the model in section 1.5.1, individual teacher incentives (κ_j) could then improve student achievement either on the intensive margin (with higher levels of effort or time spent on different tasks) or the extensive margin, through recruitment or retention effects.

There is now a well-established literature on performance-related pay for teachers, which generally finds that performance-related pay can lead to improvements in student achievement, operating through higher teacher effort on the intensive margin (Atkinson et al. (2009); Lavy (2009); Fryer et al. (2012); Figlio and Kenny (2007)). There is also growing evidence that performance-related pay or payments for high-quality or high-ability teachers can increase retention and thus improve student achievement on the extensive margin (Jones (2013); Hoxby and Leigh (2005); Dee and Wyckoff (2015); Swain et al. (2019)). There is, however, a clear risk that performance-related pay could crowd out intrinsic motivations (Tirole and Bénabou (2006); Burgess and Ratto (2003)) or reduce efforts on non-targeted tasks (Jones (2013)).

There is also a growing literature on the role of additional retention payments in shortage subjects or hard to staff areas. Clotfelter et al. (2008) and Feng and Sass (2018a) both find retention bonuses or payments can lead to significant improvements in retention in shortage subjects, such as maths, science and special education. Steele et al. (2010) found that a fellowship scheme in California offering payments of up to $20,000 increased participants' willingness to teach in such high-poverty schools by around 28 per cent. Cowan and Goldhaber (2016) find that bonuses of about $5,000 increased the number of "board-certified" teachers (those with extra training and expertise) at high-poverty schools, both through existing teachers seeking board certification and from board-certified teachers moving to high-poverty schools.

Therefore, there is clear evidence that financial incentives can be highly effective in reducing attrition among existing teachers in shortage subjects and they can also be effective in encouraging high-performing teachers to move to and stay at high-poverty schools. All these effects seem to be driven by higher retention and existing teachers being more willing to stay in teaching or move to different schools. Where we have much less evidence is the extent to which targeted payments can be used to improve the number and quality of teachers through the recruitment channel. The recruitment effect could in principle be quite different from retention effects as individuals will be evaluating a wide range of potential occupational options at the time of initial occupational choices, whilst existing teachers are likely to have built up significant levels of occupation-specific human capital and have a much clearer idea of the cost of effort as a teacher.

This is also an important challenge in England given the scale of challenges in the teacher labour market in England. Increases in pupil numbers have led to increasing demand for teachers. However, recruitment targets have been persistently missed over time, particularly in subjects such as maths and science (Sibieta (2018)). In response to such concerns, policymakers introduced large tax-free bursaries for teachers training in shortage subjects, with progressively higher payments for those with high degree classifications. These can be up to £25,000 up front for graduates with first-class degrees training in subjects such as physics and maths.

In chapter 3, I exploit the introduction of these bursaries and the variation by subject and undergraduate degree classification to study the effect of targeted recruitment incentives on the willingness of individuals to train as a teacher and their prior educational attainment (as a proxy for quality). I find no evidence of any significant impacts on the quantity or quality of trainees. This contrasts with the literature on retention incentives, which generally seem to have a positive effect. I argue that the differential effectiveness of recruitment and retention incentives is driven by differences in who the marginal teacher is in each case and the options under consideration. At the recruitment stage, individuals are choosing between a wide range of different occupations with large differences and uncertainty in expected lifetime earnings. At the retention stage, individuals are choosing whether to stay in teaching or not. Such individuals are likely to have built up large amounts of specific occupational human capital, be comparing against a more specific range of outside options and have a clearer idea of the expected effort costs of being a teacher. A retention incentive may be more effective at persuading a teacher to stay with known earnings and costs, as opposed to a recruitment incentive on top of a wide range of uncertain lifetime earnings options.

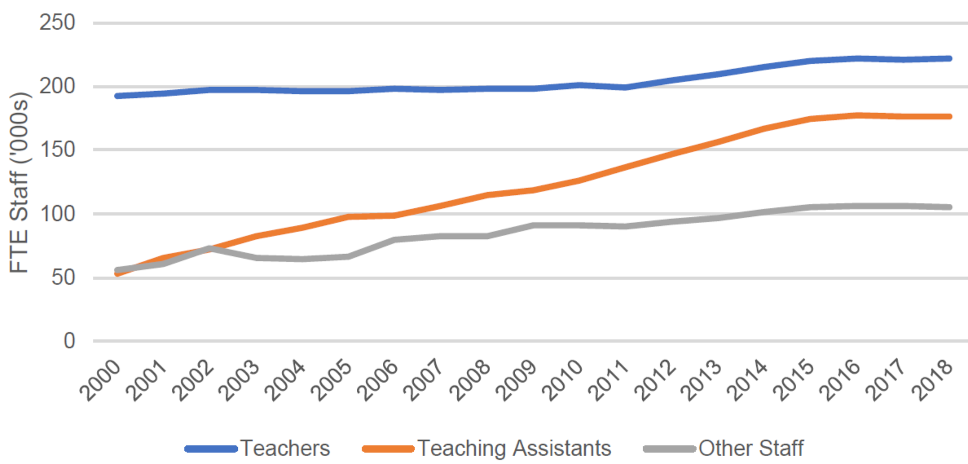
1.6 Productivity of other resources

Teachers represent the largest expenditure of schools, so it is not surprising that the academic literature has focused on the effects of teachers. However, spending on teachers now only represents about half of school spending in England. Spending on other types of staff (such as teaching assistants and administrative staff) represents about 25% of school spending, with non-staff costs taking up the other 25%. Indeed, most of the large increases in spending per pupil over the 2000s were used to expand the numbers of teaching assistants in primary schools and other staff in secondary schools (Sibieta (2015a)). As shown in Figure 1.6(a), between 2000 and 2010 the number of teaching assistants in primary schools more than doubled from 50,000 to over 125,000. Despite the squeeze on funding after 2010, numbers continued to increase to nearly 180,000 by

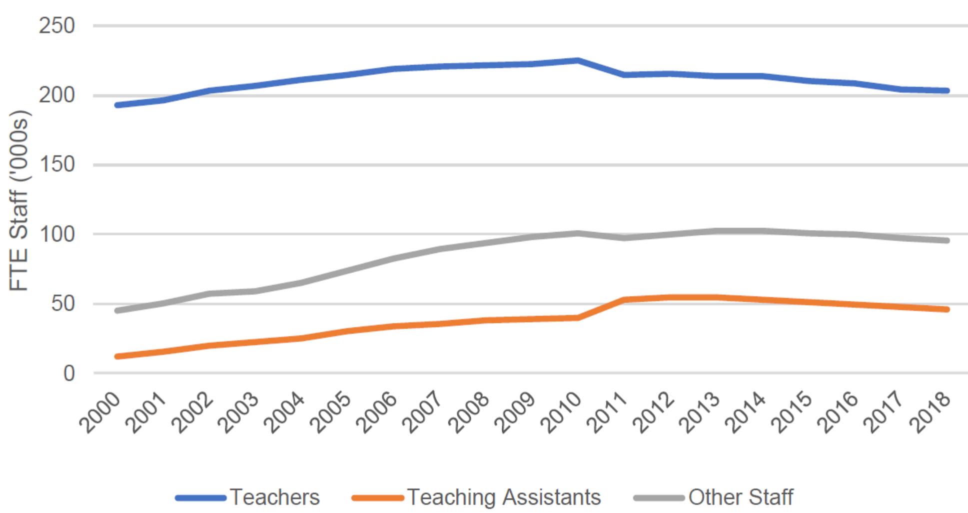
2018. This is a much faster increase than in teachers or other types of staff. In secondary schools (Figure 1.6(b), there are far fewer teaching assistants, though their number did still grow rapidly up to 2010. A large part of the extra funding for deprived schools has also been used to employ higher numbers of teaching assistants (Sibieta (2015a)).

Figure 1.6: Full-time-equivalent staff in state-funded schools in England over time

a) Primary Schools



b) Secondary Schools



Notes and sources: Department for Education, 'School workforce in England: November 2017' (https://www.gov.uk/government/statistics/school-workforce-in-england-november-2017); Department for Children, Schools and Families, 'School workforce in England: January 2008'. Years relate to January up to 2010 and November thereafter. 'Other staff' is imputed before 2011 based on the level of other and auxiliary staff recorded in 2011 and the growth rate in other staff up to 2011 (data on auxiliary staff were not recorded until 2011).

Given growth in numbers over time, understanding the productivity of investments in teaching assistants represents an important and under-studied issue. Their main roles are to provide one-to-one assistance to specific pupils who are under-performing or who have special educational needs, and to provide general support to teachers. Unfortunately, the evidence on the impact of teaching assistants is mixed, and depends on the way which they are used. Blatchford et al. (2011) and Farrell et al. (2010) find that pupils attached to teaching assistants make little extra educational progress than other equivalent pupils, and can sometimes make less progress. More recent work has emphasised the value of a model where teachers and teaching assistants share instructional responsibility in the classroom (Friend and Cook (2016)). Andersen et al. (2020) and Penney (2018) both find more positive effects of teaching assistants. There is also clear evidence that structured, high-dosage tutoring delivered by teaching assistants can improve student achievement (Fryer (2014, 2016); NFER (2014); Gorard et al. (2014)).

1.6.1 Modeling the productivity of teaching assistants

How can we reconcile these findings and, more generally, how should we think about the effectiveness of investments in teaching assistants? Equation (1.8) show a simple extension of the production function from section 1.3. In particular, equation (1.8) assumes that student achievement is a function of total factor productivity (TFP) at school s (A_s), family investments, (X_s^P), effective investments in teachers (the quantity of teachers per pupil, Q_s^T, multiplied by the average quality of teachers, $\bar{T}_s$), effective investments in other resources (the quantity per pupil, Q_s^R, multiplied by their average quality, $\bar{R}_s$) and a vector of parameters (β). A simple example of such a production function would be a Cobb-Douglas production function, where β represents the elasticities of outputs with respect to the effective units of inputs. I assume that TFP (A_s) reflects the state of knowledge and technology at school (s), which can be improved through investments in training, increased competitive pressures or improvements in the overall state of knowledge in the system.

$$Y_s = A_s f\left(X_s^P,\, \bar{T}_s Q_s^T,\, \bar{R}_s Q_s^R,\, \beta\right)\ . \tag{1.8}$$

This allows one to explicitly consider the driving factors of the effectiveness of investments in teaching assistants and other inputs. Equation (1.9) shows the marginal impact of increased investment in other resources, which is the product of total factor productivity, the average quality of other resources and the first derivative with respect to other inputs. The productivity of investments in other resources could be low if they are focused on schools with low TFP, the average quality of other resources is low or if the marginal impact of effective units is low.

$$\frac{\partial Y_s}{\partial Q_S^R} = A_s . Q_S^R . f'_{Q^R}\left(X_s^P,\, \bar{T}_s Q_s^T,\, \bar{R}_s Q_S^R,\, \beta\right)\ . \tag{1.9}$$

There are also a number of potential biases in estimating the impact of teaching assistants and other resources. Estimates could be negatively biased if investments in teaching assistants and other resources are focused on pupils or schools with low parental inputs, schools or pupils with low levels of teacher quality or if effective teaching inputs responds to other investments (such as reduced time with teachers or assignment to lower quality teachers). The empirical literature on teaching assistants acknowledges and seeks to control for all of these factors. Indeed, Blatchford et al. (2011) work explicitly argues that the impact of teaching assistants is low precisely because access to highly qualified teachers is reduced. However, the data available on parental inputs is limited, making it hard to be certain that estimates are unbiased.

This model suggests that one could improve the effectiveness of investments in teaching assistants through either improving their quality or improving TFP. One can think of ways to improve teaching assistant quality in the same way as for teacher quality, through higher pay to attract/retain higher quality staff or incentives to attract staff and increase effort. Given that teaching assistant wages are relatively low and they tend to be lowly qualified (Cribb et al. (2014)), there would seem scope for improvements. However, the labour market for teaching assistants in England is entirely decentralised, with few reforms in the last 20 years, making an empirical study of the role of pay levels quite challenging. Performance-related pay could also generate perverse incentives based on the assignment of teaching assistants to individual pupils and given uncertainty about the contribution of teaching assistants to student achievement.

One can then think of teaching assistant led trials as attempts to increase the quality of other resources (through training and development of teaching assistants) and increases in TFP through better deployment at individual schools. The fact that this work tends to find highly positive findings suggests there is considerable scope to improve the effectiveness of investments in teaching assistants. This is likely to be more significant in deprived areas given their high propensity to use extra funding to employ more teaching assistants. Efforts to improve the effectiveness of teaching assistants are also likely to be more effective at the area-level rather than the school-level, given economies of scale and potential learning across schools.

In chapter 4, I evaluate a campaign across South and West Yorkshire in England to improve the ways in which teaching assistants are used, through a combination of greater training to improve teaching assistant quality and better deployment. South and West Yorkshire represents an ideal test case for such a campaign given it is relatively deprived and under-performing relative to the the rest of England. However, evaluating an area-wide campaign is challenging from an empirical perspective due to unobservable heterogeneity. The treatment area is often deliberately chosen as one that is relatively under-performing or likely to benefit from the policy, and it is rarely possible to control for all relevant factors that determine both the treatment status and potential outcomes. One commonly employed technique to evaluate area-wide campaigns and policies is difference-in-differences (DiD). However, this is vulnerable to a violation of the common trends assumptions.

I therefore make use of the recently developed technique of synthetic control analysis, which has been employed to understand a range of policies and phenomena, such as terrorism, German reunification and smoking taxes (Abadie and Gardeazabal (2003); Abadie et al. (2010, 2015)). This seeks to re-weight a set of aggregate control units to mimic the outcomes of a treatment area over a pre-treatment phase. Through balancing time trends in pre-treatment outcomes, this seeks to account for both observable and unobservable determinants of outcomes. The weighted outcome in the treatment phase then forms the synthetic control.

In this case, I estimate a set of time-constant weights across other local authorities that best matches the pre-treatment trends in outcomes for South and West Yorkshire. This produces a good match in pre-treatment outcomes and pupil characteristics. There is then evidence of an improvement in English test scores of around 0.03-0.04 standard deviations, with little evidence of an improvement in maths scores. An effect of 0.03 should be seen as relatively large given that it represents an improvement in test scores across all 43,000 pupils over a large region. I also argue that the mechanism driving the improvement is likely to be teaching assistants spending time with more pupils, rather than just focusing on individual pupils. This can be thought of as an improvement in TFP through better deployment of teaching assistants.

Chapter 2

Constrained Optimisation? Teacher salaries, school resources and student achievement

Abstract

Should schools increase teachers' salaries to improve pupil attainment? We study the potential implications of an individual school offering higher teacher salaries from within a fixed budget by exploiting a natural experiment that forces some schools within a local area to pay teachers according to higher salary scales, but does not offer any extra funding. We show schools follow this regulation and pay their teachers more. The characteristics of teachers are largely unaffected, but teachers at high pay schools are less likely to be absent. Teacher and assistant numbers are largely unchanged. Instead, schools balance their budgets by making sizable reductions in other expenditures, particularly spending on equipment and services, amounting to around 4% of non-instructional spending. There is no evidence of any overall effect on pupil attainment. We argue that the positive effect of higher teacher pay is almost exactly countered by the negative effects of reductions in other expenditure.[1]

Keywords: Teacher Wages, School Resources, Student Achievement

JEL Codes: I20, I21, J24

[1]This represents joint work with Ellen Greaves. We are grateful to the ESRC for funding (small grants programme: ES/J006076/1; Centre for the Microeconomic Analysis of Public Policy: RES-544-28-0001 & ES/M010147/1). We thank Rebecca Allen, Jack Britton, Alex Bryson, Steve Gibbons, Carol Propper, Imran Rasul, Nikki Shure and seminar participants at the IFS, IoE, RES, EEA and PEUK for comments. Provision of data from the Department for Education is much appreciated. All errors remain our own.

2.1 Introduction

Schools across developed countries are gaining more autonomy over budget decisions and teacher salaries, meaning it is increasingly important to identify the effect of school spending choices on pupil attainment. In England, recent reforms have given state-funded schools more autonomy over the level and structure of teacher pay. In the US, charter schools have more freedom to set teacher pay outside regulations imposed on public schools, whilst the Race to the Top initiative encouraged states and school districts to decentralise more budgetary decisions to public school principals. Previous research has studied the effect of increasing the overall level of resources, with most recent work finding positive effects of higher resources (Jackson et al. (2015), Gibbons et al. (2018), Hyman (2017), Lafortune et al. (2018), Jackson et al. (2018), Baker (2019). Other evidence suggests offering higher teacher salaries across large geographic areas may improve pupil attainment (Loeb and Page (2000), Gilpin (2012), Leigh (2012), Hendricks (2014), Britton and Propper (2016a) and Tran (2017)). However, there is no existing evidence of the effects of an individual school offering higher salaries and the underlying mechanisms through which this may affect pupil attainment.

In this paper, we study a natural experiment that effectively forces some schools within a local area to pay teachers according to higher salary scales, holding school budgets constant.This allows us to produce convincing estimates of the effects of raising salaries at individual schools, and the underlying mechanisms driving the reduced form effects: teacher sorting and efficiency wage concerns, countered by reductions in other expenditure.

In contrast, existing estimates of the effects of teacher salaries (Loeb and Page (2000), Gilpin (2012), Leigh (2012), Hendricks (2014), Britton and Propper (2016a) and Tran (2017)) typically consider the impact of raising teacher salaries with a commensurate increase in funding, which is not necessarily relevant for individual schools. Through studying the effects of differences in teacher and outside wages over wide geographical areas, the underlying mechanisms in these studies include teacher occupational choices (in addition to the sorting of teachers across areas/schools and efficiency wages) which is largely irrelevant for individual schools when considering whether to raise teacher pay.

The natural experiment we exploit is a sharp geographical discontinuity in teacher salary scales in England. During our period of study, teachers in England were paid according to centrally determined salary scales. These increase with experience, but are also higher in the London area to compensate for a higher cost of living. There are four pay zones: inner London, outer London, fringe London and the rest of England. In the case of inner and outer London, higher teacher salaries come with higher levels of funding, and the boundaries of the pay zone coincide with other administrative geography that are likely to affect pupil attainment (boundaries of school districts and zoning for housing development). We instead focus on the boundary between fringe London and the rest of England. For a given level of experience, teacher salary scales are about £1,000, or about 5%, higher in fringe London as compared with the rest of England. This boundary often runs within school districts (or local authorities as they are called in England), meaning that education policy is relatively constant for schools in close proximity and either side of the boundary. Teachers are free to live either side of the boundary, the cost of living is likely to be constant within a small distance to the boundary and schools and communities are likely to be similar either of the boundary (which we confirm). Conditions and other benefits are also constant across the boundary given nationally governed pay and conditions. We therefore interpret the boundary as representing an exogenous increase in the salary that must be paid to a teacher of a given level of experience. We show that schools inside the fringe London boundary don't receive proportionately higher funding to account for the regulated higher teacher salary scales. Schools must therefore pay the higher teacher salaries from within their fixed budget, which allows us to study the resource choices made by schools. Anomalies in funding differences across areas have already been studied in England to show that increases in funding (without increases in teacher salaries) can improve pupil attainment (Gibbons et al. (2018)).

We confirm that teacher pay follows the regulation and that schools in the "high pay" area spend a greater share of their budgets on teacher salaries as a result, with offsetting reductions in non-staff expenditures to balance the budget. There is little evidence that this leads to observable differences in the sorts of teachers at individual schools, although we have imperfect measures of teacher quality. We do find indirect evidence to support an efficiency wage mechanism, however, with teachers on the high pay side of the boundary less likely to be absent or sick. This is a valuable finding as there is little existing evidence of efficiency wage effects for teachers. Finally, there is little evidence of any net positive effect on pupil attainment from the combination of higher teacher salaries and reduced non-teacher spending. Due to the relative precision of our estimates, in our preferred specification we are able to rule out quantitatively small positive estimates of 0.046 and 0.036 standard deviations in English and maths at age 11, respectively.

This combination of results suggest that schools optimally adjust resources so that the combined effect of higher regulated teacher salary scales and consequent changes in other expenditure has few implications for student achievement. Given that the pay regulation has been in place since the 1970s, our estimates of a near zero policy effect represent a long-run equilibrium.

The near zero total policy effects could be because the positive effects of higher salaries and negative effects of reduced non-teacher expenditures are both small. However, it could also be that larger effects offset each other. We argue that the latter is a more plausible explanation. The recent literature on school resources shows significant effects of changes in school expenditures (Jackson (2018)), including negative effects of reductions in school resources overall (Jackson et al. (2018)). There is also a well-established literature showing that teacher sorting is sensitive to financial incentives (Clotfelter et al. (2008) and Steele et al. (2010)). Cabrera and Webbink (2019) argue that such incentives only improve pupil achievement if they lead to changes in teacher characteristics linked to pupil attainment, such as teacher experience. Whilst we don't find any evidence of changes in teacher experience, we do find evidence of reduced teacher absence as a result of higher salaries, which is linked to higher student achievement (Herrmann and Rockoff (2012)).

Our results imply that the net benefits of a school offering higher teacher salaries from within a fixed budget are likely to be minimal. This is relevant to both regulated and autonomous schools in England, as both school types have similar freedoms on non-teacher expenditures, and, since 2013, almost complete autonomy over teacher pay and progression. Our results also imply that existing school resource decisions are relatively efficient as schools are able to adjust other resources in ways that lead to little overall change in pupil attainment.

These results are mainly relevant for schools with autonomy over teacher pay and the setting of non-instructional resources. However, our exploration of the size of the underlying mechanisms increases the relevance to all school systems, e.g. how teacher salaries can affect teacher absence and the likely negative effects of cuts to non-instructional spending. The main limitation is that the higher teacher salaries and cuts to non-instructional spending could affect wider outcomes we are not able to observe, such as other aspects of teacher characteristics or pupils' mental health.

This chapter proceeds as follows: Section 2.2 motivates our empirical approach by presenting a brief model of how student achievement relates to teacher wages and school resources. Section 2.3 summarises the key institutional details relating to schools and teachers in England. Section 2.4 describes our empirical strategy and data sources. Section 2.5 presents our empirical results. Section 2.6 concludes.

2.2 Teacher Wages, School Resources and Student Achievement

This section presents a model of student achievement in relation to teacher salaries and school resource decisions, which is very similar to that presented in chapter 1. This motivates our focus on the effect of changing teacher wages at individual schools within a fixed budget and how this differs from the current literature on the effects of teacher wages. We also detail the mechanisms through which this resource allocation affects student achievement.

Consider the level of student achievement at an individual school within a large geographic area such as a state or city-region. Schools are indexed by s and there are many schools within the area. Taking a standard school production function approach (Todd and Wolpin (2003); Hanushek (2006)), as shown in equation 2.1, we assume average student achievement (Y_s) at school s is a function of the characteristics of pupils at the school (X_s^P), the pupil:teacher ratio at the school (Q_S^T), the quantity of other resources per pupil (Q_S^O) and the average quality of teachers at the school $(\bar{T}_s)$. We also assume that schools are provided with a fixed level of funding per pupil from government (B_s). Schools maximise Y_s subject to a budget constraint, such that spending on teachers and spending on other goods is less than B_s (equation 2.2).

Average teacher quality $(\bar{T}_s)$ is assumed to be a function of teacher wages at the school (W_s^T), the average wage at other schools in the area (W_{-s}^T), the outside wage for teachers (W^O) and pupil characteristics (X_s^P) (equation 2.3).

$$max\; Y_s = f\left(X_s^P,\, Q_s^T\, Q_S^O,\, \bar{T}_s\right)\; s.t. \tag{2.1}$$

$$B_s = Q_s^T W_s^T + Q_S^O P^O \tag{2.2}$$

$$\bar{T}_s = g\left(W_s^T,\, W_{-s}^T,\, W^O,\, X_S^P\right). \tag{2.3}$$

$$X_s^P, P^O, B_s, W_{-s}^T, W^O\; given$$

We assume schools are able to set their own teacher wage rate, reflecting evidence that suggests they possess significant monopsonistic power over wages (Ransom and Sims (2010)) and the increasing autonomy over teacher salaries possessed by schools in England, the US and other countries. We assume that they are price takers for other inputs and that non-wage benefits (such as conditions of employment) are taken as fixed due to regulation[2]. Budgets are assumed to be provided by government and taken as fixed by schools when making resource decisions.

This model incorporates a number of potential mechanisms by which changes in relative teacher wages can affect teacher quality and thereby student achievement. First, as per a standard Roy model (Roy

[2]In England, this will reflect national conditions of service. In the US, conditions of service are generally constant within school districts, which Loeb and Page (2000) difference out through school district fixed effects.

(1951)), changes in relative teacher wages will affect the average quality of individuals who chose to become teachers through their occupational choices.[3] Second, relative teacher wages could affect teacher effort levels through efficiency wage concerns (Shapiro and Stiglitz (1984)). Third, the level of teacher wages at individual schools will affect the sorting of teachers between schools. Schools choosing higher teacher salaries are likely to attract more applicants. As long as schools can observe potential teacher quality and high quality teacher prefer higher pay levels, this should allow higher pay schools to employ higher quality teachers.

To date, the literature has mainly focused on how teacher wages affect student achievement through the first two mechanisms (occupational choices and efficiency wages). Gilpin (2012), Leigh (2012), Loeb and Page (2000), Britton and Propper (2016a) and Tran (2017) focus on differences in teacher wages across large geographical areas and are thus likely to exclude teacher sorting. Hendricks (2014) focuses exclusively on the occupational choices mechanism. de Ree et al. (2018) isolate the effect of higher teacher salaries on pupil attainment for existing teachers, and so through efficiency wages only. In contrast, the effect of an individual school changing teacher pay will reflect a combination of the second two mechanisms (efficiency wages and sorting across schools), as changes in teacher pay at one school are unlikely to affect occupational choices.

Our interest lies in the marginal policy effect of changes in teacher wages at individual schools on overall student achievement, holding the budget fixed, which has not been estimated in the literature to date. This overall effect will incorporate a combination of different mechanisms, potentially working in opposite directions. Pupil attainment could be boosted by increases in teacher quality ($\bar{T}_s$) driven by a combination of efficiency wage and sorting effects. Clotfelter et al. (2008) and Steele et al. (2010) find that teacher sorting responds to financial incentives, for example, although some evidence suggests that the sorting of teachers to schools is more sensitive to school and pupil characteristics than wages (Hanushek et al. (2004); Bonesronning et al. (2005)). There is no existing evidence on the direct implications of teacher sorting for student achievement. The overall effect will also include potential negative effects of reduced spending on other resources (Q_S^O), given that the budget must balance. The new literature on school resources suggests that such negative effects could be important (Jackson et al. (2015); Jackson (2018)). The existing work on teacher pay assumes, either explicitly or implicitly, that school budgets co-vary with changes in teacher pay (e.g. Loeb and Page (2000) state that pay higher salaries are also likely to provide higher levels of school funding). Also underlying our effect is any attempt by other schools or policymakers to respond to pay differences across schools, through pay or non-pay factors. We argue below that such effects are small in practice.

The main implication from this analysis is whether shifting school budgets towards higher teacher salaries can improve pupil attainment, i.e. whether the efficiency wage and sorting effects outweigh the negative resource effects. If there are large net positive or negative effects, this would imply schools are currently behaving inefficiently, as a simple shift in resources can improve pupil attainment. A near zero effect would imply schools are allocating current resources in an efficient manner. Furthermore, how other resources adjust to fund higher teacher salaries could in principle reveal new evidence on schools' resource preferences.

Consideration of this model therefore highlights that changes in teacher wages can operate through three primary channels: occupational choices; efficiency wage concerns; sorting of teachers across schools, with the first unlikely to be relevant in our setting. The model also indicates that the offer of higher teacher

[3] As discussed below, this mechanism is not relevant for our study, but we include it to compare our setting to the previous literature on the impact of teacher salaries on pupil attainment.

wages by schools is likely to be an endogenous response to their situation. To identify the impact of higher teacher wages on student achievement one therefore cannot simply compare student achievement at schools offering different levels of teacher wages. Ideally, one would use a randomised experiment that forced some schools to offer higher teacher salaries, holding budgets fixed. In the next section, we describe a natural experiment in England that comes close to replicating such a scenario.

2.3 Institutional Background

Our empirical strategy makes use of sharp geographical discontinuities in teacher pay levels in England. In this section, we describe how schooling and the teacher labour market operate in England, concluding with the implications for our empirical strategy.

2.3.1 Schools and Teachers in England

The majority of pupils in England attend primary schools from ages 4-11 before attending secondary school after age 11 through to age 16. We focus on primary schools rather than secondary schools as they are more numerous, which increases the precision of our results. All pupils in state-funded schools in England must sit external assessments at the end of primary school in Maths and English, known as Key Stage 2 (KS2) tests. These are our main measures of student achievement[4].

There are two main types of school in England: maintained schools and Academies. The main differences are that Academies have more freedoms over school organisation, can deviate from the national curriculum and, before 2013, had more freedoms over teacher pay. Academies are very similar to US charter schools. However, it is only since September 2010 that primary schools could apply to become Academies. Over the period covered by our data (2006 to 2011), most state-funded primary schools were maintained schools and were the responsibility of the 151 Local Education Authorities (LEAs) across England.

LEAs are similar to US school districts. They are responsible for providing budgets to individual schools, coordinating admissions, assisting with the governance of the school and providing some central services for all schools in their area (e.g. support for children with special educational needs). Although LEAs are responsible for funding schools, this money is not raised through local taxation. It primarily comes from a grant from central government (called the Dedicated Schools Grant)[5]. This grant is supposed to reflect the differences in the needs and costs of providing schooling in different areas. Indeed, there is an "Area Cost Adjustment" to account for differences in costs, though prior work has suggested this has not always been perfectly aligned with actual differences in costs (Gibbons et al. (2018)). LEAs are then responsible for how to distribute this grant to different schools in their area. Each sets its own school funding formula, with the most important factors being pupil numbers and the socio-economic mix of pupils at the school (Sibieta (2015b)).

Individual schools are responsible for resource decisions, given the fixed budget from the LEA. In particular, it is individual schools who make hiring decisions for teachers and other staff and how much to spend on different types of resources, subject to regulations. For example, there are maximum class sizes for

[4]Pupils in private or independent schools do not have to sit these tests, but account for only about 6% of each cohort of pupils.

[5]LEAs are able to add to this grant if they wish, but very few do in practice (Sibieta (2015b))

under 7s and there exists national pay and conditions for teachers, which are reflected in the School Teachers' Pay and Conditions Document (STPCD). All schools (maintained schools and Academies) have considerable freedom as to how they employ other members of staff (e.g. teaching assistants and administrative staff), who can be employed on temporary or fixed term contracts. All schools also have freedom on spending on non-staff resources, such as books, services and energy.

2.3.2 Teacher Pay

During the period we study, teachers in England were paid according to a national salary scale that had nine different points (M1-6, U1-3). In principle, schools had some discretion about how quickly their teachers move up the pay scale. In practice, position on the pay scale was almost entirely determined by years of experience. Schools could choose to use some additional payments to pay teachers above the salary scales if they wished[6], but these flexibilities were rarely used by schools.

Since September 2010, many primary schools have converted to Academy status and thus gained additional freedom over teacher pay, though few will have had any opportunity to use these freedoms before our final outcomes are measured in May 2011. From September 2013, all state-funded schools in England were given more autonomy over teacher pay, with pay scales replaced by minimum and maximum levels and all schools are required to have their own individual pay policies. Our research describes the likely implications of a school choosing to use these freedoms to offer higher salary levels, which is now a relevant choice for all schools in England.

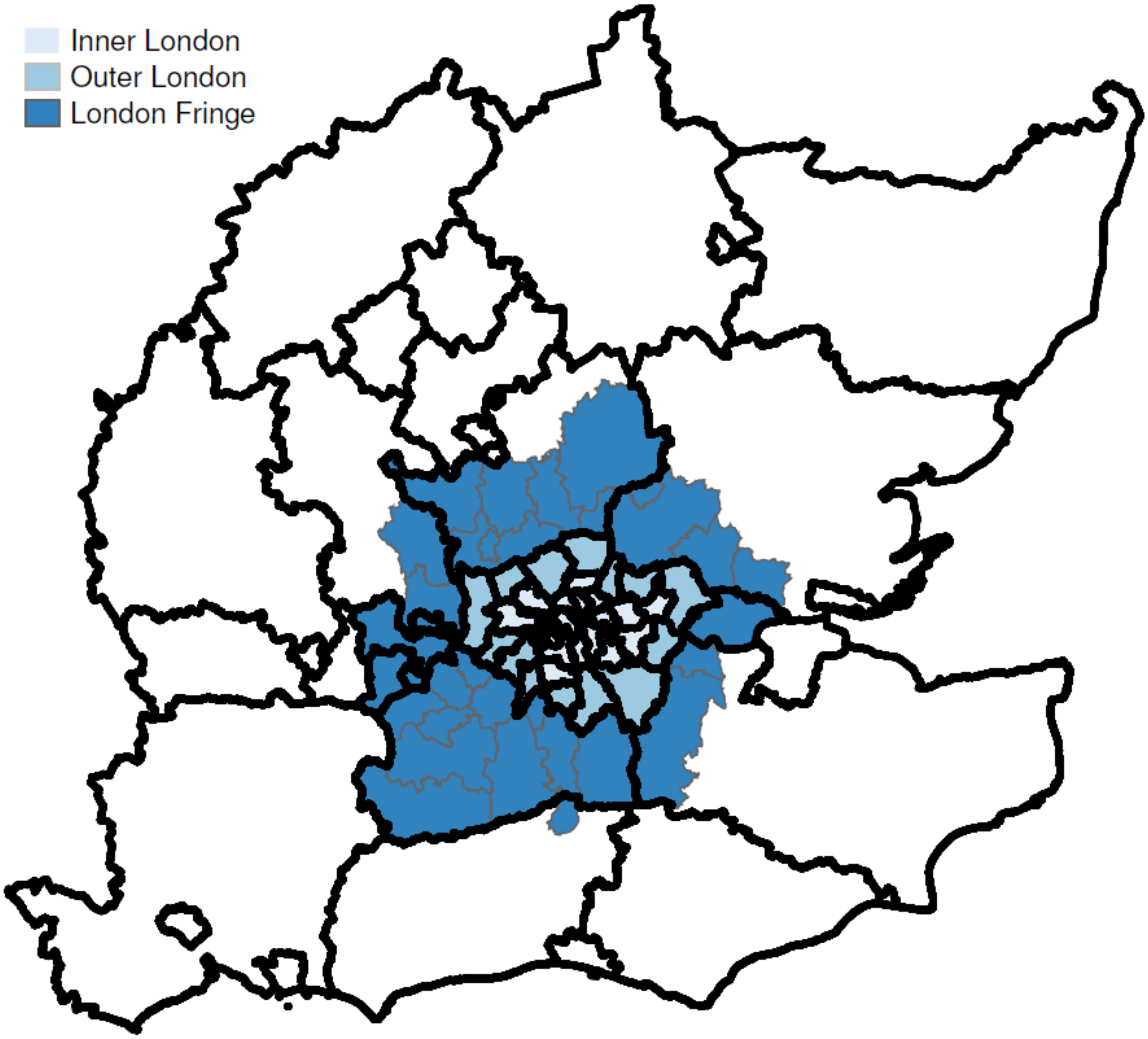
In order to reflect the higher cost of living in or near London, the level of the pay scales varies across the London area. Specifically, there are four different pay zones: inner London (highest pay zone); outer London; the fringe area of London; and, the rest of England and Wales (lowest pay zone). These are shown in Appendix Figure A1, excluding the rest of England and Wales. The boundary of the inner London and outer London pay boundary both coincide with the boundaries between LEAs. The outer London boundary also coincides with the boundary of the "green belt" (which severely reduces housing development on the low-pay side of the border). As a result, the inner and outer London boundary could be correlated with differences in schools policy and family background. By contrast, the fringe boundary (shown in Figure 2.1) mostly runs within LEAs, along local authority district boundaries (e.g. within Kent or Buckinghamshire). Local authority districts are smaller local administrative units that are responsible for a number of non-education functions., e.g. waste collection. As a result, there is no reason to expect education policy or pupil characteristics to differ either side of the fringe boundary (which we confirm empirically).

The boundaries of the pay zones and the system of London allowances were introduced in 1974 following the recommendations of the Pay Board and Houghton Committee reports of the same year (Zabalza and Williams (1979)). The fringe zone was established in order to prevent a large step change in teacher salaries at the outer London boundary. The fringe boundary has not changed since it was first established and has largely remained as a fixed amount on top of the pay scales for the rest of England and Wales. Given that the fringe pay zone has existed for more than 30 years, our results represents the the long-term equilibrium.

Our empirical strategy relies on the differences in pay scales across the fringe boundary. Figure 2.2

[6] Additional payments include recruitment and retention payments, teaching and learning responsibility payments and payments for teachers working with children with special educational needs.

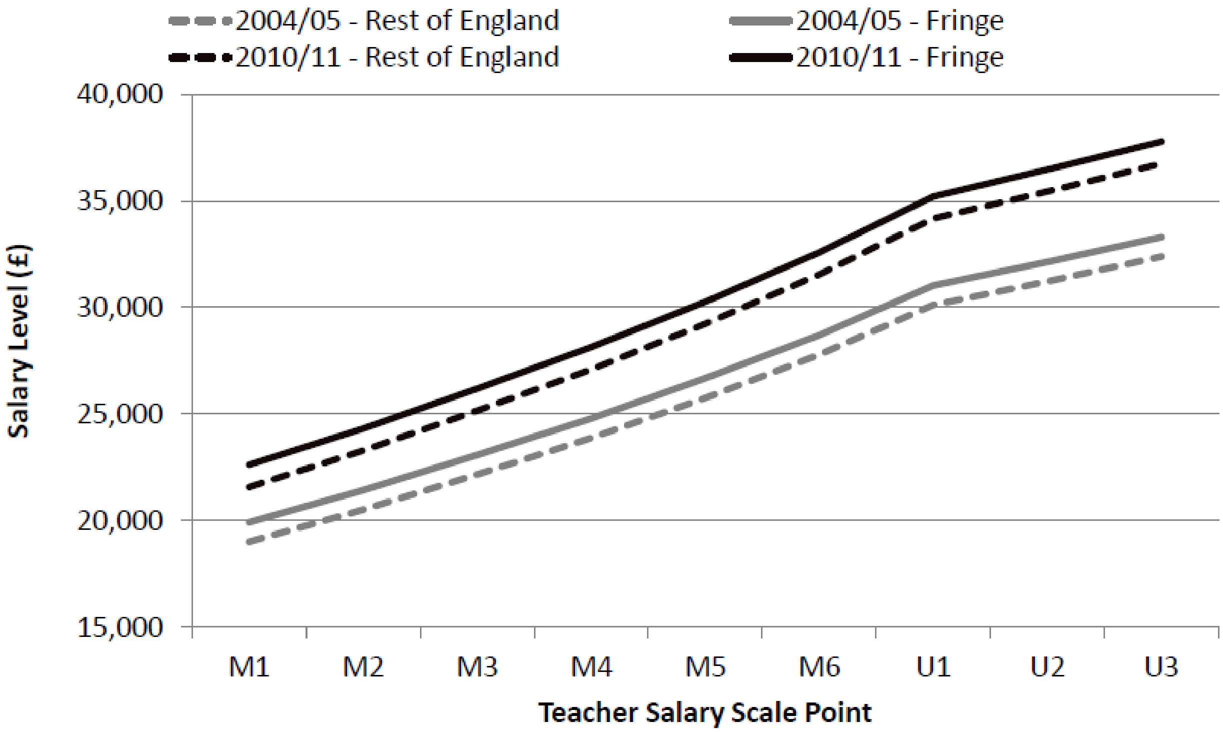
Figure 2.1: Fringe London Pay Boundary



Notes and Source: School Teacher Pay and Conditions Document; solid lack lines represent LEA boundaries.

shows the level of each point of the seven points in the pay scale either side of the fringe boundary over the period covered by our data (2006 to 2011). The differences across the boundary correspond to about £1,000 right across the pay scale throughout the period, or about 3-5% of teacher salary levels in the rest of England. These differences are within the range of increases in teacher salaries that individual schools could decide to award from within their fixed budgets, and so are policy relevant. It is unlikely that schools would be able to unilaterally offer substantially higher salaries.

Figure 2.2: Level of Teacher Salary Scales across Fringe and Rest of England (2004/05 and 2010/11)



Source: School Teachers Pay and Conditions Document.

2.3.3 Implications for School Resource Decisions and Student Achievement

We interpret the pay boundary as representing a difference in the minimum salary levels that must be paid to teachers with a given level of experience. We analyse the net implications of this rule for student achievement and detail the mechanisms driving this effect, including: actual teacher pay levels; school expenditure decisions; teacher sorting; and, teacher absences.

As argued in Section 2.2, the implications for school resource decisions and student achievement will depend on differences in school budgets at the boundary. Given that we find no difference in school budgets at the boundary, schools on the high pay side effectively face a higher price of employing teachers with no compensating change in their income. It therefore forms an ideal natural experiment to study the implications of increases in teacher pay from within fixed budgets.

The fact that there is no difference in funding per pupil at the fringe boundary is somewhat surprising. At the inner and outer London boundaries, schools do receive higher levels of funding (13-14%) to compensate for higher teacher salaries. It is not entirely clear why there is no evidence of any compensation for schools at the fringe boundary. It may be that because the differences in salaries are relatively small, local authorities felt they could be absorbed by schools. This is no longer the case, as new simpler local authority funding formulae were introduced in 2013, which explicitly recognise the fringe boundary. This, and the introduction of greater freedoms on teacher pay are the main reasons we do not extend our analysis beyond 2011.

Absent any restrictions on teacher numbers, the combination of higher teacher pay and unchanged budgets should lead to re-enforcing substitution and income effects that reduce teacher numbers (assuming teacher expenditure is a normal good). Regulation on teacher numbers (such as maximum class sizes for pupils aged 5-7) seems likely to limit any reduction in spending on teachers, however, and may even lead to an increase in spending if the restrictions are strong enough and the income and substitution effects are weak. Teachers are also a discrete good and schools might not be willing to reduce teachers numbers in response to a 5% increase in teacher salaries. However, there are ways that schools could reduce expenditure on teachers, e.g. limiting additional payments, hiring teachers with different experience or slowing teachers' progress through the salary scales. Schools on the low-pay side of the boundary could further smooth the actual difference in teacher pay across the boundary by paying their teachers more. We test for such effects by examining the actual differences in teacher pay across the boundary, the level of additional payments and average position on the salary scale.

In principle, schools could consider adjusting wider conditions and non-pecuniary rewards to respond to the pay regulations. However, this is unlikely to be feasible in practice, given statutory national pay and conditions. Any changes to conditions of service could also be implemented by schools on both side of the boundary.

We therefore interpret the net impact on pupil attainment as the effect of higher teacher salaries and offsetting reductions in other expenditure, exploring these mechanisms in as much detail as possible given the available data.

2.4 Data and Empirical Methods

Our identification strategy relies on a sharp geographical discontinuity in teacher pay scales. In particular, we compare resource choices, teacher characteristics and student achievement at schools either side and within close proximity of the fringe London teacher pay zone boundary. We therefore require schools to be well balanced in areas in close proximity to the boundary. This section outlines our identification strategy, data sources and summary statistics that suggest our identification assumption is credible.

2.4.1 Empirical strategy

For ease of exposition, let us assume that the production function in equation 2.1 is additive and separable. Taking the difference in mean pupil attainment across schools in the high (H) and low-pay (L) regions gives:

$$\overline{Y}_H - \overline{Y}_L = \beta_p \left(\overline{X_H^P} - \overline{X_L^P}\right) + \beta_Q \left(\overline{Q_H^T} - \overline{Q_L^T}\right) + \beta_O \left(\overline{Q_H^O} - \overline{Q_L^O}\right) + \beta_T \left(\overline{T}_H - \overline{T}_L\right) + \left(\overline{\varepsilon}_H - \overline{\varepsilon}_L\right) \tag{2.4}$$

where X^P denotes observable pupil characteristics, Q^T denotes the pupil:teacher ratio, Q^O denotes the ratio of other inputs to pupil numbers, T denotes average teacher quality, ε denotes unobservable pupil characteristics and where and $\overline{x}$ denotes the mean of x across schools.

By comparing schools in very close proximity to the boundary, we expect there to be very small differences in observable pupil characteristics ($\overline{X_H^P} - \overline{X_L^P}$) and unobservable characteristics ($\overline{\varepsilon}_H - \overline{\varepsilon}_L$). We test this assumption by examining the differences in a range of pupil characteristics at schools either side of the boundary, which suggests very few (if any) differences.

Our preferred specification estimates the difference in student achievement and resource choices using schools within 2 km of the boundary. We account for pupil and school characteristics using Fully-Interacted Linear Matching (FILM) (Blundell et al. (2005)). FILM differs from standard OLS regression in that FILM linearly interacts the treatment effect with all pupil and school characteristics. This then provides an impact estimate for all schools in the sample given their characteristics, which we average across all schools in the high pay region to correspond to the average treatment effect on the treated (ATT). The main advantage of FILM compared to OLS is that it is more flexible in allowing the treatment effect to vary with school characteristics. In our case, the FILM estimates are also generally more precise than both OLS and propensity score matching.

We test the robustness of our results by comparing schools within various distances to the boundary (1, 2 or 3km) and using various methods to control for observable characteristics (raw differences, OLS and kernel matching). We also use local linear regression methods often used in regression discontinuity design (Lee and Lemieux (2010)) to confirm that student achievement is similar across the boundary no matter what distance we use. Data is pooled across years, but the point estimates are very similar when we estimate separately by year.

2.4.2 Data

We link together data from a number of administrative data sets over various years. In particular, we use data from the National Pupil Database from 2006 to 2011, which contains the test results and observable characteristics for every pupil in state-funded schools in England. Our main outcomes are the school-level average points scores in Key Stage 2 Maths and English, standardised at the national level.[7] We disregard data for 2010 as a large number of schools boycotted Key Stage 2 national examinations in that year. Our sample consists of schools with non-missing Key Stage 2 results who remain in the sample for all years from 2006 to 2011 (excluding 2010) and who are close to the fringe boundary.

We derive various school level characteristics: number of pupils; proportion of pupil eligible for free school meals (FSM); proportion of pupil with English as an additional language (EAL); proportion of pupils with special educational needs (SEN), with and without statements; and, proportion of pupils from non-white ethnic backgrounds. Eligibility for free school meals is a rather coarse measure of deprivation, so we also use other measures of deprivation based on the area in which pupils live: average percentile rank on the Index of Multiple Deprivation and average percentile rank on the Income Deprivation Affecting Children Index (IDACI).

We use funding and expenditure levels defined in Section 251 outturn data, which reports funding and expenditure levels for each financial year (April to March) for all maintained schools in England. Information on staffing levels is taken from the Local Education Authority School Information Service (LEASIS) and its later replacements. We also make use of Consistent Financial Reporting data (CFR), which shows spending per pupil on different types of inputs, grouped as follows: teachers; teaching assistants; all other staff; services (e.g. catering and energy); and, equipment (e.g. books, maintenance and other learning resources).

As Key Stage 2 tests are taken in the summer of each year, we link these results to school characteristics defined in January of the same year (taken from the Spring Census), staffing levels defined for the same academic year (LEASIS) and to funding/expenditure levels in the financial year most recently ended.

[7] Externally assessed Science tests were stopped from 2009 onwards.

For teacher pay levels, we make use of the School Workforce Census that contains the pay, experience, turnover, absence and broad characteristics for all employees in schools across England from 2010 onward.[8] Our final sample consists of 238 primary schools (111 on the high-pay side of the boundary and 127 on the lower-pay side) which have a common sample of all outcomes.

2.4.3 Descriptive statistics

As discussed in section 2.4.1, small or insignificant differences in observable characteristics across the boundary would make our identification assumptions more plausible. Table 2.1 (panel A) shows the average characteristics of schools within 2km of the fringe London pay boundary and compares the characteristics of all schools just inside and outside each boundary.[9]

Table 2.1: Balance of pupil characteristics and summary statistics across Fringe London Boundary (2km)

	Within 2km		
	Inside	**Outside**	**Difference**
A) Pupil Characteristics			
Prop. FSM	0.076 (0.072)	0.077 (0.084)	-0.001
Prop. SEN (no statement)	0.206 (0.101)	0.207 (0.115)	-0.001
Prop. SEN (statement)	0.020 (0.023)	0.016 (0.018)	0.004***
Prop. EAL	0.070 (0.106)	0.067 (0.101)	0.003
Prop. non-white	0.162 (0.121)	0.160 (0.133)	0.002
Number of Pupils	259.82 (119.8)	250.17 (110.2)	9.647
IMD Rank	0.731 (0.172)	0.704 (0.174)	0.026**
IDACI Rank	0.656 (0.173)	0.649 (0.184)	0.007
Pseudo R-Squared		*0.054*	
Likelihood Ratio Test (p-value)		*<0.01*	
B) Funding and Expenditure			
Total grant funding per pupil (log)	8.201 (0.179)	8.196 (0.190)	0.005
Total expenditure per pupil (log)	8.195 (0.179)	8.192 (0.195)	0.003
C) KS2 Outcomes			
English fine points score (std)	0.123 (0.356)	0.104 (0.357)	0.019
Maths fine points score (std)	0.084 (0.348)	0.098 (0.335)	-0.013
Number of observations	600	677	
Number of Schools	120	136	

Notes: *** denotes where difference between schools on inside and outside of boundary are significant at the 1%, ** at 5%, and * at 10% level. Standard deviations are in parentheses. The unit of analysis is the school. The sample includes all primary schools that are present in the National Pupil Database in all of the academic years 2005/2006, 2006/2007, 2007/2008,2008/2009, 2010/2011 and where the school is within 2km of the fringe London pay boundary. The likelihood ratio tests the null hypothesis that the differences in school characteristics are jointly zero. The Pseudo R-squared is taken from a probit regression of an indicator of whether schools are in the high-pay area on the set of school characteristics reported in panel A.

[8] We make use of the School Workforce Census in 2011 only, given uncertainty regarding the quality of data in 2010.

[9] The choice of distance does not change this finding. See Table A1 and Table A2 for 1km and 3km respectively. Figures are pooled across all years. Changes over time in these characteristics for schools within 2km of the boundary are presented in an online appendix (Figure A3)

At the fringe area boundary, it is clear that schools are largely balanced in observable characteristics. Although a likelihood ratio test suggests that we should reject the null hypothesis that the differences are jointly zero, the actual differences are very small in absolute value. The low value of the pseudo R-squared (0.054) also suggests that these covariates explain very little of the variation in terms of whether schools are in high or low pay areas. Furthermore, there are no clear differences for individual years or differential trends across time (Figure A3).

Panel B shows that there are no significant differences in raw funding levels at the fringe boundary, which suggests schools must pay these higher teacher salaries within fixed budgets. Therefore, there is a relatively good balance in observable characteristics along the fringe boundary and no differences in school budgets.

To give a brief preview of our main findings on student achievement, Panel C shows the small raw differences in average student achievement at schools either side of the boundary, measured by the average score of pupils in age 11 tests in Maths and English (standardised at the national level each year).

2.5 Empirical Analysis

This section presents our main empirical results for the differences in resources, teacher characteristics and student achievement at the fringe London pay boundary.

2.5.1 Resources

Table 2.2 shows the differences in measures of total funding and expenditure per pupil between schools either side of the fringe pay boundary (and controlling for observable characteristics using FILM). This is shown for schools within 1, 2 and 3km of the boundary. The first row shows there are small and generally statistically insignificant differences in total grant funding per pupil (always less than £56 per pupil). The second row then examines differences in total income per pupil, which incorporates any self-generated income (either through parent donations or renting out facilities). Such income is minimal for most schools, and differences in total income per pupil are again small and generally statistically insignificant. In principle, income need not exactly equate to expenditure each year as schools can run small surpluses and deficits. The third row, however, confirms that there are no large differences in total expenditure per pupil. The difference in total expenditure for schools within 3km is statistically significant at the 10% level. However, this is small and actually suggests that schools on the high-pay side of the boundary spend less in total (£75 per pupil). Therefore, these results provide no evidence to suggest that schools on the high pay side of the boundary experience any increase in funding to compensate them for having to pay teachers according to higher salary scales.

Table 2.3 shows how schools on the high-pay side the boundary adjust resources in light of the higher salary scales and no difference in funding. We start by examining teacher remuneration (panel A). Note that data is only available for 2011 here. The first row shows the difference in base salary levels across the boundary (i.e. before any additional payments). For our preferred specification, the estimated difference in base salary levels (£550) is slightly below the difference in salary scales (around £1,000 for 2011). When we include all payments (including additional payments above base salary) in row 2, these differences more closely match the difference in salary scales (with estimates ranging from £645 to £1,130). In row 3, we

Table 2.2: Difference in funding and expenditure across Fringe/Rest of England Boundary 2006 to 2011: various distances to pay boundary

Outcome	(1) **Within 1km**	(2) **Within 2km**	(3) **Within 3km**
Grant funding per pupil (£)	-5.55	-22.63	-55.54
	[41.56]	[31.53]	[26.61]*
Total income per pupil (£)	-23.40	-36.90	-58.16
	[47.44]	[33.60]	[28.39]*
Total expenditure per pupil (£)	-67.56	-69.10	-74.97
	[48.03]	[36.65]	[29.53]*
School and Year Controls		Yes	
Observations (schools)	542 (111)	1,152 (236)	1,651 (341)

Notes: *** denotes significance at 1%, ** at 5%, and * at 10% level. Standard errors clustered at the local authority level. All columns report Fully Interacted Linear Matching estimates. The unit of analysis is the school. The sample includes all primary schools that are present in the National Pupil Database in all of the academic years 2005/2006, 2006/2007, 2007/2008,2008/2009, 2010/2011 and where each dependent variable is observed, within 1km, 2km or 3km of the fringe London pay boundary. The coefficient reported represents the increase in income or expenditure per pupil associated with the high-pay side of the boundary. School controls include distance to the boundary, characteristics of the school: percentage of pupils eligible for free school meals, with English as an additional language, that are non-white, and have a special educational needs, the number of pupils in the school, dummy variables for region (North-East London, South-East London, South-West London), rank of index of multiple deprivation and rank of income deprivation affecting children index.

examine whether there are any differences in the average salary scale point of teacher (a good proxy for teacher experience given the largely mechanical relationship between teacher salary scale and experience in operation over this period). There is no evidence of any difference in average salary scale point across the boundary. This suggests more experienced teachers do not sort into schools on the high pay side of the boundary, either through the supply or demand side.

Panel B investigates whether schools adjust staffing levels at the pay zone boundary. This shows that there are small and statistically insignificant differences in pupil:teacher ratios for schools within 1km and 2km of the pay boundary, but a significantly higher pupil:teacher ratio for schools within 3km of the pay boundary. This provides some evidence that schools choose to employ relatively fewer teachers on the high pay side of the boundary, but the estimates are not stable across distance cut-offs.

The lack of consistent evidence for a reduction in teacher numbers suggests that either the substitution or income effects are weak, or that schools are unable to reduce teacher numbers due to regulation (e.g. maximum class sizes for under 7s). Given that expenditure on teachers is relatively lumpy, it might also be hard for schools to reduce teacher expenditure by a small amount.

In the next row, we look at numbers of teaching assistants per pupil. Teaching assistants are generally low-skilled staff who assist teaching during lessons or with administrative tasks (Sibieta (2015b)). Although not statistically significant, the point estimates for schools within 1km and 2km of the pay boundary are consistent with schools choosing higher ratios of pupils to assistants on the high-pay side of the boundary, suggesting that schools might be responding by cutting assistant numbers. However, as for the pupil:teacher ratio, schools within 3km of the boundary exhibit a different pattern.

Table 2.3: Difference in input choices across Fringe/Rest of England Boundary 2006 to 2011: various distances to pay boundary

Outcome	(1) **Within 1km**	(2) **Within 2km**	(3) **Within 3km**
A) Teacher Remuneration (2011 only)			
Teacher Salary (£)	632.55	550.82	537.22
	[991.63]	[601.9]	[461.68]
Teacher Total Pay (£)	1129.67	687.19	645.42
	[1540.39]	[781.32]	[582.89]
Average salary Scale Point (1-9)	-0.11	0.10	0.03
	[0.24]	[0.28]	[0.21]
B) Staff Ratios (years pooled)			
Pupil:Teacher Ratio	0.11	-0.02	0.38
	[0.33]	[0.24]	[0.19]*
Pupil: Assistant Ratio	22.56	14.20	-12.79
	[16.63]	[12.23]	[13.23]
C) Spending on different inputs (per pupil) (£)			
Teachers	29.00	34.33	11.58
	[32.02]	[22.94]	[18.57]
Teaching Assistants	14.20	-25.82	-21.79
	[20.33]	[18.40]	[13.39]
Other staff	3.36	-20.78	-16.63
	[15.32]	[10.80]	[8.80]
Services	-43.96	-10.83	-14.48
	[16.72]**	[11.74]	[10.59]
Equipment	-49.55	-29.64	-27.25
	[16.16]**	[10.34]**	[9.23]**
School and Year Controls		Yes	
Observations (schools)	542 (111)	1,152 (236)	1,651 (341)
2011 Schools	100	232	329

Notes: *** denotes significance at 1%, ** at 5%, and * at 10% level. Standard errors clustered at the local authority level. All columns report Fully Interacted Linear Matching estimates. The unit of analysis is the school. The sample includes all primary schools that are present in the National Pupil Database in all of the academic years 2005/2006, 2006/2007, 2007/2008,2008/2009, 2010/2011 and where each dependent variable is observed, within 1km, 2km or 3km of the fringe London pay boundary. The coefficient reported represents the change in the resource margin associated with the high-pay side of the boundary. School controls include distance to the boundary, characteristics of the school: percentage of pupils eligible for free school meals, with English as an additional language, that are non-white, and have a special educational needs, the number of pupils in the school, dummy variables for region (North-East London, South-East London, South-West London), rank of index of multiple deprivation and rank of income deprivation affecting children index.

Panel C then examines spending per pupil on teachers, teaching assistants, other staff, services and equipment. Here, there is suggestive evidence of slightly higher spending per pupil on teachers (reflecting higher salaries and little change in quantity) and typically a small reduction in spending on teaching assistants and other staff, per pupil. Spending on services per pupil is also lower on the high-pay side of the boundary (although not significantly so) which is noteworthy as this category of expenditure includes items that are typically difficult to shift (for example energy and catering expenditure). There are significant reductions in spending on equipment (learning resources, information communication technology and spending on the school premises) per pupil on the high-pay side of the boundary, around £30 for schools within 2km of the boundary. These differences equate to reductions in other expenditure per pupil of about 3% and 6%, respectively.

Therefore, there is no evidence of any additional funding or total expenditure for schools on the high-pay side of the fringe boundary. Despite this, there is also no evidence of pay smoothing, with differences in actual teacher pay roughly in line with the salary scales. There is also no evidence of any differences in the composition of teachers in terms of their salary scale point (a good proxy for experience) or any changes in teacher numbers. Instead, there are small reductions in numbers of assistants per pupil and much larger reductions in other expenditure (particularly equipment) per pupil. Schools thus seem to respond to the higher salary scales and fixed budgets by paying the higher salaries and cutting non-instructional expenditure in order to do so.[10]

2.5.2 Mechanisms

Table 2.4 presents evidence for positive mechanisms which might offset the negative impact of lower non-teacher spending in schools on the high-pay side of the boundary. Unfortunately it is not possible to directly study teacher quality on either side of the boundary as administrative data in England has no link between teachers and pupils.[11] To provide some evidence of teachers' responses to the variation in pay levels, however, Table 2.4 shows two measures of teacher absence and two measures of teacher turnover. The proportion of teachers that are absent at the time of administrative data collection is around 6 percentage points, or 12%, lower on the high pay side of the boundary. The average level of absence is also lower in schools on the high-pay side of the boundary, although not significantly so. This may suggest that teachers on the high-pay side of the boundary respond to efficiency wages, or that teacher well-being is higher. The variation in teacher absence does not appear to translate into lower teacher turnover, however, with the proportion of new entrants to the school not significantly different across the boundary.

2.5.3 Student Achievement

Table 2.5 shows the estimates for the differences in average pupil attainment in Maths and English. These are measured in terms of national standard deviations, so can be interpreted in effect size terms. This shows little evidence of any positive difference in student achievement at the fringe pay zone boundary. For English, the point estimate is 0.015 standard deviations for our preferred specification of being within 2km of the pay

[10]This contrasts with Cruz (2018) who finds that school districts in Brazil only increase teacher salaries when they receive extra funding. However, the context of these findings is somewhat different. Cruz (2018) considers the effect of a threshold for minimum spend on teachers combined with a funding reform. We consider the effects of higher minimum salaries for teachers.

[11]The administrative data contains some information on the highest level of qualification of teachers, but this information is incomplete, and a relatively poor proxy for teaching quality.

Table 2.4: Difference in teacher responses (2011): various distances to pay boundary

Outcome	(1) Within 1km	(2) Within 2km	(3) Within 3km
A) Teacher absence (2011 only)			
Mean absence	-1.42	-0.81	-0.72
	[0.90]	[0.64]	[0.49]
Prop. teachers absent	-5.50	-5.95	-5.56
	[4.80]	[2.88]*	[2.31]*
B) Teacher turnover (2011 only)			
Prop. teachers < 1 year	0.026	-0.006	-0.006
	[0.029]	[0.019]	[0.016]
Prop. teachers < 2 years	-0.007	-0.016	-0.007
	[0.041]	[0.028]	[0.022]
2011 Schools	100	232	329

Notes: *** denotes significance at 1%, ** at 5%, and * at 10% level. Standard errors clustered at the local authority level. All columns report Fully Interacted Linear Matching estimates. The unit of analysis is the school. The sample includes all primary schools that are present in the National Pupil Database in all of the academic years 2005/2006, 2006/2007, 2007/2008,2008/2009, 2010/2011 and where each dependent variable is observed, within 1km, 2km or 3km of the fringe London pay boundary. The coefficient reported represents the change in the resource margin associated with the high-pay side of the boundary. School controls include distance to the boundary, characteristics of the school: percentage of pupils eligible for free school meals, with English as an additional language, that are non-white, and have a special educational needs, the number of pupils in the school, dummy variables for region (North-East London, South-East London, South-West London), rank of index of multiple deprivation and rank of income deprivation affecting children index.

zone boundary and an alternative specification within 3km of the boundary, and slightly higher (0.033) for 1km.

The point estimates for Maths are generally very small across all specifications, with an estimate of -0.007 for our preferred specification. The estimates are also relatively precise. The 95% confidence intervals implied by our estimates mean that for our preferred specification we are able to rule out quantitatively small effects of 0.056 and 0.036 standard deviations in English and Maths, respectively.

Combined with our previous results, we find no empirical evidence that the offer of higher teacher salaries combined with reductions in other expenditure have any positive implications for student achievement. This suggests that schools optimally adjust other resources in response to the pay rules and lack of funding.

This could be because a positive effect of higher teacher salaries (perhaps through lower teacher absences or efficiency wage mechanisms more generally) is exactly canceled out by a negative effect of reductions in other resources. Recent evidence suggests that school resources have a significant impact on pupil attainment (Jackson (2018)). We therefore hypothesise that schools efficiently balance the negative impact of reductions in other spending with the positive impact on teachers from marginally higher salaries.

As we have argued, however, it is the total effect that is the most relevant policy parameter for a school considering whether or not to offer higher salaries from within a fixed budget.

Table 2.5: Difference in student achievement across Fringe/Rest of England Boundary 2006 to 2011: various distances to pay boundary

Outcome	(1) Within 1km	(2) Within 2km	(3) Within 3km
KS2 Fine Points Score (std)			
English	0.033	0.015	0.015
	[0.031]	[0.021]	[0.018]
Maths	0.018	-0.007	-0.006
	[0.028]	[0.022]	[0.020]
School and Year Controls		Yes	
Observations (schools)	542 (111)	1,152 (236)	1,651 (341)

Notes: *** denotes significance at 1%, ** at 5%, and * at 10% level. Standard errors clustered at the local authority level. All columns report Fully Interacted Linear Matching estimates. The unit of analysis is the school. The sample includes all primary schools that are present in the National Pupil Database in all of the academic years 2005/2006, 2006/2007, 2007/2008,2008/2009, 2010/2011 and where each dependent variable is observed, within 1km, 2km or 3km of the fringe London pay boundary. The coefficient reported represents the change in the KS2 fine point score in each subject (standardised at the national level) associated with the high-pay side of the boundary. School controls include distance to the boundary, characteristics of the school: percentage of pupils eligible for free school meals, with English as an additional language, that are non-white, and have a special educational needs, the number of pupils in the school, dummy variables for region (North-East London, South-East London, South-West London), rank of index of multiple deprivation and rank of income deprivation affecting children index.

2.5.4 Robustness checks

Tables A3 and A4 show our estimates of the effect of the pay differential at the fringe boundary for different measures of distance to the boundary, in raw and conditional terms (using a range of ways to control for observable characteristics) and across all years for Maths and English, respectively. This shows that the differences in student achievement are largely stable over time, and are qualitatively unchanged by how and whether we control for observable characteristics.

Figures A4 and A5 estimate the raw difference in student achievement over a much longer time horizon for schools within 2km of the boundary (1995 to 2011). Importantly, the pay zone boundary existed for all years covered by this data. As this makes use of older school-level data, we have to use a different measure of student achievement (the proportion of pupils achieving the expected level in English and Maths). Although the estimates are clearly more imprecise, the point estimates remain close to zero throughout the period. This suggests the long-term equilibrium we observe in the later 2000s has persisted since at least the early 1990s.

Figures A6 and A7 replicate local linear regression methods recommended by Lee and Lemieux (2010) for regression discontinuity designs. This illustrates how outcomes vary with distance to the boundary. In particular, we show the local averages for schools in bins of size 200m either side of the fringe boundary (black dots) up to 3 km from the boundary, as well as estimates of the relationship between distance to the boundary and each outcome based on a linear specification (dashed line) and a 7th order polynomial (solid line), each with a break at the discontinuity. We show this for English (Figure A6) and Maths (Figure A7), with data pooled across years. In both cases, there is no clear or consistent relationship between test scores and distance from the pay boundary (at least within 3km either side of the pay boundaries). The relationship between distance to the boundary and attainment is best described by a flat line with noise, with the high order polynomial oscillating around the linear estimates. Indeed, in a linear regression we are unable to reject the null hypothesis that the slope coefficients on distance are zero either side of the boundary. There is also little evidence to suggest a positive jump at the pay boundary. This flexible approach confirms our main finding that there is no difference in student achievement at the pay zone boundary.

2.6 Conclusion

In response to increasing school autonomy and funding constraints, this paper provides the first evidence of the impact of increasing teacher salaries (and decreasing other spending) on pupil attainment. We use a natural experiment that forces some schools within a small local area to offer higher salaries with no compensating change in school budgets. Teacher pay follows the regulation, despite opportunities for schools to undo the difference in salary scales, suggesting that schools have some monopsonistic wage-setting powers. Schools do not, however, cut back on teacher numbers. Instead, schools use their budgetary autonomy to make large reductions in non-instructional spending to provide the higher teacher salaries, which reveals information about schools' resource preferences.

We find no evidence of teacher sorting effects, measured by the average level of experience or higher teacher turnover. However, we do find evidence of potential efficiency wage effects through reduced teacher absence, equivalent to around a 12% reduction.

The total combined effect on student achievement of higher teacher salaries, reduced teacher absence and reduced non-instructional spending is estimated to be very close to zero. Schools seem to adjust resources optimally in response to the pay regulation and no extra funding, effectively moving around the Pareto frontier. Given recent evidence showing a positive effect of school resources on human capital (Jackson et al. (2015); Jackson (2018)) and negative effects of teacher absence (Herrmann and Rockoff (2012)), we interpret our results as representing offsetting negative resource and positive effects from reductions in teacher absence.

This suggests that offering small increases in teacher salaries from within a fixed budget are unlikely to be a good use of new freedoms on teacher pay for schools in England or the US. This finding is relevant both to autonomous and regulated schools in England, particularly given that all schools have similar freedoms on non-instructional spending and have had flexibility in teacher pay awards since 2013. The results are generalisable to educational systems with similar flexibilities and comparable levels of teacher pay relative to outside wages.

Comparing our results with previous work on teacher pay (Loeb and Page (2000); Gilpin (2012); Leigh (2012); Hendricks (2014); Britton and Propper (2016a)) suggests that teacher quality is more sensitive to pay levels when individuals make their occupational choices, rather than when they decide where to teach. In future work, it will be important to understand the reasons driving why the sorting of teachers across schools might not be sensitive to pay levels. This could be because teacher decisions are more sensitive to pupil characteristics or non-pecuniary rewards (Hanushek et al. (2004); Bonesronning et al. (2005)) or because potential teacher quality is not observable among a pool of applicants (Delfgaauw and Dur (2007); Bó et al. (2013)). Investigating this explanation would require a more detailed understanding about how schools make hiring decisions by collecting data on the characteristics of the pool of applicants for individual teacher positions. Furthermore, although we observe no difference in student achievement at the pay boundary, it might be that changes in non-instructional spending have implications for other non-achievement outcomes, for example pupil behaviour or health, which would require further data collection and research.

Chapter 3

The effect of cash incentives on the number of new teachers and their aptitude

Abstract

Good teachers are important for the production of human capital. There are significant problems in attracting teachers in some subjects, particularly maths and science. This is almost certainly due to the lack of variation in teacher wages across subjects. Existing evidence has shown how cash incentives can be used to retain teachers in specific subjects and how they can be used to attract teachers to specific types of schools. However, there is little evidence regarding the effectiveness of incentives to recruit new teachers in hard-to-staff subject areas. In this paper, I evaluate the effect of an up-front cash payment worth up to £25,000 for teachers training in hard-to-staff subjects and who have high levels of college attainment. Using a triple-difference approach, I find no impact on the number of teachers or the distribution of educational attainment among teachers in hard-to-staff subjects.[1]

JEL Classifications: I2, J3, J4

Keywords: Teacher Wages, Teacher Quality, Incentives

[1] I am grateful to the ESRC for financial support as part of the UCL IoE Doctoral Training Centre and the ESRC Centre for the Micro economic Analysis of Public Policy at IFS (grant number RES-544-28-0001). I am also grateful to the Office for Manpower Economics for funding an earlier piece of research that led to this paper. I am also grateful to Neil Amin-Smith and Ellen Greaves for their contribution to the earlier piece of work, and to Chris Belfield, Jack Britton, Nikki Shure, Alex Bryson and RES conference participants for comments and feedback. This paper makes use of three sources of data provided by the Higher Education Statistics Agency (HESA). In particular, the HESA Student record 2006/07–2014/15, the HESA Destinations of Leavers from Higher Education record 2006/07–2014/15 and the HESA Longitudinal Destinations of Leavers from 2006/07, 2008/09 and 2010/11. These data are Copyright Higher Education Statistics Agency Limited. Neither the Higher Education Statistics Agency Limited nor HESA Services Limited can accept responsibility for any inferences or conclusions derived by third parties from data or other information supplied by HESA Services. All errors remain my own.

3.1 Introduction

The presence of good teachers can have a substantial effect on the stock of human capital. Chetty et al. (2014) find that replacing bad teachers with average teachers can improve students later life income by $250,000 per classroom over a teacher's lifetime. Unfortunately, it is hard to predict who is likely to be a good teacher on the basis of observable characteristics, with the important exception of detailed measures of past educational attainment (Ehrenberg and Brewer (1994); Wayne and Youngs (2003); Kane et al. (2008)). This creates an interest in how incentives can be designed to recruit and retain high-quality teachers. Here, evidence has shown that comparisons between teacher salaries and outside options can have meaningful effects on the overall number and quality of teachers. Increases in teacher salaries have been shown to have a positive effect on average teacher skill levels (Leigh (2012); Nickell and Quintini (2002)) and their overall quality through their ability to raise student achievement (Loeb and Page (2000); Britton and Propper (2016b)). At the same time, compression in the teacher wage distribution over time has reduced the number of highly skilled teachers in the profession (Hoxby and Leigh (2005)).

There are particular challenges in recruiting and retaining in teachers in some specific subjects. Empirical evidence suggests that graduate earnings vary significantly according to field of study (Britton et al. (2016b); Kirkeboen et al. (2016)), with higher averages wages and better opportunities for highly-skilled individuals in subjects such as physics, maths and economics. In contrast, the teacher wage distribution is generally more compressed in most developed countries, with little variation across subject or based on skill level (Allen and Sims (2018)). This makes it particularly hard to recruit and retain teachers in areas where graduates have much better outside options. For example, in 2015-16, the UK National Audit Office (NAO) reported that the proportion of teacher training places filled was 71% in physics, 70% in computing and 87% in modern foreign languages (NAO (2016)). Such targets have been missed persistently over time and as a result, pupils are much more likely to be taught by a teacher without a college degree in that subject. For example, around half of physics and maths teachers had a degree in that subject in 2016, compared with 80% of English and history teachers (Sibieta (2018)). Such figures raise obvious concerns as to whether all pupils are receiving a good quality of education in shortage subjects. Such concerns are not specific to England, with shortages in maths and science subjects reported across a wide range of countries.

With these concerns in mind, policymakers in England in 2012 greatly increased cash payments made to teachers training in hard-to-staff subject areas and introduced even larger payments for those with higher college attainment, with the highest cash payments increasing from £9,000 to £25,000 for a highly skilled teacher in hard-to-staff subjects. In this paper, I use a triple-difference approach based on this policy change to estimate the effect of up-front cash payments on the number and skill level of new teachers in hard-to-staff subject areas.

There is now a wide literature estimating the effects of recruitment and retention cash incentives for teachers, which vary in size and scope Clotfelter et al. (2008) find that a $1,800 bonus for maths, science and special education teachers can reduce teacher attrition by 17% in high-poverty schools in North Carolina. Between 1984 and 2001, Florida implemented student loan forgiveness (up to $2,500 annually or $10,000 over a lifetime) and tuition cost reimbursement (up to $700 annually or $2,800 over a lifetime) schemes for teachers in shortage subjects. One-time retention bonuses of up to $1,200 were paid for teachers in shortage subjects in the year 2000. Feng and Sass (2018b) shows that the loan forgiveness programme reduced teacher attrition by about about 11% for maths teachers and 9% for science teachers. Given the average annual award was about $1,200, these results are fully in line with the effects estimated in North Carolina. They find even larger effects of the one-time bonuses of $1,200 in 2000, which reduced attrition by up to 25 per cent. The tuition reimbursements were also found to be effective in terms of boosting the probability of eligible teachers to become certified to teach in shortage subjects.

In California, the Graduate Teacher Fellowship offered $20,000 to high-ability individuals who had trained as a teacher and who committed to teach in high poverty schools for four years, which increased participants willingness to teach in such high-poverty schools by around 28% (Steele et al. (2010)). There is also clear evidence that incentives for highly effective teachers can boost retention in high-poverty schools (Dee and Wyckoff (2015); Springer et al. (2016)) and that cash incentives can increase teacher certification rates (Cowan and Goldhaber (2016)). There is therefore relatively consistent evidence that cash incentives can boost retention rates in hard-to-staff or shortage subjects, and that they can be used to attract teacher in high-poverty schools.

All such evidence, however, relates to existing teachers or those that have already gone through training. There is relatively little evidence on the effectiveness of cash incentives as a means to increase recruitment to teaching, particularly in hard-to-staff subjects. The effects of cash incentives on new and existing teachers could also differ due to the fact that the marginal teacher is likely to be different in each case. In the case of existing teachers, the marginal individual is likely to be someone who has built up specific human capital as a teacher and who has already revealed themselves to have a disposition towards a career in teaching. In this case, incentives are designed to either counteract negative aspects of working conditions (e.g. teaching at a high poverty school) or stopping teachers in shortage subjects from leaving for a new non-teaching job. In contrast, the marginal teacher targeted by recruitment incentives is someone who would otherwise have chosen a non-teaching occupation. Such individuals will have built up less specific human capital as teachers and probably have less natural disposition to become a teacher. The results of this paper therefore go beyond the existing literature by considering the recruitment margin instead of retention and school switches.

As in most countries, teachers in England specialise to teach specific subjects for children in secondary or high school. The majority of such teachers (60%) join after taking a one year teacher training course following a standard 3-year college degree. Teachers specialise in specific subjects, which is usually a subject they studied in college or strongly related to it (e.g. economic college graduates can qualify to teach maths). During this one year of training, teacher trainees are entitled to a bursary to help cover living costs. Up to 2012, teacher trainees were entitled to bursaries of either £9,000, £6,000 or nothing depending on whether their subject of study was considered a priority or not, with no variation by their level of college attainment. From 2012, bursaries were substantially increased for high priority subjects, with higher amounts now on offer for trainees with high levels of college attainment. For example, from 2012 trainee maths and most science teachers with a first class degree (highest level of college attainment) could receive up to £20,000, with gradually lower amounts for lower levels of college attainment, down to zero for those who achieve just above the pass mark. There were also smaller increases for lower priority subjects and the introduction of variation by college attainment for these lower priority subjects too. Trainee teachers in non priority subjects continued to receive no bursary. I exploit changes in these bursaries over time, subject and level of college attainment to estimate the impact of cash incentives on individuals' willingness to train as a teacher and prior educational attainment of new teachers.

First, I study the effect of the change in bursaries on individuals' willingness to train as teacher by using a triple-difference. I examine if the gradient in willingness to train as a teacher by college attainment increased disproportionately in high-priority subject areas as compared with lower priority subject areas. This shows that prior to the introduction of bursaries, individuals with degrees in non priority subjects were more likely to become a teacher if they had high levels of college attainment, with the opposite true for high priority subjects. This is unsurprising given the better labour market opportunities for individuals with college degrees in shortage subjects. The propensity of individuals with the highest levels of college attainment to become a teacher increases over the period covered by the data (2006 to 2014). However, there is no evidence that the gradient by college attainment increases more for higher priority subjects following the introduction of the higher value bursaries in 2012. This suggests that up front cash incentives have little, if any, impact on the propensity of highly skilled individuals to train as teachers in shortage subjects.

Second, I estimate the effect of the change in bursaries on individuals' educational attainment on entry to college (as a proxy for their overall skill level). This relies on a difference-in-differences approach by subject, comparing the periods before and after the shift in bursaries in 2012. This shows that prior to the change in bursaries in 2012, average educational attainment amongst trainees was actually higher in shortage subjects like physics and maths than in non-priority subjects like history and English. This is largely because the average college graduate in shortage subjects had higher level of educational attainment. In fact, I find that the distribution of educational attainment among teachers in shortage subjects closely resembles the distribution among all college graduates in that subject. Whilst there is a quantity problem in shortage subjects, there is no evidence of a quality problem. Following the change to bursaries in 2012, there is no evidence of any differential change in the distribution of educational attainment across subjects. Therefore, as well as finding no evidence of an effect of cash payments on the quantity of new teachers, I find no impact on their skill levels either.

In the first case, identification relies on there being no differential shocks or policy changes by subject and college attainment that also take effect from 2012 onwards. In the second case, it relies on a stronger assumption that there are no differential shocks by subject over time. I present evidence that suggests both are valid. Whilst there are differences in average graduate earnings and variation in graduate earnings by subject of study, these relative differences have remained steady over time. Changes in tuition fees in 2012 were common across students and are only likely to have a small effect on expected student loan repayments given the amount of debt accumulated before starting a teacher training course. The main policy change that could represent a threat to identification is the expansion of on-the-job training routes over time, which have expanded from 23% of secondary school trainees in 2010 to about 36% in 2014. However, this growth has been gradual over time, rather than representing a decisive shift in 2012. Furthermore, I show that the college attainment of new teachers by subject is relatively similar across training routes. A final potential problem is that the data only records individuals' college major, rather than the subject they train to teacher. This is a potential source of bias if bursaries encouraged individuals with college degrees in other subjects to train in high priority subjects (e.g. biology graduates to train as chemistry teachers). However, numbers of trainees and their skill levels are largely stable across individual subjects, suggesting this is unlikely to be a source of bias.

A further potential limitation of the analysis is the relatively wide confidence intervals around individual estimates, given the need to use a wild cluster bootstrap approach to conduct inference with relatively few clusters (15 subjects). However, the stability of the results by subject and subject grouping by individual year, and the narrower confidence intervals based on pre and post period analysis, lend credibility to results.

The main implication of the results is that up front cash incentives are likely to be an ineffective means to attract new teachers to shortage subjects or as a means to increase their skill levels. In 2014-15, such incentives cost around £145m in England and probably represent poor value for money. Other empirical evidence suggests that retention incentives are likely to be more effective.

The differential effectiveness of recruitment and retention incentives are relatively easy to rationalise. Recruitment incentives are effectively targeted at all graduates, only a small group of whom will become teachers (less than 2% six months after graduation). Incentives would thus need to very large indeed to induce a significant change in the proportion willing to become a teacher. Retention incentives are targeted at a much smaller set of existing teachers who have already revealed a disposition towards teaching. There is also more uncertainty about future rewards at the point of graduation, particularly in terms of non-wage benefits (e.g. working conditions). Retention incentives can therefore be used to counteract known assessments of the benefits of teaching against outside options.

The rest of this paper proceeds as follows. Section 3.2 gives details of the institutional background for teacher training, bursaries and fees in England. Section 3.3 sets out the empirical framework for estimating the introduction of high-value, targeted bursaries on the quantity and aptitude of teacher trainees. Section 3.4 provides descriptive evidence on the quantity and aptitude of individuals training to be teachers in different subjects, and compares this with labour market opportunities for graduates in different subjects. Section 3.5 presents the main empirical results and discusses their policy implications. Section 3.6 concludes.

3.2 Teacher training and bursaries in England

Individuals wanting to teach a specific subject in secondary schools in England (covering ages 11-18 and thus similar to high schools in the US) generally have to complete an additional year of training or education after a college degree in a specific subject[2]. The most commonly used route is a one-year course called a Postgraduate Certificate of Education (PGCE) at a higher education institution. This and other formal education routes accounted for about 64% of those training to be secondary school teacher in 2015. This is down from around 77% in 2010 due to growth in employment-based routes that combine on- and off-the-job training. For example, in 2012 the government replaced the existing employment-based route with a new one called 'School Direct', which continued the growth in employment-based training routes. Another high-profile employment-based route that has grown over time is Teach First, which aims to attract high-attaining graduates, provide them with an intensive course and place them in challenging settings. This was introduced in the early 2000s and the number teach first trainees was about 1,600 in 2015 , up from about 400 in 2008 and 700 in 2011. However, most of the recent growth can be accounted for by growth in primary school trainees, which were only recruited from 2011 onward and represented about 500 in 2015[3].

Over the last few years, the government has persistently missed targets for the number of teacher trainees, with the problem worse in some specific subjects. For example, in 2016, the National Audit Office reported that the proportion of teacher training places filled was 71% in physics, 70% in computing and 87% in languages (NAO (2016)). Such targets have been missed persistently over time and as a result, pupils

[2]Those who want to teach at primary school level in England (ages 4-11) can undertake a specific undergraduate degree to become a qualified teacher, which accounted for about 30% of new primary school trainees in 2015-16. However, those who want to teach a specific subject at secondary school level (ages 11-18) almost all complete some kind of postgraduate education or training, with only about 5% undertaking undergraduate teacher training.

[3]https://www.gov.uk/government/statistics/initial-teacher-training-trainee-number-census-2015-to-2016

learning such subjects are much more likely to be taught by a teacher without a degree qualification in that subject. For example, around half of physics and languages teachers had a degree in that subject in 2016, compared with around 80% of English and history teachers (Sibieta (2018)). This raises obvious concerns as to whether all pupils are receiving a good quality of education in shortage subjects. Such concerns are not specific to England, with shortages in maths and science subjects reported across almost all US states[4]

3.2.1 Bursaries

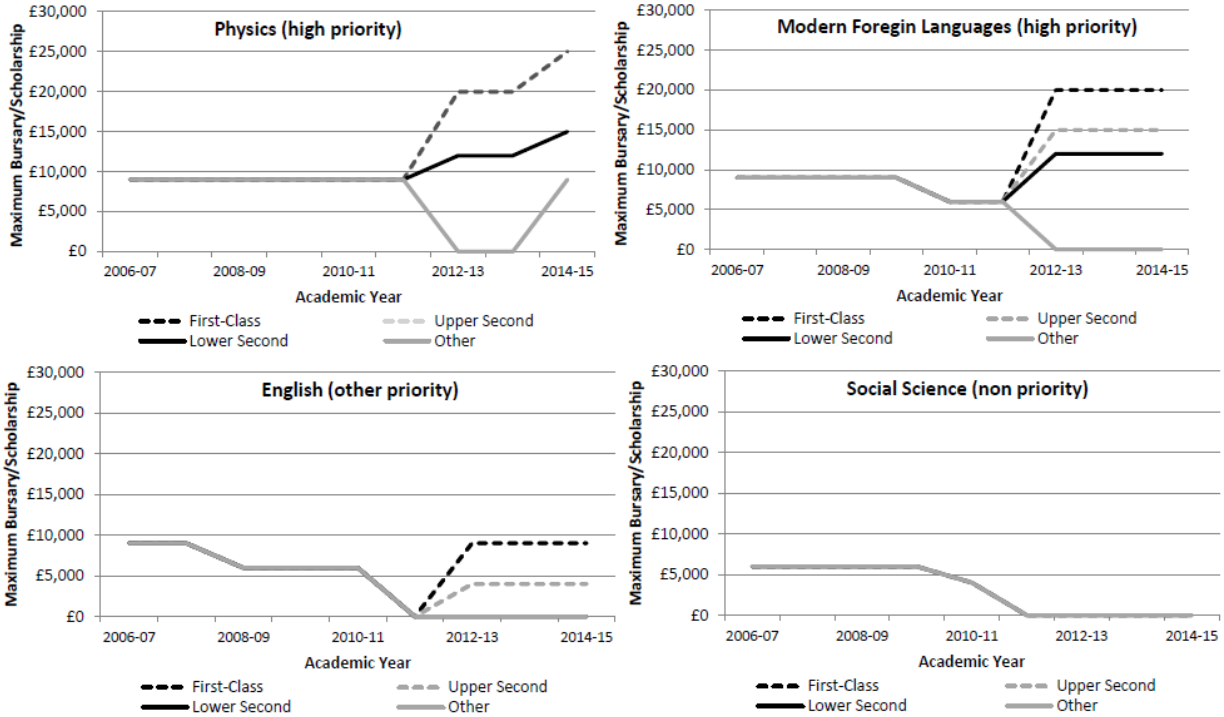
The main policy targeted at addressing such shortages in England has been the substantial expansion of – and increased variation in - bursaries for teachers whilst in training. One can view these bursaries as targeted reductions in the costs of study to become a teacher in a particular subject and as a means to counteract the higher outside wages on offer to teachers in shortage subjects. Bursaries vary by subject and the amount is determined by which subject students are training to teach, which generally ties closely to actual subject of study for secondary school teachers. They are only available to individuals training through postgraduate training routes in formal education (i.e PGCE trainees and those in the unsalaried version of School Direct).

The level and variation in these bursaries has changed substantially over time. Figure 3.1 shows the maximum level of bursaries/scholarships individuals would have been entitled to for selected subjects that exemplify the variation over time and subject (Physics, Modern Foreign Languages, English and Social Science). In 2006, teachers training in priority subjects were eligible for £9,000, whilst those training in other subjects were eligible for £6,000. At this point, a large number of subjects were classified as "priority," including maths, science, English, drama, ICT, design and technology, modern languages, religious education and music. As a result, three of the four subjects we show here attracted a bursary of £9,000. Over the next few years, the range of priority subjects changed slightly over time (e.g. English ceased to be a priority subject in 2008). In 2010, an additional distinction was introduced such that subjects classed as "high priority" would be eligible for £9,000, "other priority" subjects would be eligible for £6,000 and "non priority" eligible for the lower amount of £4,000. In 2011, bursaries were not available for non-priority subjects and some subjects, such as English, lost priority status. Throughout the period, however, there was no variation at all by college attainment.

Wholesale reform of the bursaries system was then introduced for students starting in 2012, with substantial increases for trainees in high-priority subjects with high levels of college attainment. For example, students with a first class or upper second class college degree (the highest classifications) training as physics teachers were now eligible for around £20,000 compared with £9,000 the year before. Students with a lower second class degree also saw eligibility increase to around to £12,000. Those training in physics with lower degree classifications (third class or pass mark) ceased to be eligible. A similar pattern was introduced for other subjects with high-priority status like Modern Foreign Languages, and in this case the changes with the previous year are actually larger as they were coming from a lower base. Trainees in other priority subjects, like English, continued to be eligible, but only if they received an upper second class or higher and the amounts on offer were much lower (£9,000 or less). Trainees in other subjects were not eligible for any bursary at all, as compared with £6,000 before 2011. In all cases, the level of bursaries and priority status of subject was set with respect to expected difficulties in recruiting sufficient numbers of trainees in each subject (based on past experience, demographic trends and curriculum policy).

[4]https://www2.ed.gov/about/offices/list/ope/pol/ateachershortageareasreport2017-18.pdf

Figure 3.1: Maximum bursary/scholarship amounts over time for selected subjects



Notes and Source: Initial Teacher Training Bursaries Funding Manual (various years); Bursaries and Scholarships are all tax-free

The level and variation by subject and degree classification changed slightly over the next few years, with increases for high-priority subjects like Physics in 2014 (and the re-introduction of smaller bursaries for those with low degree classifications). However, the broad structure of high bursaries targeted at high-priority subjects and high degree classifications remained in place.

3.2.2 Fees

At the same time as the changes in bursaries, tuition fees were increased from £3,375 for students starting higher education courses in September 2011 to £9,000 for students starting courses in September 2012. Teaching grants were cut at the same time to leave overall resources largely unchanged. Whilst higher education institutions had the freedom to charge lower amounts, practically all charged these maximum fees before and after the change. These fees applied to both undergraduate and PGCE teaching courses alike (Britton et al. (2019)).

Students from the UK and EU were eligible for loans to cover these fees, with very generous repayment terms after graduation. Fees are repaid at the rate of 9% on income above £15,000 for students starting courses up to 2011 and 9% above £21,000 for courses starting after 2012, with outstanding loans canceled after 25 years for loans taken out up to 2011 and after 30 years for loans taken out after 2012. Interest was charged at the rate of inflation up to 2011, with a variable interest rate of inflation + 0-3% for loans after 2012 (depending on actual income). Further details about the changes are described in Crawford and Jin

(2014).

Students starting 3 year undergraduate courses up to 2011-12 (and thus potentially starting PGCE courses up to 2014-15) would still have accumulated student debt under the old system, but would have faced higher fees for PGCE courses starting from 2012-13. Undergraduates starting 3 year courses after 2012-13 would have faced the new system for all their studies.

Despite this very large change in fees in 2012, there are good reasons to believe that these changes would have had little impact on teacher training decisions, particularly the differences by subject and degree classifications, which represent the key differences from the perspective of identifying the impact of bursary changes. First, the generous repayment terms meant that very few students would have faced the upfront changes in fees (unless they chose not to take out a loan). Second, this change would have affected all leavers from higher education equally from 2012 onwards. Third, undergraduates leaving higher education will have already built up significant levels of student debt before embarking on a PGCE course.

Students starting courses up to 2011-12 (and thus graduating from courses up to 2013-14) could borrow up to about £8,300 per year including fee and maintenance loans (or £10,300 inside London), making for a potential total debt of about £25,000 before starting a PGCE course. Modeling by Allen et al. (2014) shows that teachers on a typical wage profile would only have paid off undergraduate and PGCE loans in their mid-40s under the pre-2012 system, just before the 25 year limit for the cancellation of debt. Students starting PGCE courses from 2012 onwards would therefore only begin to start repaying the increased fees from £3,375 to £9,000 in their mid-40s and might not even finish repaying this given the very generous repayment terms. Students starting PGCE courses from 2015-16 (the last period covered by the data used in the analysis) would have faced fees of £9,000 per year for their undergraduate courses too and would thus have built up student debt of about £15,000 or more per year (or over £45,000 in total). Allen et al. (2014) show that such students would be unlikely to repay their undergraduate loan over a typical career for a teacher and would therefore not be likely to face any extra repayments as a result of taking a PGCE course.

In principle, the fee changes in 2012-13 could have affected undergraduate decisions about where and what to study, and thus the distribution of graduates potentially starting PGCE courses in 2015-16. However, Britton et al. (2019) show that the distribution of students by subject changed very little in the first year of the new system, though subsequent changes were more dramatic as cheaper to teach courses in the arts, humanities and social sciences expanded in relative terms over time.

In summary, whilst the changes in fees in 2012 were substantial, the actual effects on decisions are likely to have been minimal. The changes affected everyone alike and teachers were already only just about repaying likely student debt under the old system before the cancellation of debt after 25 years. Any increases in repayments under the new system would only have led to increases in repayments for teachers during their late 40s and would thus be heavily discounted too.

3.2.3 Implications

The sudden and large change in bursaries in 2012-13 for students in specific subjects with particular degree classifications offers an opportunity to examine the effects of targeted financial incentives to increase the number and quality of teachers in specific subjects. The next section describes the empirical methodology and data sources before then examining whether there is evidence of a differential shift in the willingness of teachers with high degree classifications in high-priority subjects when the new more generous and targeted bursaries were introduced in 2012. I also examine whether there is a differential shift in the distribution of teacher skill level by subject when the more generous and differentiated bursaries were introduced in 2012.

3.3 Empirical Methodology

This section describes the main methods for estimating the effect of the introduction of high-value targeted bursaries on the propensity of individuals to train as a teacher and their skill level. The section starts by setting out the main estimating equations and identifying assumptions, before describing the data and empirical results in the following sections.

3.3.1 Number of trainees

The main approach for estimating the effect of bursaries on the propensity to train as a teacher can be summarised as examining whether the gradient by college degree classification in a high priority subject increases by more than it does for non-priority subjects after the introduction of high-value, targeted bursaries. In essence the approach is a triple-difference across subject, time and degree classification. The main estimating equations are shown in equations (3.1) and (3.2), where an individual individual (i) with a degree in subject (j) graduating from college or university at time (t) with degree classification (k) chooses to train as a teacher $T_{ijkt} = 1$ if and only if the latent variable $T^*_{ijkt} > 0$.

$$T_{ijkt} = 1\left[T^*_{ijkt} > 0\right]$$

$$T^*_{ijkt} = \sum_k \left(HP_j * Class_{ijkt} * year_t \tau^H_{tk} + OP_j * Class_{ijkt} * year_t \tau^O_{tk} + Class_{ijkt} * year_t \tau^N_{tk}\right) + ... \quad (3.1)$$
$$X_{it}\beta + \sum S_j \kappa_j + \epsilon_{ijt}$$

$$T^*_{ijkt} = \sum_k \left(HP_j * Class_{ijkt} \alpha^H_k + OP_{jk} * Class_{ijkt} \alpha^O_k + Class_{ijkt} \alpha^N_k\right) + ... \quad (3.2)$$
$$\sum_k \left(post_t * HP_j * Class_{ijkt} \lambda^H_k + OP_j * Class_{ijkt} \lambda^O_k + Class_{ijkt} \lambda^N_k\right) + X_{it}\beta + \sum_j S_j \delta_j + \varepsilon_{ijt}$$

In equation (3.1), β captures the effect of demographic characteristics (X_{it}) and κ_j captures subject fixed effects. Subjects are then grouped into their official priority status as set out earlier (high-priority (HP), other priority (OP) and non-priority subjects (N)). τ^N_{tk} captures the effect of different degree classifications (indexed by k) in non-priority subjects for each year, whilst τ^H_{tk} and τ^O_{tk} capture the differential effect of different degree classifications in high and other priority subjects, respectively, over and above the effect of degree classifications in non-priority subjects over time.

Estimates before 2012 represent the baseline assumptions for the differential effect of degree classification by subject grouping over time. I make no assumption about their relative size, which could go in a range of directions depending on the underlying willingness of different individuals to become a teacher and demand.

The main objects of interest are how the effects of degree classification change by subject group after the introduction of the high-value and targeted bursaries for cohorts entering training from 2012 onward. I therefore present the relevant coefficients as one would present an event study. This allows us to observe the full trends over the period and also to see, in addition to any major shifts after 2012, whether there were minor changes in response to smaller changes to bursaries in 2011 and 2014. Interest lies in whether the τ_{tk}^{H} and τ_{tk}^{O} increase by more after 2012 for higher degree classifications, and whether the increase is larger for high priority than non priority subjects. Such a hypothesis effectively represents a triple difference capturing the change over time in the effect of higher degree classifications for higher priority subjects.

The main identifying assumption is that there should be no other policies or labour market changes that would differentially affect high-degree classifications in high-priority subjects over the same period. An example of such a policy would be changing admission requirements for teacher training by different subjects over time. I am not aware of any such policy around admission requirements, though there was an increase in employment-based training routes for teaching which could have differentially affected individuals by degree classification and subject. I check for evidence of this as part of robustness checks. An example of a confounding labour market trend would be differential changes in outside wage opportunities by subject and skill level over time. I therefore also show changes in the distribution of wages by subject of study over time for all college graduates. As discussed in the previous section, changes to tuition fees in 2012 are unlikely to affect the expected fee costs of training to be a teacher overall or by prior subject and degree classifications.

I also estimate equation (3.2), which allows me to directly estimate the triple-differences across periods (pre and post 2012), subject priority status and degree classification, and to directly check the statistical significance of any differential effects. To examine whether there are differential effects by subject within the three subjects groupings, I also show trends over time by individual subjects in terms of the propensity of graduates with different degree classifications to train as a teacher.

As likelihood to train as a teacher is likely to be correlated within subject and there are only a small number of subjects (15 in total), inference is conducted using the wild cluster bootstrap approach, with 95% confidence intervals presented in regression results and figures (Cameron et al. (2008); Cameron and Miller (2015)). This increases the size of confidence intervals relative to robust standard errors clustered at the subject level.

3.3.2 Aptitude of trainees

To estimate the effect of changes in bursaries on the skill level of individuals training as teachers, I effectively rely on a difference-in-differences approach. In particular, I examine whether there is evidence of a differential increase in the average skill level of teachers in higher priority subjects after the introduction of high-value bursaries in 2012. The main estimating equation is shown in equation (3.3), where Q_{ijt} represents a measure of the skill level of individual (i), with a college degree in subject (j) and graduating from college at time (t) and $teach_{ijt}$ represents an indicator for whether individuals are training to be a teacher. The γ_t^N terms capture the differences in average skill level over time for teachers in non priority subjects versus other graduates. Interest then lies in whether the γ_t^H and γ_t^O terms increase after 2012. In addition, I also estimate the double-differences directly in equation (3.4) to analyse the statistical significance of any differential changes over time. In both cases, I include subject-by-year fixed effects to allow for changing skills composition of graduates by subject over time. Inference is again conducted using the wild cluster bootstrap approach.

$$Q_{ijt} = \left(HP_j * teach_{ijt} * year_t \gamma_t^H + OP_j * teach_{ijt} * year_t \gamma_t^O + teach_{ijt} * year_t \gamma_t^N\right) + X_{it}\zeta + S_{jt}\phi_{jt} + \upsilon_{ijt} \tag{3.3}$$

$$\begin{gathered} Q_{ijt} = \left(HP_j * teach_{ijt}\theta^H + OP_j * teach_{ijt}\theta^O + teach_{ijt}\theta^N\right) + ... \\ \left(post_t * HP_j * teach_{ijt}\pi^H + post_t * OP_j * teach_{ijt}\pi^O + post_t * teach_{ijt}\pi^N\right) + X_{ijt}\zeta + S_{jt}\xi_{jt} + \nu_{ijt} \end{gathered} \tag{3.4}$$

The identifying assumption here is stronger than for changes in the propensity of individuals to become a teacher. This is because I am no longer able to rely on variation by degree classification as one would not necessarily expect such a policy to affect average skill level within college degree classification over time, particularly as it is targeted at increasing numbers with particular college degree classifications over time. Here, the identifying assumption is that there are no other policies or changes in wage opportunities differentially affecting subjects over time by skill level over the same period. I find little evidence of this in analysis of wage trends by subject over time. However, it is certainly possible that the increased prevalence of employment-based routes affected the composition of teachers training in different routes. This is discussed as part of the robustness checks.

3.4 Data and descriptive trends

This section describes the data sources used in the analysis before showing the observed numbers of teachers by subject over time, the relative wages of teachers as compared with graduates who studied different subjects and the distribution of the educational achievement of trainee teachers by subject over time.

3.4.1 Data sources and sample selection

The main source of data is the Destination of Leavers from Higher Education (DLHE) survey. The DLHE survey is a survey of all graduates leaving higher education across the UK each year. Although not all graduates respond to the survey, response rates are relatively high, with approximately 79% of graduates responding to the latest survey[5]. The survey records graduates' main activity six months after finishing their course. For those still in education, I am able to classify courses by their level of study and subject, which allows me to identify individuals training to be a teacher straight after finishing university. The DLHE data can be linked to Higher Education Statistical Agency (HESA) record data, such as subject of study, institution, college degree classification, key characteristics about individuals (e.g. gender and age) and examination results on entry into higher education. This enables me to document changes in the number of individuals training to be a teacher by prior subject of study, their degree classification and their aptitude, and how this compares with all recent graduates by subject of study.

[5] https://www.hesa.ac.uk/stats-dlhe

I use two main proxies for skills and aptitude. First, I use undergraduate degree classification, where the highest classification is a first class degree, followed by upper second class, lower second class and third class. The lowest classification is a pass mark. These marks are determined by individual universities, with some limited moderation across institutions. Given that bursaries were specifically targeted at different degree classifications, I use this as the main measure of eligibility for bursaries over time.

Degree classification is a relatively coarse measure of educational achievement and there is evidence that similar degree classifications from different institutions are valued differently in the labour market (Britton et al. (2016b)). I therefore also examine a more finely grained measure of educational achievement, which is consistent across individuals. In particular, I use individuals' UCAS Tariff Score as a proxy for skills. The UCAS Tariff score converts individuals' different educational qualifications on entry to higher education into a points score (e.g. an A at A-Level attracts 120 points and a B 100 points, with further examples shown in Appendix Table B1). Offers for higher education are often specified in terms of a certain minimum UCAS tariff score. This UCAS Tariff Score is generally only available for individuals on undergraduate courses, so I cannot use it to analyse the destinations of individuals who graduated from PGCE courses.

My preferred sample represents individuals who graduated from full-time undergraduate degrees in England between 2006-07 and 2014-15 (and thus entered PGCE courses between 2007-08 and 2015-16). Sample sizes by year of graduation are shown in column (1) of Table 3.1, which clearly grow over time, reflecting rising participation in higher education and improvements in the survey. Panel (b) then shows that about 2-3% of such graduates are observed on a postgraduate teacher training course (mostly those taking PGCEs). This group represents the main focus of the the analysis as this is the main group who are eligible for bursaries. For a similar reason, the analysis focuses on graduates in England as bursaries have only changed substantially for teachers training in England[6].

The rest of Table 3.1 shows how the sample sizes and propensity to train as a teacher changes under different assumptions. When narrowing down to individuals aged under 25 (column (2)), the propensity to train as a teacher increases as young people are slightly more likely to train as a teacher, probably because they are more likely to be following a standard educational route. Column (3) adds leavers from part-time courses, which reduces the propensity to train as a teacher, probably because individuals taking part-time courses are more likely to be older students with a specific career path in mind. Column (4) adds postgraduate courses, excluding teacher training, which further reduces the propensity to train as a teacher. Again, this is probably because individuals taking postgraduate courses are more likely to have a specific career path in mind and have already rejected the option to train as a teacher. Expanding to students across the UK in column (5) does not change the propensity to train as a teacher as students across the four countries have a similar propensity to train as a teacher. Whilst the main specification focuses on leavers from full-time undergraduate courses, I do perform robustness checks with these different samples. This confirms that the sample selection assumptions do not affect the main results.

Table 3.2 compares the average characteristics of individuals observed on postgraduate teacher training courses six months after graduation with the main comparison sample (all individuals leaving full-time undergraduate courses in England). Trainee teachers are slightly younger, on average, than all full-time undergraduate leavers (an average age of 23 for trainee teachers compared with 24 for all undergraduate leavers

[6]Strictly speaking, we are not capturing this perfectly as this group represents those who graduated in England, rather than those observed in England 6 months later. We do not observe the latter in the data for those in teacher training. However, there is a very correlation between studying England and activity 6 months later

Table 3.1: Sample sizes and proportion of trainee teachers over time

	(1) FT Undergraduate Leavers (Eng)	(2) FT Undergraduate Leavers (Eng, $<$ 25)	(3) FT & PT Undergraduate Leavers (Eng)	(4) FT & PT Higher Education Leavers (Eng)	(5) FT & PT Higher Education Leavers (UK)
A) Number of graduates over time					
2006-07	153,587	127,401	179,345	224,008	270,695
2007-08	163,309	135,439	195,731	234,455	281,963
2008-09	166,958	139,203	200,078	242,406	290,603
2009-10	176,594	147,181	213,726	262,036	309,884
2010-11	184,585	154,759	220,957	273,925	324,223
2011-12	189,991	159,652	232,286	285,972	338,335
2012-13	200,353	169,544	242,593	299,347	352,185
2013-14	205,993	176,538	241,115	295,117	349,047
2014-15	186,922	159,943	219,716	274,543	326,155
B) Percentage Training as Teachers 6 months after Graduating					
2006-07	2.6%	2.8%	2.4%	2.0%	2.1%
2007-08	2.6%	2.8%	2.4%	2.1%	2.1%
2008-09	2.7%	2.8%	2.4%	2.1%	2.1%
2009-10	2.2%	2.3%	2.0%	1.7%	1.7%
2010-11	1.9%	2.0%	1.7%	1.5%	1.5%
2011-12	2.8%	2.8%	2.4%	2.1%	2.1%
2012-13	2.7%	2.8%	2.4%	2.1%	2.1%
2013-14	2.7%	2.8%	2.5%	2.1%	2.2%
2014-15	2.7%	2.8%	2.5%	2.1%	2.2%

Notes: Author's calculations using HESA Destinations of Leavers from Higher Education. Teachers in training are defined as those undertaking a postgraduate teacher training course at the survey date.

) and more likely to be female (76% compared with 56%). Trainee teachers also have slightly higher degree classifications, on average, with a larger share possessing an upper second class degree or above. However, the distribution of aptitude (as captured by UCAS tariff scores) is very similar across both groups.

Figure 3.2 shows the total number of graduates from higher education in England observed on teacher training courses in the data 6-months after graduating, the total number in the main sample (full-time undergraduates) and the actual number of individuals on postgraduate teacher training courses each year from 2010-11 onward. As can be seen, the data includes about 5000-6000 trainees each year, which falls by about 1,000 when narrowing down to full-time undergraduates leavers. This represents about one third of all individuals on formal postgraduate teacher training courses each year. This is relatively low, the reason being that not all individuals train as a teacher straight after finishing their undergraduate degree, with some seeking other employment first.

As a robustness check, I therefore also use another version of the data which surveys graduates 3 years later. Response rates are much lower, leading to lower sample sizes and this survey is only undertaken every 2 years. I pool data across three cohorts of this data (2006/07, 2008/09 and 2010/11 leavers) to examine the characteristics of individuals who train to be a teacher later after finishing their studies. This shows that there is indeed further growth in trainee numbers, but that the average aptitude of teacher trainees 6 months after graduating from an undergraduate degree is very similar to those observed as actual teachers 3

Table 3.2: Characteristics of trainee teachers and all graduates

	Main Sample (FT undergrad. leavers in England)	**Trainee teachers (6 months after graduation)**
Individual Characteristics		
Age	24.01 (5.91)	23.19 (5.13)
Female	0.57 (0.50)	0.76 (0.43)
Degree Classification		
First class	0.16 (0.37)	0.18 (0.38)
Upper Second	0.45 (0.50)	0.59 (0.49)
Lower Second	0.21 (0.41)	0.21 (0.40)
Third or Below	0.07 (0.26)	0.02 (0.13)
Unknown	0.10 (0.30)	0.01 (0.09)
UCAS Tariff Scores		
Average Scores	335.82 (129.53)	340.65 (118.53)
Missing Tariff Score	0.26 (0.44)	0.17 (0.38)
UCAS Tariff - P10	180	200
UCAS Tariff - P25	250	260
UCAS Tariff - P50	330	340
UCAS Tariff - P75	420	420
UCAS Tariff - P90	500	490

Notes: Author's calculations using HESA Destinations of Leavers from Higher Education. Teachers in training are defined as those undertaking a postgraduate teacher training course at the survey date. Standard deviations in parentheses

years later. As a further robustness check, I examine the numbers and educational achievement of teachers training through employment-based routes, whose number have grow over time.

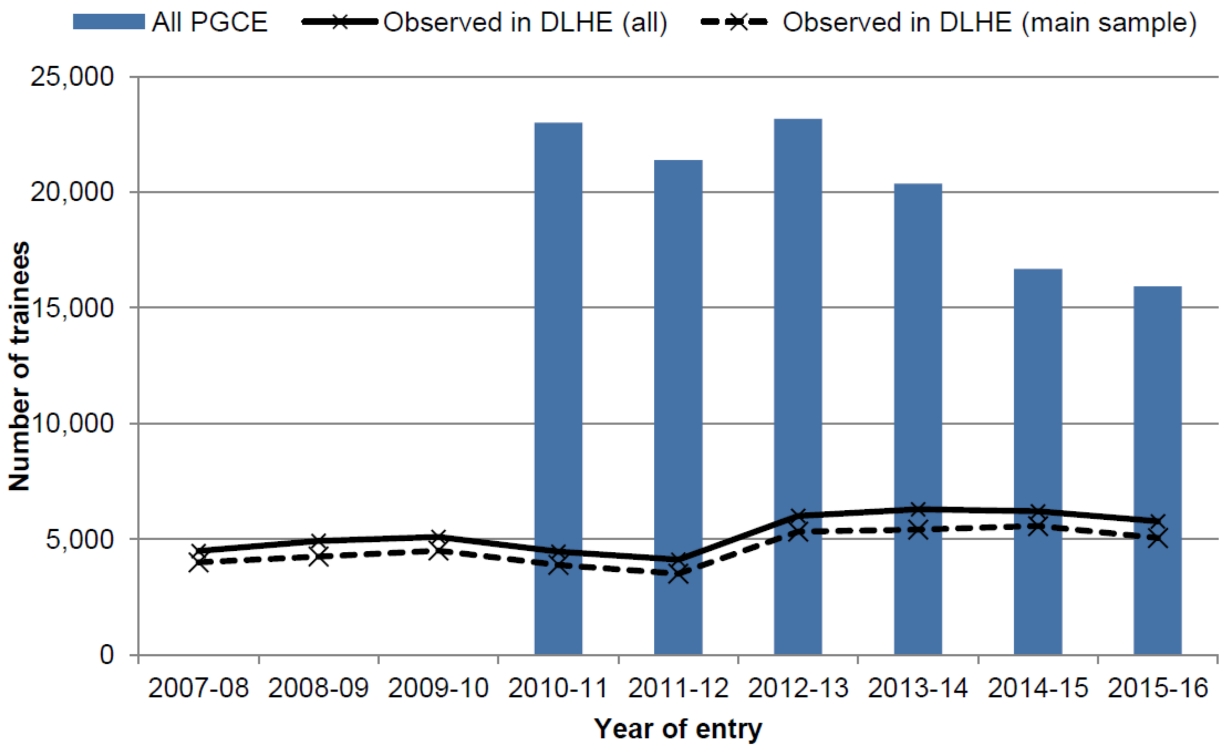
3.4.2 Numbers of teachers by subject over time

To examine differences by subject of study, I classify individuals' undergraduate degrees into 15 different subjects of study based on the two-digit JACS code (detailed classification in Appendix Table B2). Individuals can be allocated to more than one subject if they are studying a joint degree.

Figure 3.3 shows the number of individuals observed on postgraduate teacher training courses by subject over time, with subjects grouped according to whether they were high-priority, other priority or non-priority in 2014-15. The dashed line shows the number of such individuals where prior educational attainment is non-missing, which in most cases makes very little difference to sample sizes. The grey line then shows the number trainees as a share of total graduates in each share over time. Results for computing, medical science and design and technology are not shown due to small sample sizes.

Looking across subjects, biology was the subject with the highest number of individuals on teacher training courses (with nearly 1,000 per year) closely followed by those with degrees in the arts, social sciences, education, history and English, all with around 500 per year. These are all other priority or non-priority subjects. Numbers are much lower, by contrast, in high priority subjects, with only around 100 per year with degrees in physics and chemistry and 200 per year with degrees in maths and languages. This illustrates some of the problem in recruiting sufficient quantities of teachers in such subjects.

Figure 3.2: Numbers of PGCE teacher trainees over time



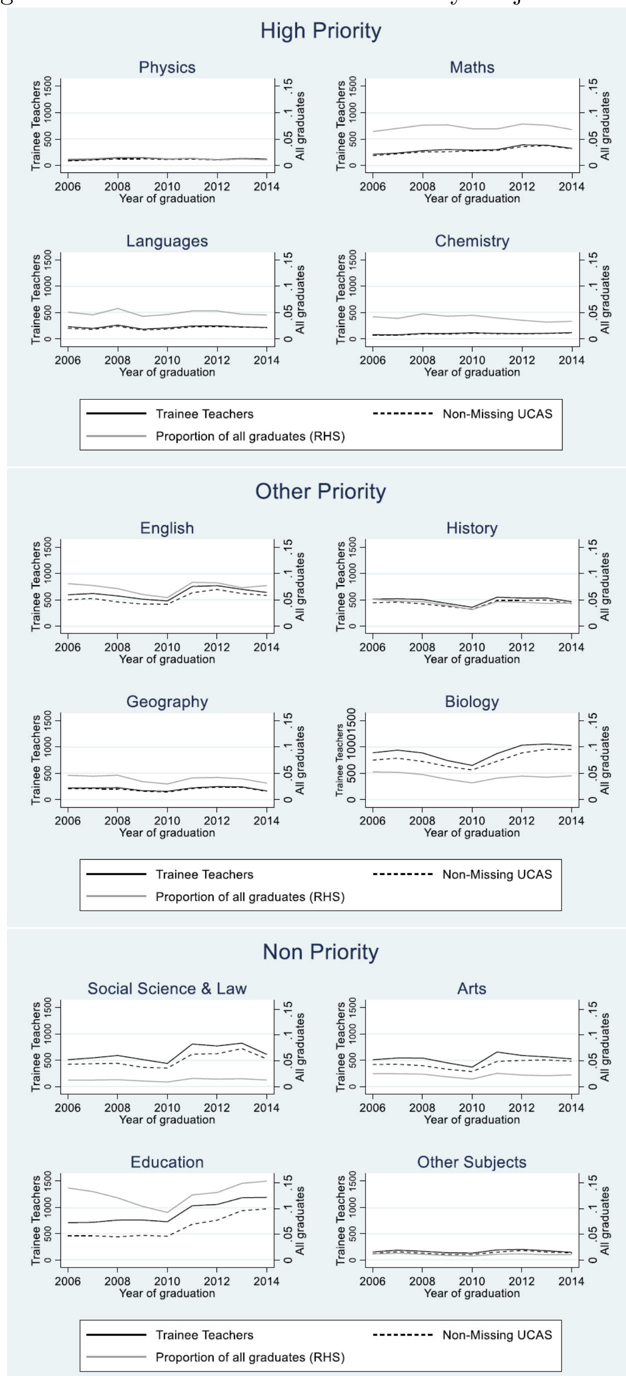
Notes and Source: Initial Teacher Training Census (various years); author's calculations using HESA Destinations of Leavers from Higher Education. Teachers in training are defined as those undertaking a postgraduate teacher training course at the survey date.

If we look at the propensity of graduates from different subjects to train to be a teacher, we see that this is either just above or just below 5% of graduates in most subjects. There are, however, some notable exceptions, with only about 1% of graduates in the high-priority subject of physics. Whilst in the arts and social sciences (non-priority subjects), we see relatively low shares of graduates going into teaching, but this still leads to high numbers of teachers with degrees in these subjects as they are relatively popular degrees. A larger share (10-15%) of individuals with degrees in education related subjects tend to train as teachers, which is unsurprising.

In terms of trends over time, there is little evidence of systematic trends in the numbers of trainee teachers or propensity of graduates in each subject to train as a teacher. The only exception to this is that there is a dip and then recovery in numbers in some subjects for those graduating from courses between 2009-10 and 2011-12, most notably in English, history and biology. It is not clear exactly what is driving this pattern, but it is notable that such patterns correlate with the ending and re-introduction of bursaries in such subjects. This highlights the importance of examining the specific trends in bursaries for each subject, with the policy change not simply being higher levels for higher priority subjects after 2012.

Strictly speaking, eligibility for bursaries depends on which subjects individuals train to teach, rather than their subject of degree. In the HESA data, one only observes subject of prior study, rather than training subject. Such differences are still of considerable interest given that the target of the policy is aimed at

Figure 3.3: Number of teacher trainees by subject over time



Notes and sources: Author's calculations using HESA Destinations of Leavers from Higher Education. Trainee Teachers defined as those on a postgraduate teacher training course at the survey date. Subject defined by JACs code of prior undergraduate degree.

increasing the numbers and quality of teachers with expertise in specific subjects. Furthermore, there is generally a close correlation between subject of prior study and training subject as individuals prior subject of study has a strong bearing on one's ability to train to teach that subject. As can be seen from Table B2, the specific subjects included for each subject definition are relatively broad too (e.g. physics includes engineering and materials sciences).

A potential concern with my method is that the bursaries policy may induce more individuals to train in high-priority subjects outside their area of prior study. For example, more biology graduates may train as chemistry teachers as the latter is a high-priority subject and the former is not. Figure B1 therefore shows trends in the number of trainees by the actual subject they are training to teach. This shows that all subjects were on a downward trend between 2009 and 2014, and there is little difference across subject. There is also no evidence of any differential change in trends after 2012 for high-priority subjects like maths and languages as compared with other subjects like English, history and geography. Science also seems to follow the same trend for other subjects, though this will include high-priority subjects like chemistry and physics and other subjects like biology. Unfortunately, one can only look at science subjects together as science cannot be broken down by individual subject before 2014. Its possible that there is a compositional shift towards chemistry and physics within the science group, but this seems unlikely, particularly as the overall trends match those seen in DLHE.

3.4.3 Relative teacher wages by subject of study

There are probably a wide range of factors contributing to teacher shortages by subject, but one of the most important considerations is likely to be the structure of teacher wages and variation in the level of outside wage opportunities for individuals with degrees in different subjects. For most of the recent past, teachers in England have been paid according to centrally determined salary schedules, where teacher pay rises with experience and is also higher in the London area to compensate for a higher cost of living (Greaves and Sibieta (2019)). It also varies very little, if at all, by subject (Allen and Sims (2018)) In contrast, there is significant variation in graduate wages by subject of study, which is increasingly well recognised in the empirical literature (Britton et al. (2016b); Kirkeboen et al. (2016)).

To illustrate this, I make use of the Quarterly Labour Force Survey (LFS) to describe changes in graduate wages by major subject over time. The LFS is the main source of data for official labour market statistics, with about 150,000 adults surveyed each quarter. I focus on the annual gross earnings of graduates aged 22-30 to give an indication of the likely salary graduates with degree in different subjects can expect early in their career. I pool individuals over each calendar year to ensure sufficient sample sizes for each subject grouping.

Figure 3.4(a) shows the average (median) early career salaries of graduates with degrees in specific selected subjects relative to that of teachers over time, and panel (b) shows real-terms changes over time relative to 2011 (the year just before the introduction of high-value targeted bursaries). Perhaps unsurprisingly, high priority subjects (coloured blue) are precisely those with the highest outside wage opportunities. However, it is also clear that the average salaries of teachers are relatively high. For example, the average salary for teachers was about £25,000 is 2015, which is only about 8% lower than the average salary of £27,000 in physics and about 10% lower than the average salary of £28,000 for maths and computing graduates. In contrast, the average wage of English and social science graduates is about 8% lower at £23,000, about 12%

lower in biology, about 16% lower in humanities and nearly 30% lower in arts subjects.

If we looks at trends over time, we see that relative teacher wages have changed over time, but there is relatively little heterogeneity by subject. Up to 2005, there was significant real-terms growth in graduates wages, with slightly faster growth in teacher wages narrowing relative wages across most subjects to a similar degree. Between 2005 and 2009 (covering most of the Great Recession), graduate and teacher wages were stagnant, keeping relative wages relatively constant. Since 2010, there have been falls in graduate wages across all subjects, as well as falls in teacher wages. Concentrating on changes since 2011 (just before the changes to bursaries), graduate wages fell by around 5-10% in real-terms between 2011 and 2015 across most subjects. Faster falls amongst teachers (around 13%) have then led to growth in the wages of graduates in most subjects relative to teachers. However, because the falls in graduate wages have been relatively similar across subjects, the growth in relative wages has also been similar.

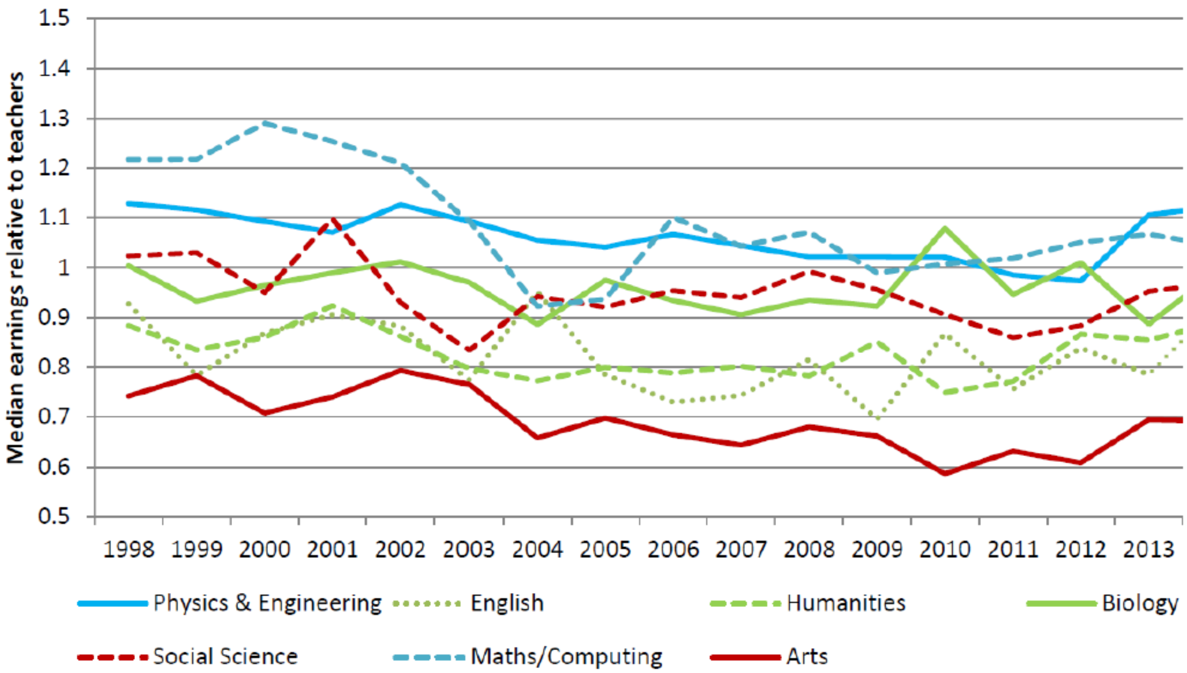
Such similar trends in average wages are reassuring from an identification perspective. However, identification also requires similar trends in relative wages for high and low skilled individuals across subjects. For example, if high skilled physics graduates were experiencing faster growth in wages relative to high-skilled English graduates, this could bias estimates of the impact of bursaries. Figure 3.5 therefore shows the level of the 75th and 25th percentile of graduates wages across subject relative to the subject-specific median over time, as well as the equivalent for teachers.

What is immediately clear is just how little variation there is in teacher salaries relative to that of graduates in different subjects. The relative gap to the median is about 15% at both the 25th and 75th percentile for teachers, but generally over 30% across most subjects. As a result, the interquartile range is about 35% for teachers, compared with about 80% across most subjects of study. Although there is clearly significant noise in the data, these differences appear relatively constant over time. The larger range of wages outside of teaching will reflect a combination of greater variation within and across occupations. However, both sources of variation are likely to be important for individuals considering a career in teaching. For example. a highly-skilled physics graduate has a range of attractive outside wage opportunities across different occupations, making a career in teaching relatively unattractive.

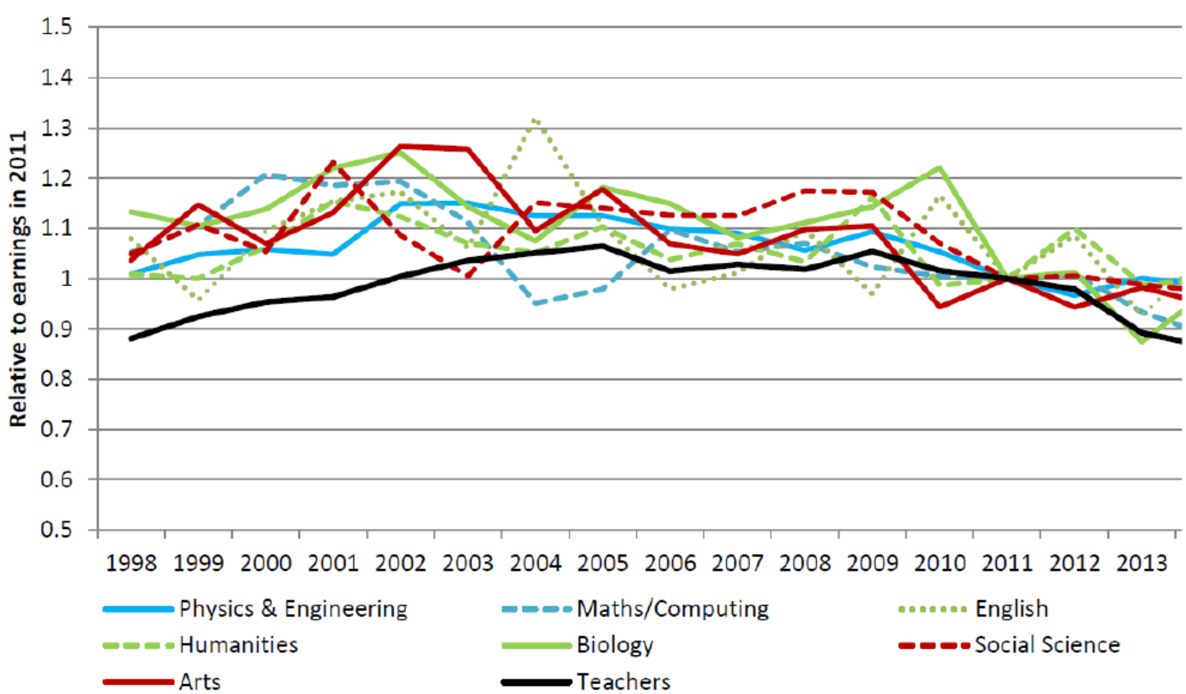
The most obvious policy response to such concerns would be for teacher salaries to vary by subject, and potentially within subject to attract more highly qualified individuals. Schools in England have been given increasing levels of autonomy over teacher salaries over time that would allow them to do just that. All schools can use targeted recruitment and retention payments, the large number of schools converting to Academy status can deviate from national pay structures and since 2013 all schools have to design their own pay structures. Despite this, the average level of teacher pay by subject varies very little, both early in teachers careers and as they gain more experience (Allen and Sims (2018)) One explanation for an aversion to varying teacher salary by subject could be an aversion to inequality within schools. Whatever the reason, the lack of variation in teacher pay and persistent shortages by subject creates an interest as to whether other means could be used to tackle teacher shortages by subject.

Figure 3.4: Average early career gross earnings of graduates by subject of study relative to teachers over time

a) Median Earnings Relative to Teachers

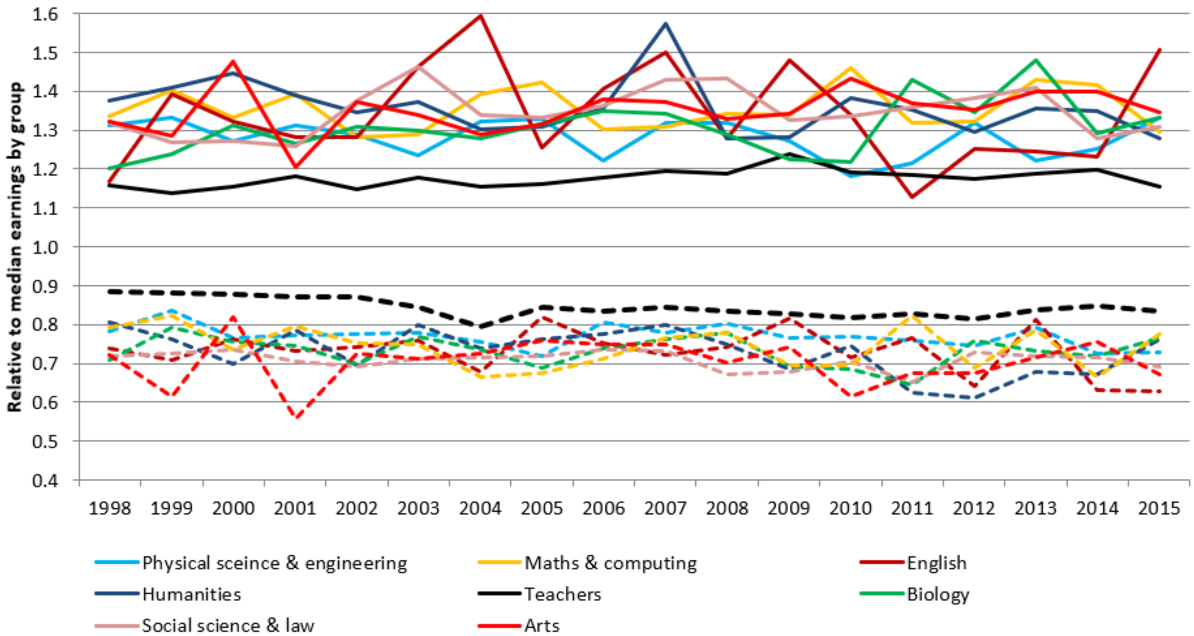


b) Gross salary levels over time (relative to 2011)



Notes and sources: Authors calculations using Quarterly Labour Force Survey (various years).

Figure 3.5: Early career gross earnings of graduates, P75 and P50 relative to median by group



Notes and sources: Authors calculations using Quarterly Labour Force Survey (various years).

3.5 Empirical Results

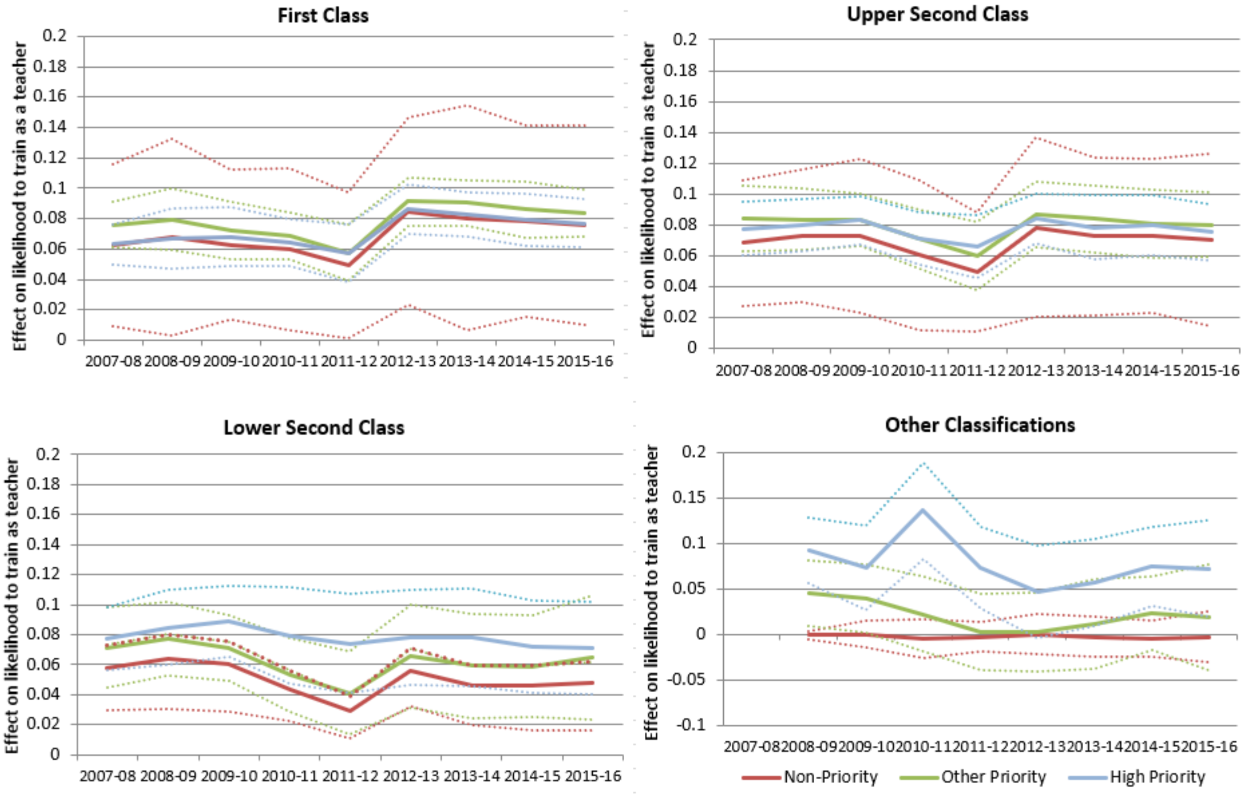
3.5.1 Number of trainees

Given the expansion of bursaries in 2012 was targeted at graduates with high degree classifications in high-priority subjects, the natural first question to ask is whether the policy increased the propensity of such individuals to train as a teacher. I present three pieces of evidence to suggest this is not the case: year-by-year regression estimates of the effect of subject priority status by college degree classification; trends over time in the raw propensity to train as a teacher by individual subject of study and degree classification; and, regression estimates of the differential effect of subject priority status by college degree classification before and after the change in 2012.

Figure 3.6 begins by illustrating the main estimates of equation (3.1), with the solid vertical line illustrating the changes to bursaries for entrants in 2012. In all cases, the effects are shown relative to the omitted category of a low degree classification at the start of the period (entrants in 2007-08). This means that for priority subjects we add together the baseline effect for non-priority subjects to the extra effect for degrees in higher priority subjects.

Looking across degree classifications, college graduates are generally more likely to train as a teacher if they have a degree in a priority subject, reflecting greater demand for such subjects from schools and training providers. However, there is a different gradient by degree classification. For non priority subjects, the propensity to train as a teacher generally increases with college degree classification. Individuals with a lower second class degree were around 5 percentage points more likely to train as a teacher in 2014 than those with a low degree classification, and those with an upper second class or first class degree around 7 percentage

Figure 3.6: Effect of degree classification on likelihood to train as a teacher by subject priority status and by year, with 95% confidence intervals



Notes and sources: Author's calculations using HESA Destinations of Leavers from Higher Education. Estimated effects taken from probit regression on year-by-year effects of degree classification interacted with subject priority status. Other classification effects shown relative to 2007-08, with all other estimated effects showing the total effect incorporating the baseline effect of other classifications each year. Controls for age, subject, gender and region. 95% confidence intervals are estimated using a wild cluster bootstrap with Webb weights

points more likely. For high priority subjects, there is no such pattern, with their relative likelihood to train as a teacher similar by degree classification in individual years (around 7 percentage points more likely to train as a teacher as compared with the omitted category). This is almost certainly a reflection of the better wage opportunities for graduates in high-priority subjects outside of teaching.

Turning to trends over time for low degree classifications, there is no net change over time for non-priority subjects. Those with low degrees in other and high priority subjects remain more likely to train as a teacher, though this difference declines over time, with almost all of this decline occurring prior to the change in bursaries in 2012. After 2012, the changes are relatively small, with those with low degree classifications in other and high priority subjects becoming about 2 percentage points more likely to train as a teacher relative to those with low degrees in non-priority subjects. For lower second-class degrees, the next lowest classification, there is a fall in the propensity to train as a teacher for non and other priority subjects between 2007 and 2011, with a recovery after 2011 of about 2 percentage points. For high priority subjects, there is a small and gradual decline over time and no real recovery after 2011.

For those with upper second class degrees, there are very similar patterns over time across subject groupings. There is a slight decline over time for all subject groupings up to 2011 of about 2 percentage points, followed by an uptick in 2012. The size of this uptick is very similar across all subject groupings (about 2 percentage points), suggesting it is unlikely to be driven by the change in bursaries for entrants in 2012. For those with the highest degree classification (first class), there is a similar pattern with a gradual decline up to 2011, followed by an uptick in 2012. However, this uptick is again of a similar value across the subject groupings, around 3 percentage points.

In summary, there is no evidence of a differential increase in the propensity to train as teachers from 2012 for those with high degree classifications in priority subjects. There is a similar increase across all subject groupings for high degree classifications. For low degree classifications, the trends are similar across subject groupings or slightly more positive for high-priority subjects (as in the case of those with lower second class degrees). Whilst there is no evidence of differential trends, the relatively wide confidence intervals does limit the power of this analysis.

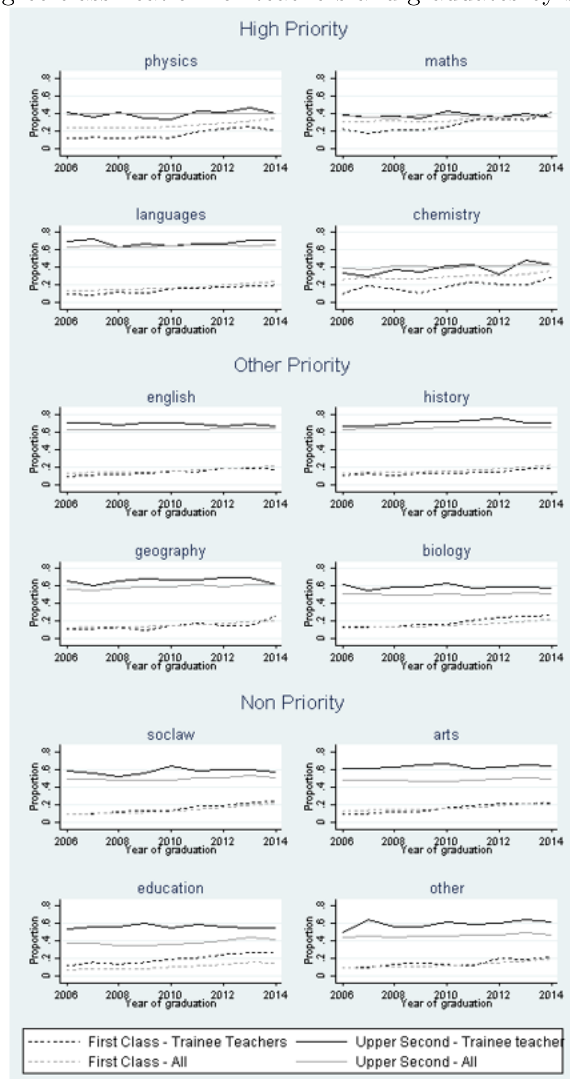
Such trends across subject groupings could mask important differences across subjects within each grouping. Figure 3.7 therefore shows the proportion of teacher trainees with first class and upper second class degrees by subject over time, together with the equivalent figures for all graduates. This shows that the proportion of teacher trainees with an upper second closely matches that of all graduates in high-priority subjects over time. However, a lower share of teacher trainees had a first class degree as compared with all graduates, particularly in physics and chemistry. Such under-representation is exactly what one would predict given the differences in the distribution of salaries by subject observed earlier. This difference shrinks slightly over time for maths and physics. However, most of this occurs well before the change in bursaries in 2012. Across other priority subjects, the proportion of teacher trainees with an upper second is higher than that amongst all graduates in such subjects, whilst the proportion with a first class degree is similar. For non-priority subjects, there are similar trends to other priority subjects, with evidence of positive selection of those with upper second degrees and similar proportions of those with first class degrees. In both cases, there is no evidence of differential trends over time for teacher trainees in other or non priority subjects.

The third piece of evidence comes from estimates of equation (3.2), which are shown in Table 3.3. The first set of columns shows the results for all full-time graduates from undergraduate courses in England, and then split for men and women separately for models (2) and (3). In each case, the first column shows estimates of the α terms in equation (3.2), the differential impact of degree classification by subject grouping before the change in the bursaries policy. The second column then shows estimates of the λ terms (the change in the impact of degree classification by subject grouping after bursaries changed in 2012). The probability of training as a teacher is modeled as a Probit model.

As with the trends over time, estimates of the α terms show a clear gradient in the impact of degree classification by subject grouping prior to 2012. Prior to the introduction of the bursaries in 2012, the effect of degree classification on likelihood to train as a teacher is markedly different by the priority status of subjects. Amongst low priority subjects, it is those with the highest degree classifications that are most likely to train. Whilst amongst the highest priority subjects, this is reversed, with lower degree classifications more likely to train as a teacher.

The next column then shows how these differential effects change after 2012, with most of the changes very small. Amongst non-priority subjects, there was a slight increase amongst the highest degree classifications (0.5 percentage points) and no change amongst those with low classifications. This makes for

Figure 3.7: Degree classification for teachers and graduates by subject over time



Notes and sources: Author's calculations using HESA Destinations of Leavers from Higher Education. Trainee Teachers defined as those on a postgraduate teacher training course at the survey date. Subject defined by JACs code of prior undergraduate degree

Table 3.3: Effect of degree class by subject priority on likelihood to train as a teacher

	(1) All		(2) Men		(3) Women	
Degree Class & Subject Group	**2007/08 to 2011/12**	**Change in effect after 2012/13**	**2007/08 to 2011/12**	**Change in effect after 2012/13**	**2007/08 to 2011/12**	**Change in effect after 2012/13**
Non-priority (Baseline)						
First Class	0.055	0.005*	0.023*	0.005**	0.076	0.006*
	[-0.007;0.117]	[-0.011;0.021]	[0.005;0.040]	[0.000;0.010]	[0.002;0.150]	[-0.011;0.023]
Upper Second	0.037**	0.003	0.020**	0.003*	0.048**	0.003
	[-0.013;0.086]	[-0.007;0.012]	[0.004;0.035]	[-0.001;0.007]	[-0.015;0.110]	[-0.009;0.015]
Lower Second	0.045*	0	0.026**	0	0.057*	-0.001
	[0.010;0.080]	[-0.006;0.005]	[0.016;0.036]	[-0.003;0.004]	[0.017;0.097]	[-0.009;0.007]
Other Classification	[omitted]	-0.001	[omitted]	-0.001	[omitted]	-0.001
		[-0.012;0.010]		[-0.006;0.003]		[-0.013;0.012]
Other Priority (Relative to Baseline)						
First Class	0.01	-0.001	0.012*	-0.002	0.011	-0.002
	[-0.008;0.028]	[-0.011;0.008]	[0.001;0.024]	[-0.008;0.005]	[-0.011;0.034]	[-0.015;0.012]
Upper Second	0.011*	-0.001	0.008	0.000**	0.015**	-0.003
	[-0.010;0.032]	[-0.008;0.006]	[-0.004;0.019]	[-0.004;0.005]	[-0.012;0.042]	[-0.012;0.007]
Lower Second	0.012	0	0.006	-0.001	0.018	0.001
	[-0.013;0.037]	[-0.013;0.013]	[-0.006;0.017]	[-0.008;0.007]	[-0.016;0.052]	[-0.017;0.019]
Other Classification	0.028	-0.005	0.012	-0.004	0.035	-0.005
	[-0.006;0.062]	[-0.017;0.007]	[-0.002;0.026]	[-0.009;0.001]	[-0.009;0.078]	[-0.025;0.015]
High Priority (Relative to Baseline)						
First Class	0.004	-0.002	0.002	-0.001	0.011*	-0.003
	[-0.016;0.024]	[-0.010;0.006]	[-0.016;0.020]	[-0.008;0.006]	[-0.020;0.043]	[-0.014;0.008]
Upper Second	0.010**	-0.003**	0.006	-0.002***	0.022**	-0.004**
	[-0.011;0.032]	[-0.007;0.001]	[-0.009;0.021]	[-0.004;0.000]	[-0.009;0.054]	[-0.010;0.002]
Lower Second	0.028**	-0.001	0.015*	-0.002**	0.048**	-0.001
	[-0.003;0.059]	[-0.007;0.005]	[-0.006;0.036]	[-0.005;0.002]	[0.012;0.083]	[-0.014;0.012]
Other Classification	0.095*	-0.005	0.043	-0.003	0.151**	-0.008*
	[0.054;0.136]	[-0.026;0.015]	[0.021;0.065]	[-0.017;0.011]	[0.102;0.201]	[-0.037;0.021]
Individual Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,628,292		700,636		927,656	
Pseudo R-Squared	0.136		0.101		0.131	

Notes: Author's calculations using HESA Destinations of Leavers from Higher Education. Teachers in training are defined as those undertaking a postgraduate teacher training course at the survey date. Sample only includes full-time undergraduate leavers from higher education institutions in England. Individual covariates include age, gender, region and subject fixed effects. 95% confidence intervals are shown in brackets and estimated using a wild cluster bootstrap with Webb weights.

a net increase of 0.5 percentage point in the gap between the highest and lowest degree classifications. For other priority subjects, there are falls in the propensity to train as a teacher amongst all degree classifications with smaller falls amongst the highest degree classifications (0.1 for first class and upper second class degrees compared with 0.5 percentage points for low degree classifications). There is an almost identical pattern across degree classifications for high-priority subjects. In both cases, the difference between the change for high and low degree classifications is about 0.4 percentage points.

These results suggest that other factors, such as improving alternative labour market opportunities, are pushing down the propensity of individuals with degrees in priority subjects to train as a teacher. The fact that these falls are slightly smaller for high degree classifications on its own might suggest the bursaries were slightly protective. However, these differences are small and the changes are almost identical to that seen amongst non-priority subjects, where eligibility for bursaries either stayed the same or declined. Whilst the confidence intervals for the baseline effects of degree classifications by subject grouping are relatively wide, the changes after 2012 are estimated with more precision. All this confirms that there is very little evidence that the introduction of high-value and targeted bursaries had any impact on the propensity of graduates targeted the most to train as a teacher.

The next four columns repeat these estimates for men and women separately. Here, there is a similar baseline pattern in the impact of degree classification by subject grouping across men and women, with a greater propensity amongst those with high degree classifications in non-priority subjects which then reverses in high-priority subjects. The magnitude of the effects are clearly larger for women, reflecting the fact that they are more likely to train as a teacher. Even so, the relative effect of degree classification by subject grouping is larger for women. This can be seen most clearly for high-priority subjects. Women with low degree classifications in high-priority subjects are 15 percentage points more likely to train as a teacher relative to those with low degrees in non-priority subjects, which compares with an increase of 1 percentage point for women with first class degrees in high-priority subjects. This compares with 4 percentage points and zero for men.

If we then look at changes over time, there is no evidence of differential changes for men and women. There are falls in the propensity to train as a teacher amongst men and women with degrees in high and other priority subjects, with slightly smaller falls amongst those with the highest degree classifications. And then for those with degrees in non-priority subjects, there is then a faster rise amongst those with higher degree classifications.

The picture by men and women is therefore a bit different in terms of baseline differences by degree classifications, with bigger differences for women than men. However, there is no evidence of differential changes by degree classification amongst high priority subjects after the introduction of high-value, targeted bursaries in 2012.

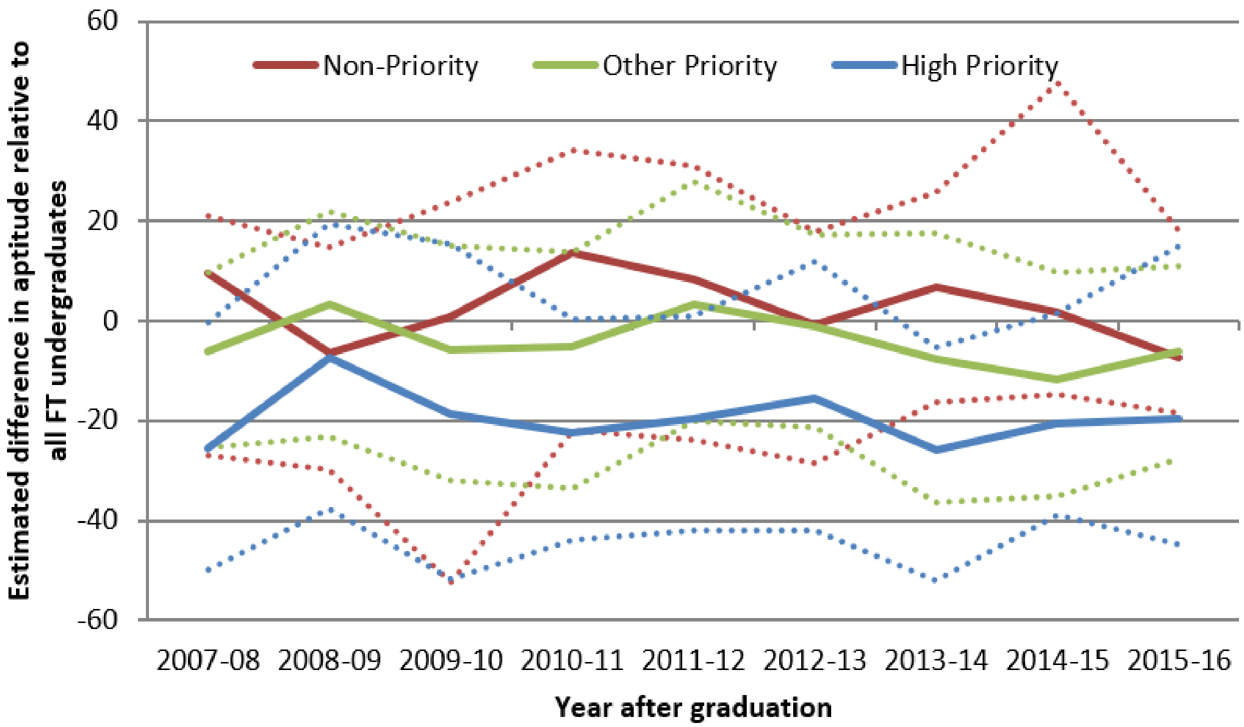
In summary, there is evidence of an increased propensity to train as a teacher for high degree classifications relative to low degree classifications after the change in bursaries in 2012. However, this is common across subject groups, suggesting little evidence that the change in bursaries had any effect on the willingness of targeted individuals to train as a teacher.

3.5.2 Aptitude of trainees

Given that a primary aim of the change in bursaries was to attract more highly skilled individuals into teaching, I now present results on the likely impact of the change in bursaries on skill levels. Here, I proxy skill levels with individuals' UCAS Tariff score, which aggregates individuals' age 18 exam results into a single metric, which is then often used to determine entry into specific higher education courses. As with results for the propensity to train as a teacher, I present three set of results: regression estimates of the differences by subject groupings over time; raw differences in skill level by individual subject; and, regression estimates for the differences in skill level by subject grouping before and after 2012.

Figure 3.8 shows estimates of the γ terms in equation (3.4), i.e. the differential skill level of teacher trainees by subject group over time, controlling for individual characteristics and subject by year fixed effects for graduates as a whole. Before the change in bursaries in 2012, it is clear that the average teacher trainee in non-priority and other priority subjects had similar levels of skills to overall graduates in that subject, with the estimated difference close to zero from 2006 through to 2010. In contract, teacher trainees in high priority subjects had lower skill levels than all graduates in these subjects, about 10-15 points lower (equivalent to about one tenth of a standard deviation). After 2010, the skill level of teacher trainees then declined over time across subject groupings, with the the gap reaching about -10 points for non priority and other priority subjects, and about -25 points for high priority subjects. There is little evidence of any differential change by subject grouping after the changes in bursaries in 2012. Whilst the trends appear similar across subject groupings, it is important to acknowledge that the confidence intervals are relatively wide.

Figure 3.8: Effect of degree classification on likelihood to train as a teacher by subject priority status and by year, with 95% confidence intervals

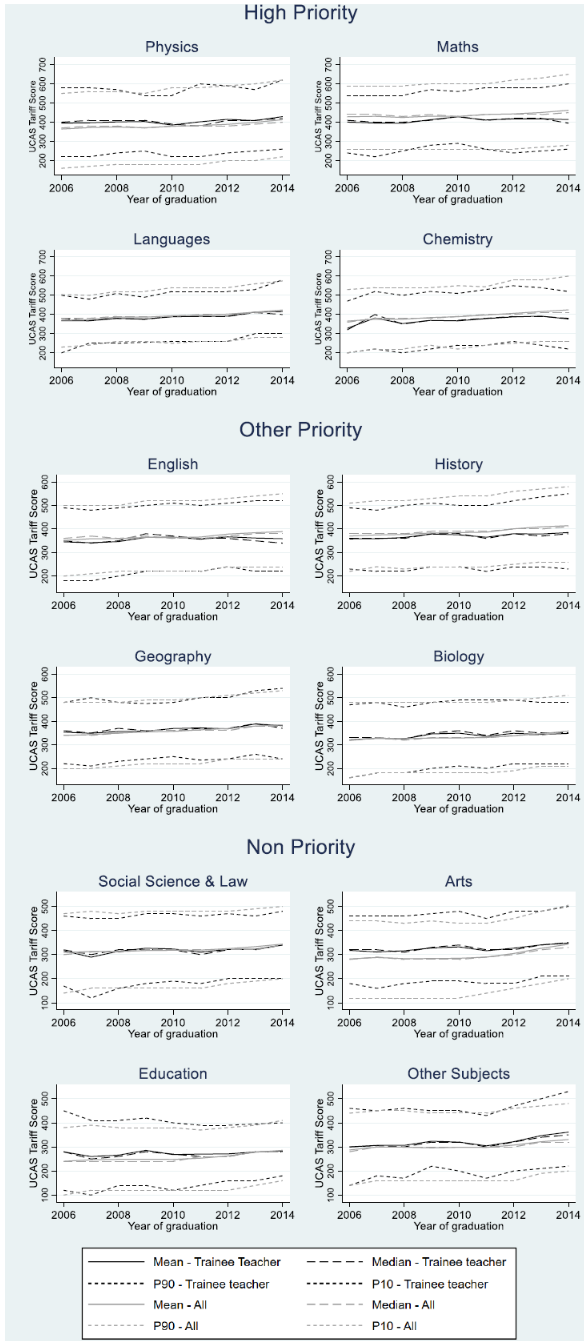


Notes and sources: Author's calculations using HESA Destinations of Leavers from Higher Education. Trainee Teachers defined as those on a postgraduate teacher training course at the survey date. Subject defined by JACs code of prior undergraduate degree. Estimated effects taken from OLS regression on year-by-year effects of subject priority status. All estimated effects are shown relative to other FT undergraduates who graduated in the same year. Regressions also include controls for age, subject by year effects, gender and region. Dotted lines indicate 95% confidence intervals estimated using a wild cluster bootstrap with Webb weights.

Figure 3.9 shows the median, 10th and 90th percentile of the UCAS tariff score over time for trainee teachers by prior subjects of study, as well as the equivalent for all graduates in each subject. There are differences in skill level by subject, but this shows that aptitude is actually higher in shortage subjects like physics, reflecting higher overall graduate aptitude in such subjects. It is also notable that the 90th percentile for teachers is very close to the 90th percentile for all graduates by subject, suggesting no differential absence of highly skilled teachers by subject. The distribution of UCAS tariff scores trainee teachers closely matches that of all graduates in each subject over time, with no evidence of differential trends for trainee teachers in high-priority subjects after bursaries were made more generous in 2012.

Finally, Table 3.4 shows estimates of the differential change in skill level by subject grouping, comparing the pre-2012 period with the post-2012 period. Results are shown for all full-time graduates, as well as men and women separately. Before 2012, there are clear differences by subject grouping, with those training in non priority subjects of slightly higher aptitude (5 UCAS tariff points) than average. Those training to be a teacher in a higher priority subject are generally of lower aptitude, by about 18 UCAS Tariff points for high priority subjects (about one eighth of a standard deviation). These differences are slightly larger for men than they are for women.

Figure 3.9: UCAS Tariff Scores for teachers and graduates by subject over time



Notes and sources: Author's calculations using HESA Destinations of Leavers from Higher Education. Trainee Teachers defined as those on a postgraduate teacher training course at the survey date. Subject defined by JACs code of prior undergraduate degree

Table 3.4: Estimated effect of subject priority status on UCAS tariff scores of teachers in training

	(1) All		**(2) Men**		**(3) Women**	
	2007/08 to 2011/12	**Change in effect after 2012/13**	**2007/08 to 2011/12**	**Change in effect after 2012/13**	**2007/08 to 2011/12**	**Change in effect after 2012/13**
Subject group						
Non Priority	4.8	-4.7	7.9	-5.8	3.3	-4.1
	[-30.1;23.9]	[-14.2;25.1]	[-59.0;45.2]	[-27.3;38.2]	[-24.0;16.7]	[-11.6;23.9]
Non Priority	-2.2	-4.4	-10.6	-1.7	-1.1	-5.1
	[-26.9;13.8]	[-16.7; 8.1]	[-50.1;16.2]	[-29.7;22.6]	[-20.1;12.0]	[-14.9; 5.2]
High Priority	-18.3	-2.1	-22.9	0.9	-15.0**	-4.2
	[-42.0; 6.8]	[-15.2; 9.3]	[-64.3;15.6]	[-17.2;26.5]	[-30.3;-1.7]	[-17.4; 6.8]
Individual Covariates	Yes		Yes		Yes	
Observations	1,202,397		541,539		660,858	
Pseudo R-squared	0.159		0.181		0.155	

Notes: Author's calculations using HESA Destinations of Leavers from Higher Education. Teachers in training are defined as those undertaking a postgraduate teacher training course at the survey date. Sample only includes full-time undergraduate leavers from higher education institutions in England. Individual covariates include age, gender, region and subject-by-year fixed effects. 95% confidence intervals are shown in brackets and estimated using a wild cluster bootstrap with Webb weights.

After 2012, there is no evidence of significant differential change by subject grouping. Across all groups there is a very small reduction in UCAS tariff scores, with these very slightly larger for non and other priority subjects. Whilst the confidence intervals are relatively wide, the actual reductions and differences across subjects are extremely small in absolute value. These results therefore suggest there is little evidence that changes to bursaries have led to an improvement in the aptitude of teacher trainees in targeted subjects.

In summary, there is no evidence of differential change in the skill levels of teacher trainees in high-priority subjects following the introduction of high-value bursaries in 2012. This suggests these targeted cash payments are a relatively ineffective means for increasing the skill level of teacher trainees in these shortage subjects.

3.5.3 Robustness checks

Estimates of the impact of the change in bursaries on the propensity and aptitude of individuals training as teachers are based on whether full-time undergraduates are observed on a formal postgraduate teacher training course 6 months after graduating. To confirm the robustness of the findings, I examine three potential sources of bias: the sample; the focus on destinations 6 months after graduation; and, the changing importance of other routes into teaching.

3.5.3.1 Choice of sample

First, I examine the potential impact of the choice of sample. Table B3 repeats the main results for the changing impact of degree classification by subject grouping on the propensity to train as a teacher across a range of samples. The first set of estimates repeat those from Table 3.3, whilst the second set of results narrows down the sample to under 25s, who are likely to have followed a more standard educational track. The results for under 25s are near identical to the main sample. Individuals with high degree classifications become more likely to train as a teacher relative to those with lower degree classifications after the change in bursaries in 2012. However, the differential changes are nearly identical across subject groupings. In the third set of results, I return to the main sample and add individuals who have studied part-time. Here, the main results change slightly. For non-priority and high priority subjects, those with high degree classifications become about 0.3 percentage points more likely to train as a teacher relative to lower classifications. However, for other priority subjects, there is almost no gradient. In any case, there is no evidence of an increasing differential change for high-degree classifications in high-priority subjects relative to non-priority subjects.

In the fourth set of estimates, I add individuals who completed a postgraduate course. Individuals with a postgraduate qualification would be entitled to the same bursary as an undergraduate with a first-class degree and I thus treat them as having a first-class equivalent degree for the analysis. Here there is a slight decrease in the change in the effect of high vs low degree classifications for high priority subjects (from 0.3 ppts to 0.2ppts). However, the change in the relative effect of degree classifications remains the same for non-priority subjects. Therefore, if anything, there is a larger effect of high degree classifications for low priority subjects when postgraduate degrees are included. These figures then remain largely the same when I include all graduates from across the UK in the final set of estimates.

Table B4 repeats this exercise where the outcome is skill level (as proxied by UCAS tariff scores) rather than propensity to train as a teacher. The under 25 sample gives similar results to the main sample.

When I expand the sample to include part-time students, a slight gap opens up between high-priority and non-priority subjects, with average aptitude in high-priority subjects growing by around 5 UCAS points relative to low-priority subjects after 2012. It then grows further to 7 UCAS points when I include postgraduates, though shrinks again when we look at all graduates from across the UK. However, these gaps are not statistically or economically significant. They are equivalent to less than half a grade at A-Level (the difference between each grade is 20 points).

In summary, it seems as if the choice of sample has little effect on the overall pattern of results, either in terms of the propensity to teach or average aptitude.

3.5.3.2 Time since graduation

The second potential source of bias is the fact that the data only includes graduate destinations 6 months after graduation, and we know this group only accounts for about one third of new entrants to formal postgraduate teacher training courses each year. Table 3.5 therefore examines how the number of teachers and their average skill level changes if we instead examine destinations 3 1/2 years after graduation. This analysis can only be performed for every other cohort and sample sizes are much lower due to lower response rates. Given that individuals may have completed a teacher training course in between the two sample points, I also expand the definition of a teacher to include those actually in a teaching post. The first panel looks at the total number of teachers by year of graduation, whilst the second panel pools data across years to look at the differences across subjects.

Table 3.5: Teacher numbers and aptitude 3.5 years after graduation

	Share of graduates observed as teachers		**Change in numbers between 6 months and 3.5 years**			**Average UCAS Tariff Score**	
	6 months after graduation	**3.5 years after graduation**	**% growth**	**Of which: leavers**	**Of which: new teachers**	**6 months after graduation**	**3.5 years after graduation**
Year of graduation							
2006-07	6.8%	7.9%	45%	30%	75%	315	319
2008-09	8.2%	10.4%	14%	34%	48%	312	320
2010-11	6.7%	9.8%	31%	33%	64%	318	320
Subject of undergraduate degree (pooled across years)							
Physics	2.7%	3.2%	+22%	-39%	+61%	370	374
Maths	12.4%	13.6%	+15%	-34%	+49%	395	386
Computing	2.9%	3.6%	+6%	-50%	+55%	258	244
Chemistry	7.7%	8.4%	+31%	-38%	+69%	342	347
Modern Languages	7.2%	11.0%	+52%	-39%	+91%	373	382
English	12.4%	18.4%	+59%	-28%	+87%	332	343
History	6.6%	11.2%	+87%	-33%	+119%	360	361
Biology	8.4%	13.0%	+54%	-37%	+91%	309	312
Geography	5.8%	9.9%	+61%	-27%	+88%	354	348
Social Science & Law	1.8%	3.7%	+102%	-44%	+146%	339	320
Arts	4.7%	9.1%	+89%	-33%	+122%	305	304

Notes: Author's calculations using HESA Destinations of Leavers from Higher Education (6 months and 3.5 years after graduation from an undergraduate degree). Teachers are defined as those in a teaching job or in training at the survey date. Leavers 3.5 years after graduation are pooled across 2006/07, 2008/09 and 2010/11

As can be seen, the number of observed teachers clearly grows between the two sample points, by an average of 30% across years. This net growth comes from larger number of new teachers than drop outs over the three years. However, it is certainly worth noting that the drop out rate is relatively high with around one third of actual or trainee teachers 6 months after graduation not in a teaching post 3 years later. Such high dropout figures have been noted elsewhere (Allen et al. (2016)). Importantly, the average skill level is very similar amongst teachers 6 months and 3 1/2 years after graduation. This suggests that a focus on teacher trainees 6 months after graduation is unlikely to be a source of bias.

If we look across subjects, there are differential patterns of growth across the two survey points. There is faster growth in teacher numbers for lower priority subjects than high priority subjects. For example, there is about 15-20% net growth in the number of physics and maths teachers, but a near doubling in teachers with social science degrees or history degrees, and around 50% growth in those with degrees in biology, English, geography and modern languages. This lower growth in physics and maths seems to result from fewer additional teachers rather than higher dropout. In terms of aptitude, there are similar average levels of skills by subject for teachers observed 6 months and 3 1/2 years after graduation. Again, this suggests a focus on teachers 6 months after graduation is unlikely to be a major source of bias.

3.5.3.3 Changes in training routes

The final potential source of bias is changes in the relative importance of alternative routes into teaching. There are three main alternative routes to becoming a secondary school teacher. First, a small number of individuals take a specific undergraduate degree, around 5% of secondary school trainees each year. The size of this route is largely unchanged over time. Furthermore, it is unlikely to affect the trends in our data as if choices to take undergraduate courses were affected by changes to bursaries of postgraduates in 2012, then we would not observe any effects until such individuals graduated in 2015. Second, individuals can train as a teacher via Teach First, which seeks to attract high-flying graduates, provide intensive training and place them in hard-to-staff school settings. It is true that by design such individuals are much more likely to have high degree classifications. Indeed, almost all of Teach First trainees have an upper second class degree each year (97% for entrants in 2015-16[7]). However, the numbers of Teach First secondary teacher trainees only increased by about 400 between 2011-12 and 2015-16[8] and numbers are still relatively small in priority subjects (e.g. 23 in physics in 2015-16), making it an unlikely source of bias for the effects of changes to bursaries. However, it is certainly worth noting this route represents an attractive opportunity for teachers in priority subjects.

The main change in routes to teaching has been the significant growth in the government-run "School Direct" programme, which was introduced in 2012 and replaced existing employment-based training routes. Individuals apply to individual schools who provide them with on-the-job training combined with formal education in higher education institutions. School Direct is split between trainees who receive a salary whilst training and some who don't (and are thus eligible for bursaries). If we look at numbers over time, total participants in employment-based training or School Direct have risen from around 6,500 in 2010-11 to reach over 10,000 in 2015-16. However, almost all the growth has been amongst primary school trainees (from around 2,000 to 5,000) with numbers of secondary school trainees growing by less (from just over 4,000

[7]https://www.gov.uk/government/statistics/initial-teacher-training-trainee-number-census-2015-to-2016
[8]https://www.gov.uk/government/collections/statistics-teacher-training

to just over 5,000 in 2015-16).

Unfortunately, data is not available to allow for a clear and consistent comparison of aptitude across formal postgraduate (or provider-led) routes and employment-based training routes. However, I am able to compare the proportions with higher degree classifications across subjects, shown in Figure B2. As one would expect, the proportion with high degree classifications is lower in high-priority subjects. If we compare across routes, there are some differences by training route with a higher proportion of trainees with high-degree classifications in employment-based routes than for provider-led routes in the case of high-priority subjects, and the reverse for lower priority subjects. However, such differences are small and unlikely to be a source of bias, particularly given the relatively small change in numbers of secondary school teacher trainees over time.

3.5.4 Mechanisms and Policy Implications

The empirical results show that the introduction of high-value recruitment incentives were relatively ineffective in changing the willingness of graduates in high-priority subjects to become teachers. The insensitivity of public service workers to financial rewards is not a new finding, with theoretical work suggesting extrinsic motivators can crowd out intrinsic motivation (Besley and Ghatak (2005, 2006); Tirole and Bénabou (2006)). There is also empirical evidence to suggest that awards and social recognition are more effective for intrinsically motivated workers (Ashraf et al. (2014); Kosfeld and Neckermann (2011)). In the context of overall lifetime earnings, the bursaries on offer were relatively small too, representing less than one year's annual earnings.

There is, however, significant evidence showing that financial incentives can be an effective means to encourage teachers to remain in the profession and to move to specific types of schools. For example, Clotfelter et al. (2008) find that a \$1,800 bonus reduced teacher attrition by 17% in high-poverty schools in North Carolina, Feng and Sass (2018b) find that a loan forgiveness programme (worth up to \$2,5000 annually) in Florida reduced teacher attrition by around 10% and a scholarship of \$20,000 over 4 years increases willingness to teach in high-poverty schools in California by around 28%. I offer two explanations for the apparent differential effectiveness of incentives for attracting new teachers and existing teachers.

First, the base population is very different. Incentives to attract new teachers are targeted at all graduates, only 2% of whom train to be a teacher six months after graduation. In the language of a Roy model of occupational choices, individuals willing to train as a teacher are already highly selected and in the the very upper tail of individuals' disposition to teaching. Small recruitment incentives are thus likely to have only a small effect on the total number of individuals willing to train as a teacher. Finding a statistically significant effect is also likely to be challenging empirically. In contrast, retention incentives are targeted at a small group of individuals who have already chosen to become a teacher.

Second, there is significant uncertainty about future earnings at the point of graduation. Individuals actual earnings in different occupations are only revealed as employers become more aware of individuals' actual skills. There is also significant uncertainty about the non-wage benefits individuals will derive from different occupations. At the point of graduation, individuals are comparing relatively uncertain future prospects. In contrast, existing teachers already know their earnings as a teacher and the non-wage benefits they have derived to date. Retention incentives might therefore be more effective for existing teachers as they are effectively compensating for known differentials relative to other occupations.

Whatever the reason for lower effectiveness of recruitment incentives, the clear policy implication is that retention incentives are likely to provide better value-for-money. At present, policymakers in England spend around £135m per year on bursaries. Diverting such funds towards retention incentives will likely have a larger impact on the total number of teachers in post by subject. Policymakers are already making small steps towards such a shift with some of the bursaries for maths teachers now being converted into early career payments of about £5,000 after individuals complete 3 and 5 years of teaching[9]. This will be implemented for teachers who started their training from September 2018 onwards.

3.6 Conclusions

There are widespread and persistent difficulties attracting teachers in maths and science subjects. This is unsurprising given that teacher wages display relatively little variation and that graduates with degrees in maths and science subjects can generally command the highest earnings in the labour market. In this paper, I have sought to better understand the role that financial incentives can play in attracting more highly skilled individuals into teaching in shortage subjects. Despite the differences in graduate earnings by subject, the distribution of teacher aptitude by subject closely resembles the distribution of overall graduate aptitude by subject. I also find little evidence that the introduction of high-value and targeted bursaries increased the propensity of highly skilled individuals to train in shortage subjects and little evidence of any increase in aptitude. Whilst the power of this analysis is limited by relatively wide confidence intervals around year-by-year estimates, the stability of the results by subject and subject grouping by individual year, and the narrower confidence intervals based on pre and post period analysis, lend credibility to results.

Whilst the findings suggest that decisions to become a teacher are relatively insensitive to financial incentives, other empirical evidence suggests that financial incentives for existing teachers can be relatively effective to prevent drop-out (Clotfelter et al. (2008); Feng and Sass (2018b)), increase professional development (Cowan and Goldhaber (2016)) or encourage teachers to move to high-poverty schools (Steele et al. (2010)). This apparent contradiction on the relative effectiveness of financial incentives can be rationalised by the fact that recruitment incentives are targeted at graduates as a whole (only a small number of whom will become teachers) and uncertainty about future rewards before graduates enter the labour market. Whatever the reason, the clear policy implication is that incentives for existing teachers might prove a more effective means than recruitment incentives to ensure sufficient numbers of teachers in shortage subjects, particularly in England where dropout and attrition rates are relatively high (Allen et al. (2016))

A further implication of my findings is that non-financial incentives might prove to be a more effective means to recruit and retain highly skilled individuals in shortage subjects. For instance, Teach First and Teach for America have proven to be relatively effective at attracting highly skilled graduates largely through strategies that raise the profile and status of their trainees. Future research should seek to test the impact of different types of non-financial incentives in recruiting highly skilled individuals

[9]https://www.gov.uk/guidance/mathematics-early-career-payments-guidance-for-teachers-and-schools

Chapter 4

Effect of an area-wide campaign to improve the use of teaching assistants: an application of synthetic control methods

Abstract

The use of teaching assistants or aides has become more prevalent across a range of countries. Empirical evidence suggests they can improve student achievement where they are used well, but are often used in poor ways. This paper uses the synthetic control approach to estimate the effects of a campaign to improve the ways in which teaching assistants are used across a large, disadvantaged area of England. The results suggest the campaign increased English scores by a modest amount of about 0.03-0.04 standard deviations, though there is no evidence of an improvement in maths. In most specifications, the positive effect on English scores is large relative to simulated, placebo treatment effects estimated across other school districts. If there is an effect, a larger effect on English scores would be consistent with existing evidence showing that there is greater scope for improving the effect of teaching assistants on literacy outcomes. A major advantage of the synthetic control approach is being able to relax assumptions on the balance of unobservables and parallel trends. In this case, the impact estimates are larger under the synthetic control approach than under other non-experimental estimators, such as matching and difference-in-differences. The synthetic control approach should be used more across education evaluations.[1]

JEL Classifications: I2, J3, J4

Keywords: Synthetic Control, Evaluation Methods, School Quality

[1]I am grateful to the ESRC for financial support as part of the UCL IoE Doctoral Training Centre and the Education Endowment Foundation for providing funding and support. I thank Barbara Sianesi and Guillermo Rodriguez-Guzman for their comments and feedback. I am also grateful to the Department for Education for provision of the National Pupil Database. This paper is partly based on an independent evaluation commissioned by the Education Endowment Foundation, which has already been published. A pre-specified Analysis plan was also published, which is largely followed. The exceptions are stated clearly. Neither the funders nor the data owner accept responsibility for any inferences or conclusions derived in this work. All errors, conclusions and interpretation remain those of the author.

4.1 Introduction

The use of teaching assistants (TAs) or teaching aides has expanded across a range of countries. As they are generally low paid, this can represent a relatively inexpensive way of increasing the number of instructional staff within classrooms. The main roles of TAs are to provide one-to-one assistance to specific pupils who are under-performing or who have special educational needs, and to provide general support to teachers. In England, the number of TAs in English schools has greatly expanded over time, particularly in primary schools. There are now about 260,000 TAs across schools in England, up from about 50,000 in the mid-1990s and they comprise over 25% of the school workforce[2]. Evidence shows that TAs are primarily used in disadvantaged schools (Cook et al. (2011)). Indeed, most of the sizable increase in school funding targeted at deprivation in England has been spent on employing extra TAs (Sibieta, 2015).

Unfortunately, the evidence on the impact of TAs is mixed, and depends on the way which they are used. Blatchford et al. (2011) and Farrell et al. (2010) find that pupils attached to TAs make little extra educational progress than other equivalent pupils, and can sometimes make less progress. Such evidence also finds that TAs in England are often used for general practical or administrative tasks, and that one-to-one time with TAs often diverts from time with a more highly qualified teacher. More recent work has emphasised the value of a model where teachers and TAs share instructional responsibility in the classroom (Friend and Cook (2016)). Andersen et al. (2020) estimates the effect of an experiment that randomly allocated additional teacher aides with and without teaching degrees to schools in Denmark. This found positive effects of both types of TAs on literacy scores, with little effect on maths scores. These effects were larger when instructional responsibility was shared between teachers and TAs. Penney (2018) re-examines the STAR experiment to find positive effects of TAs on student test scores, with larger effects for those with higher socio-economic status. There is also clear evidence that structured, high-dosage tutoring delivered by TAs can improve student achievement (Fryer (2014, 2016); NFER (2014); Gorard et al. (2014)). Changing the ways in which TAs are used is therefore key to improving their impact on student achievement.

This paper seek to estimate the effect of an area-wide campaign to improve the ways in which TAs are used across a large, disadvantaged area of England. This formed part of a national campaign started in 2015 by the Education Endowment Foundation (EEF), an education charity in England, to improve the ways in which TAs are deployed. This included a guidance report that was disseminated to all state-funded schools in England[3]. This guidance included recommendations to use TAs to deliver structured one-to-one catch up interventions, use them more in whole class teaching, ensure pupils are not diverted away from teachers and to provide better training.

EEF implemented a specific scale-up campaign across primary schools in South and West Yorkshire (S&W Yorkshire) to increase uptake of this guidance and set of recommendations. Advocacy providers (APs) were recruited for each local authority in the region. These APs were generally collaborations between organisations and individuals with significant experience in school leadership or training. APs recruited schools to their programme with a launch event at the start of the 2015-16 academic year and then ran a series of workshops for head teachers and leaders focused on implementing the guidance and recommendations. In addition, each AP ran a set of tailored activities, such as additional training or mentoring of individual schools. This paper focuses on the effect of this area-wide scale-up campaign in S&W Yorkshire.

[2]https://www.gov.uk/government/statistics/school-workforce-in-england-november-2018
[3]https://educationendowmentfoundation.org.uk/tools/guidance-reports/making-best-use-of-teaching-assistants/

Unfortunately, it is notoriously hard to estimate the effect of area-wide policy changes or initiatives. They are often deliberately targeted at under-performing areas or those likely to benefit from the policy, which means that areas affected by a policy change are often different in ways unobservable to researchers (Blundell and Dias (2009)). The most common ways in which area-wide policies are evaluated are through matching based on contemporaneous data, difference–in-differences (DiD) estimators or some combination of the two (Hutchings et al. (2012); Machin et al. (2010); Machin and McNally (2008)). Matching techniques rely on balance in unobservables, which is not always a convincing assumption, whilst DiD estimators rely on an untestable assumption of parallel trends in unobservable factors.

In order to estimate the causal effect of the TA campaign in S&W Yorkshire, I make use of recently developed Synthetic Control methods. This approach seeks to find a set of time-constant weights that best approximate outcomes in the treatment group during a pre-treatment phase. The identifying assumptions are weaker than other non-experimental methods in that Synthetic Control methods implicitly allow the effects of unobservable factors to change over time. It is therefore more suitable than DiD to cases where the common trends assumption is likely to be violated.

Synthetic control methods were pioneered in a series of papers aimed at evaluating the effects of Basque terrorism, cigarette taxes in California and the economic effects of German reunification (Abadie and Gardeazabal (2003); Abadie et al. (2010, 2015)). The methods have also been used to evaluate a diverse range of policies and phenomena, including the effects of Hugo Chavez's economic policies (Grier and Maynard (2016)); terrorism on house prices (Gautier et al. (2009)), economic liberalization (Billmeier and Nannicini (2013)), natural disasters (Cavallo et al. (2013)), employment regulation on immigrant workers (Bohn et al. (2014)) and have been used to revisit the effects of the Mariel boatlift in Florida (Peri and Yasenov (2019)). The only applications to education policy to date, however, have been to examine the effects of affirmative action and school nutrition policies (Hinrichs (2012); Bauhoff (2014))

Inference methods have been developed based on permutation or placebo tests, with p-values derived from the distribution of placebo effects estimated for control units (Abadie et al. (2010)). The methods have also been developed for the case of multiple treatment units, with Kreif et al. (2016) proposing using them as a single treatment unit and Acemoglu et al. (2016) proposing creating synthetic controls for each individual treatment unit and then aggregating.

A contribution of this paper is to show how synthetic control methods can be applied to estimate the effects of area-wide education policies and how the estimated effects differ from those that would be found through more traditional non-experimental methods. It also illustrates the effects of various methodological choices, such as the specification of control factors, and compares the Krief et al and Acemoglu et al methods for dealing with multiple treatment units.

The TA campaign in S&W Yorkshire is ideally suited to the application of Synthetic Control analysis. In summer 2015, EEF sought to recruit APs for each of the nine local authorities in S&W Yorkshire to help primary schools implement EEF's guidance on making the best use of TAs. The choice of S&W Yorkshire was very deliberate in that it was an area with high levels of educational under-performance. Advocacy providers could not be found for all local authorities, with 7 advocacy providers recruited in total. They were expected to support schools in understanding and implementing the EEF guidance. Advocacy providers differed in how they achieved this aim, but all ran events, workshops and provided tailored support to individual schools. Schools were also offered greater opportunities to participate in randomised control trials focused on TA-led structured one-to-one interventions for under-performing pupils. I estimate the effects of this policy on an intention to treat (ITT) basis as it was intended to affect all primary schools in S&W Yorkshire. I focus on estimating the effects on Key Stage 2 English and maths scores for pupils at the end of primary school (aged 11) in 2017, leaving one year between the time of treatment (2015-16 academic year) and the final outcomes (2016-17 academic year). This leaves time for schools to change their practices and for these practices to affect outcomes.

Using Synthetic Control methods, the results suggest that the campaign in S&W Yorkshire had minimal effects on maths scores at age 11, but improved English scores across the whole region by about 0.035 standard deviations. The pseudo p-value for the effect on English scores is about 0.1, placing it on the boundary of statistical significance at the 10% level. This estimate is largely unchanged when varying the set of local authorities that can form potential control units or which predictor variables we use to construct the synthetic control weights. In most cases, the p-value varies from 0.02 to 0.16. When including many years of lagged outcomes, the estimate falls and is no longer on the boundaries of statistical significance, but the estimates and inference procedure seem to perform poorly when many years of lagged outcomes are added. The Kreif approach for multiple control units performs significantly better than the Acemoglu approach in replicating outcomes during the pre-treatment phase in this context of small numbers of control units. The results differ from those that would be found using matching and DiD techniques, which tend to be negative or close to zero. This illustrates the value of the Synthetic Control approach in relaxing the common trends assumption.

The larger effects on English than maths are consistent with previous evidence showing that existing negative effects of teaching assistants are larger for English than for maths, suggesting a greater scope for improvement in English (Blatchford et al. (2011)). TA-led interventions also tend to have more positive effects for English than for maths (Alborz et al. (2009); Andersen et al. (2020)). This may suggest than English and literacy outcomes are more malleable to the roles played by TAs than maths and numeracy outcomes, with more research needed as to why this might be the case.

The results are mainly driven by improvements in English scores among less deprived pupils, those without special educational needs and those with English as an additional language. The greater improvements among less deprived pupils and those without special educational may seem slightly surprising, given these pupils are less likely to interact with TAs. However, one of the main recommendations of the guidance was to involve TAs in whole class teaching to a greater degree and reduce the extent to which they are exclusively attached to specific groups of pupils with special or additional educational needs. Schools in the trial changed their practices in line with this recommendation and the pattern of improvements by sub-group is fully consistent with this change. Indeed, on-treatment analysis suggests that any effect is more likely to be driven by the advocacy and training part of the campaign, as opposed to the greater availability of RCTs and TA-led interventions, given that schools involved in these additional trials had similar outcomes to those that weren't involved. These results are also in line with the pattern of results in Penney (2018).

This paper also relates to the literature on the effects of school resources. Previous surveys of the literature on effects of school resources found little consistent evidence of a positive effect of school resources on pupil attainment, concluding that the way in which resources are used is more important (Hanushek (2003)). The recent literature has revised this conclusion and suggests consistent evidence of a positive effect of school resources, particularly among disadvantaged pupils (Jackson et al. (2016)). However, there remains significant interest in ways in which resources can be better deployed, with the results in this paper suggesting that changing the way in which TAs are used (which account for over a quarter of school staff) can yield significant benefits in terms of pupil attainment.

The rest of this paper proceeds as follows. Section 4.2 provides more background on the TA campaign in S&W Yorkshire. Section 4.3 details the Synthetic Control approach as a response to the evaluation problem and compares this with other methods. Section 4.4 presents the main analysis, compares results with other non-experimental methods, and discusses and interprets the mechanisms driving the results. Section 4.5 concludes.

4.2 Background: Teaching Assistant Campaign

4.2.1 Teaching Assistants in England

Teaching Assistants (TAs) make up approximately one quarter of the school workforce in England. Their numbers have risen significantly over the last two decades, with around 50,000 TAs in 1995 up to 150,000 in 2005 and over 260,000 at the latest count in 2017[4]. Indeed, Sibieta (2015a) finds that most of the increase in funding targeted at disadvantaged schools over this period was spent on employing extra TAs. Part of this rise was deliberate with TAs intended to help teachers in the classroom, such as one-to-one support to individual pupils with specific needs, and to help teachers with general or administrative tasks. However, some of the rise is likely to have arisen because TAs can be employed and used by schools on flexible terms.

Unfortunately, existing empirical evidence suggests that the current deployment of TAs in schools is not improving pupil outcomes, and can even be detrimental to learning. Blatchford et al. (2011) show that pupils who spend more time with TAs make less progress in English and maths, even after controlling for pupil characteristics, with the largest negative effects for English. Part of this effect may be driven by the fact that TAs tend to be less qualified than teachers, with one third of TAs possessing degree level qualifications as compared with almost all teachers. Substituting time with a qualified teacher to time with a less qualified TA could reduce pupil attainment. However, some of the effect may also be driven by the way in which TAs are deployed. Indeed, there is now a wide evidence base suggesting positive ways TAs can be deployed. Andersen et al. (2020) finds positive effects of TAs on literacy scores in Denmark, particularly when instructional responsibility was shared between teachers and TAs. Penney (2018) use the STAR experiment to find positive effects of TAs on student test scores. There is also clear evidence that structured, high-dosage tutoring delivered by TAs can improve student achievement (Fryer (2014, 2016); NFER (2014); Gorard et al. (2014)). Furthermore, when TAs are used to provide one-to-one or small group support using structured interventions, they can generate average gains of between three and four additional months, e.g. Switch-on Reading programme (Gorard et al. (2014)) and the Catch Up Numeracy programme (NFER (2014)).

4.2.2 The Teaching Assistant Campaign

In summer 2015, the Education Endowment Foundation (EEF) launched a national campaign to improve the ways in which TAs are used[5]. They produced a guidance document setting out ways that schools can maximise the impact of teaching assistants and launched a multi-stranded campaign to increase uptake of this advice and guidance. The main recommendations were:

[4]https://www.gov.uk/government/statistics/school-workforce-in-england-november-2018
[5]https://educationendowmentfoundation.org.uk/tools/guidance-reports/making-best-use-of-teaching-assistants

1. TAs should not be used as an informal teaching resource for low-attaining pupils
2. Use TAs to add value to what teachers do, not replace them
3. Use TAs to help pupils develop independent learning skills and manage their own learning
4. Ensure TAs are fully prepared for their role in the classroom
5. Use TAs to deliver high-quality one-to-one and small group support using structured interventions
6. Adopt evidence-based interventions to support TAs in their small group and one-to-one instruction.
7. Ensure explicit connections are made between learning from everyday classroom teaching and structured intervention

Recommendations (1)-(4) focus on ways to improve the use of TAs in classroom settings and to develop strong links between teacher and TA activity, ensuring TAs are not just used for informal administrative tasks and that time with TAs does not divert from time with more qualified teachers. This is very much in line with the shared responsibility model described by Friend and Cook (2016) and Andersen et al. (2020). Recommendations (5)-(7) propose greater use of TAs to deliver structured one-to-one TA-led interventions backed up by empirical evidence. Seven specific interventions were recommended by EEF, with six of these relating to literacy or communication and one relating to numeracy. This links to the largely positive literature on the effects of intensive, high-dosage tutoring and one-to-one or small group support (Fryer (2014, 2016)).

As part of the national campaign, hard copies of the guidance report were sent to all schools in England, EEF emailed all schools with links to the guidance and there were various events and widespread social media promotion. Schools all over the country reported high levels of engagement with this campaign, with 80% of head teachers claiming to have read it and 95% of these saying the guidance had been helpful[6]. However, it is not clear how many schools actually changed their practices and behaviours in response to such general guidance.

Alongside the national campaign, EEF therefore launched a more intensive programme of advocacy and training targeted at increasing uptake of this guidance in a single region. The chosen region was South and West Yorkshire (S&W Yorkshire) and the campaign was focused on primary schools. S&W Yorkshire comprises nine different local authorities[7] covering about 350,000 pupils in 1,050 primary schools (which serve pupils between the ages of 4 and 11). S&W Yorkshire was chosen given a relatively high share of disadvantaged pupils and low levels of pupil attainment. Primary schools were chosen as the focus as these schools tend to have a higher share of TAs than secondary schools. This paper's main research question concerns estimating the impact of this intensive advocacy campaign.

In this campaign in S&W Yorkshire, advocacy providers (APs) were appointed to run programmes for schools in each local authority from September 2015 to July 2016. APs had to bid to EEF with a list of proposed activities to be run for single local authorities. In all cases, the aim was to provide a programme of events, coaching and training designed to increase uptake of the guidance and associated recommendations. Recruited advocacy providers were generally collaborations between individuals and organisations

[6]https://educationendowmentfoundation.org.uk/news/school-survey-eef-teaching-assistants-guidance/
[7]Sheffield, Rotherham, Doncaster, Barnsley, Leeds, Wakefield, Calderdale, Kirklees, Bradford

with experience of leadership within the system or who with significant experience of providing training (e.g. universities, local authorities, teaching school alliances, head teachers, school improvement consultants).

The proposed and actual activities differed across APs, which are fully documented in a process evaluation of the campaign[8]. They all included a launch event, which was used to recruit schools to each AP's advocacy campaign, alongside other correspondence. Recruited schools were then invited to a core set workshops. These were designed to engage schools with the EEF recommendations for making the best use of teaching assistants and implementing the recommendations in their school, such as presentations from the APs, presentations of exemplar practices by schools that had implemented the guidance and discussions with authors of the guidance document. These workshops were attended by head teachers or senior teachers within schools. All of the APs were focused on persuading and effecting change at the leadership level. In addition, many APs undertook additional activities to further improve uptake of the guidance, such as training, school visits and mentoring to individual schools. Most of the APs also held a celebration event at the end of the programme to showcase and embed best practice.

Given the different activities employed, the cost of the programme varied across each AP. However, it is estimated that the total cost was about £27,000-£38,000 for each AP or about £400 per recruited school[9]. This includes the cost of staff for each AP, the cost of events/workshops and any additional activities.

Primary schools in S&W Yorkshire were also offered additional opportunities to partake in trials to test the effectiveness of structured TA-led interventions. These were intended to meet additional demand for such interventions from schools in S&W Yorkshire seeking to act on the campaign recommendations. These trials were targeted at schools with high numbers of disadvantaged pupils in S&W Yorkshire. Each of these trials was subject to a separate evaluation.

Originally, there was a target to work with about 525 state-funded primary school schools across all nine local authorities (about half of the 1,050 primary schools in the area), focusing on disadvantaged and under-performing schools. Whilst the campaign managed to work with about 480 schools, advocacy providers could only be recruited in seven of the nine local authorities and the campaign as a whole was less successful in focusing on disadvantaged or under-performing schools.

4.2.3 Effect of advocacy campaign in South and West Yorkshire

This paper's main research question concerns estimating the effect of the offer of intensive advocacy and guidance across primary schools across S&W Yorkshire. Included within this is the effect of the additional opportunities to partake in trials of TA-led interventions in S&W Yorkshire. In order to estimate this effect, one requires a counterfactual group with similar characteristics and subject to similar time trends. This poses significant estimation challenges as S&W Yorkshire was chosen deliberately as a region with low attainment and an area likely to benefit from the campaign. As a result, one cannot implement a randomized controlled trial.

This is similar to the challenges in estimating the effect of school improvement programmes spread across wide and deliberately chosen areas. Such studies have generally made use of either matching or difference-in-difference techniques to find a valid counterfactual. For example, Hutchings et al. (2012) evaluate the impact of the City Challenge programmes in England by comparing changes in pupil attainment

[8]https://educationendowmentfoundation.org.uk/public/files/Campaigns/SW_Yorks_TA_final_report_final.pdf
[9]https://educationendowmentfoundation.org.uk/public/files/Campaigns/TA_campaign_IFS_report.pdf

in City Challenges areas with changes across other areas, Muijs (2015) studies the impact of primary school partnerships using propensity score matching and Machin et al. (2010) use a difference-in-differences approach to estimate the effect of the Excellence in Cities programme. In the next section, I argue that the identification assumptions underlying these methods are unlikely to hold in practice for many area-wide education initiatives. Instead, the assumptions underlying 'synthetic control' methods are likely to be more flexible and plausible. Such methods have only been used in a relatively small number of education settings, e.g. Hinrichs (2012) uses synthetic control methods to study the effect of affirmative action.

4.3 Empirical Methods

This section considers the general problem of estimating the effect of an area-wide programme. This is used to motivate the synthetic control approach and how this compares with other non-experimental estimation techniques. Section 4.4 applies synthetic control methods to estimation of the effect of the TA campaign in S&W Yorkshire and compares with results given by other non-experimental methods.

4.3.1 General problems with estimating the effect of area-wide policies

Let us assume that the production function for educational attainment takes the form of equation (4.1), where the educational attainment of individual (i) at time (t) is a function of the full history of school ($\boldsymbol{S}^{(it)}$) and parental investments ($\boldsymbol{P}^{(it)}$) up until time (t). For the sake of notational convenience, I assume that any ability endowments are included within parental investments.

$$Y_{it} = f_t\left(\boldsymbol{S}^{(it)}, \boldsymbol{P}^{(it)}\right) \tag{4.1}$$

School and parental investments can be separated out into national-level factors (n), area-level factors (a) and school-level factors (s).

$$\boldsymbol{S}^{(it)} = \left[\boldsymbol{s}_n^{(it)}, \boldsymbol{s}_a^{(it)}, \boldsymbol{s}_s^{(it)}\right] \tag{4.2}$$

$$\boldsymbol{P}^{(it)} = \left[\boldsymbol{p}_n^{(it)}, \boldsymbol{p}_a^{(it)}, \boldsymbol{p}_s^{(it)}\right] \tag{4.3}$$

National-level school investments could include the overall level of education spending or school accountability structures, whilst national parental investments are likely to include the overall state of the economy or employment. Area-wide parental investments are likely to relate to local economic or industrial trends, whilst area-wide school investments might relate to levels of local school spending or initiatives/policies of local school districts. School-specific investments are likely to include levels of teacher quality and individual school policies, whilst school-specific parental investments will reflect the collection of individual family circumstances and attitudes of families whose children attend that school.

Many policies or initiatives are targeted at the area-level. This is sometimes because they are set at the area-level by local policymakers or targeted at specific areas because they are under-performing or disadvantaged in some way. There is therefore great interest in ways to estimate the effect of area-wide policies. A standard way to do this is via the difference-in-differences (DiD) estimator. By differencing area-level outcomes before and after a treatment or policy, one can remove national or macro trends in school and parental investments over time, as well as any time constant differences across areas in parental and school investments. However, such a method is biased if there are area-specific time trends in school or parental investments.

Following Abadie et al. (2010), this problem can be easily expressed within a standard potential outcomes framework. Let there be A+ 1 regions, with the first unit $a = 0$ representing the treated area. Outcomes are observed for T periods, with the treatment taking place in region $a = 0$ at time $t = T^0 + 1$. The observed vector of outcomes for region a is $Y_a = \left(Y_{a,1}...Y_{a,T^0}...Y_{a,T}\right)'$. The observed outcome at time (t) in region (a) is a function of the outcome in the absence of the treatment (Y_{at}^N) and the effect of the treatment (α_{at}), where $D_{at} = 1$ iff $a = 0$ and $t > T^0$.

$$Y_{at} = Y_{at}^N + \alpha_{at} D_{at} \tag{4.4}$$

$$Y_{at}^N = \delta_t + \theta_t \boldsymbol{Z}_a + \lambda_t \boldsymbol{\mu}_a + \varepsilon_{at} \tag{4.5}$$

Outcomes in the absence of the treatment are a function of aggregate time effects (δ_t), incorporating national-level school and parental investments, a vector of observed time-constant area-characteristics ($\boldsymbol{Z}_a$), whose effect may vary over time (θ_t), a vector of unobserved area characteristics ($\boldsymbol{\mu}_a$), whose effect may also vary over time (λ_t) and unobserved mean zero transitory shocks (ε_{at}). This model incorporates the standard DiD model if one constrains the effect of unobserved area-factors to be constant over time. However, it also allows for area-specific trends as a result of the changing effect of area-factors over time, in which case the DiD estimator is biased as a result of a violation of the parallel trends assumption. This then motivates the more flexible synthetic control approach

4.3.2 Synthetic Control Approach

The synthetic control approach was first proposed by Abadie and Gardeazabal (2003) to analyse the economic effects of Basque Terrorism and has since been extended by Abadie et al. (2010) in their study of cigarette taxes in California. The overall idea is to estimate the effects of policies and treatments on aggregate units by finding a set of time-constant weights across control units that best approximate the treatment unit in the pre-treatment phase.

More formally, let the vector of weights used in the synthetic control approach be $W = (w_1...w_A)'$, such that $\sum_{a=1}^{A} w_a = 1$ and $w_a \geq 0$. An estimator of the outcome in the absence of the treatment is then $\hat{Y}_{0,T^0+1}^N = \sum_{a=1}^{A} w_a Y_{a,T^0+1}$ and an estimate of the treatment effect is $\hat{\alpha}_{at} = Y_{0,T^0+1} - \hat{Y}_{0,T^0+1}^N$. Abadie et al. (2010) show this is an unbiased predictor of the treatment effect if the number of pre-treatment periods is large relative to the scale of the idyosyncratic error. In this case, the synthetic control approach is able to reproduce the time-varying effects of both the observed and unobserved predictors. In particular, if one can find a set of weights such that $\sum_{a=1}^{A} w_a Z_a = Z_0$ and $\sum_{a=1}^{A} w_a Y_{a,t} = Y_{0,t}, t = 0...T^0$ then $\hat{\alpha}_{at}$ is an unbiased predictor.

In order to find the set of weights W, Abadie et al. (2010) propose minimising the following distance metric:

$$d = \sqrt{(X_1 - X_0 W)' \, V \, (X_1 - X_0 W)} \tag{4.6}$$

where X_0 is k x 1 metric of covariates, including pre-treatment outcomes and predictor variables for the treated area, with X_1 representing an equivalent k x A matrix for control areas. Both what variables go into X and the weight attached to them in the matrix V can be a subjective decision, justified by knowledge of the process driving outcomes. Abadie et al. (2010) propose choosing V and W in order to minimise the root mean squared prediction error and this approach has largely been followed in the literature.

In seeking to apply this method to estimating the effect of the teaching assistant campaign, there are two main challenges: the case of multiple treated units; and, inference. I address each issue in turn.

4.3.3 Multiple Treatment Units

The teaching assistant campaign was targeted at nine local authorities across South and West Yorkshire. One therefore needs to consider the implementation of the synthetic control approach in the context of multiple treated units[10]. Let there be A units in total, with A_0 treated units and A_1 control units. Using synthetic control methods in such a context has been considered by Acemoglu et al. (2016) and Kreif et al. (2016) in slightly different ways.

Acemoglu et al. (2016) take the set of treated units within A_0, find a synthetic control weights for each individual treated unit W_a and an associated treatment estimate $\hat{\alpha}_{at}, a = 1...A^0$. They then aggregate these treatment estimators to produce an overall estimate of the treatment effect $\hat{\alpha}_{A^0,T^{0+1}} = \sum_{a=0}^{A^0} \hat{\alpha}_{a,T^{0+1}}$.

In contrast, Kreif et al. (2016) aggregate the outcome for the treated units first and then find a synthetic control for the aggregated treated region. In particular, they aggregate equation (4.4) to get:

$$\bar{Y}_t = \bar{Y}_t^N + \bar{\alpha}_t D \tag{4.7}$$

where $\bar{Y}_t$, $\bar{Y}_t^N$and $\bar{\alpha}_t$ represent the weighted sum of the observed outcome, potential outcome and average treatment effect at time t, respectively. They then seek to find a set of weights across the remaining control units that best approximates the potential outcome across the treated region.

$$\hat{\bar{Y}}_t^N = \sum_{a=A^0+1}^{A^0+A^1} w_a Y_{t,a} \tag{4.8}$$

I apply both methods, with the Kreif et al. (2016) representing the preferred approach. This is because the Acemoglu et al. (2016) approach performs poorly in this setting, with relatively high prediction errors. This is likely the result of difficulties in creating synthetic controls for small units, subject to particular time trends and with predictor variable values close to or outside the boundaries of common support.

4.3.4 Inference

The next challenge is conducting inference. Whilst there is no uncertainty in the value of the aggregate units, there is uncertainty in the extent to which they can capture the potential outcomes in the absence of

[10]An alternative would be to use regions as our units of analysis. However, there are less than 10 of these, which would make it empirically challenging to find an appropriate set of weights.

treatment. Abadie et al. (2010) capture this uncertainty through placebo or permutation tests. They take each unit within the control group, estimate synthetic control weights for each unit and treatment effect estimators. This allows them to create a distribution of treatment estimators for the treatment period, as well as treatment effects relative to the prediction errors during the control phase. These can then be used to create standard errors, p-values or confidence intervals.

This approach is slightly complicated by the presence of multiple treated units. In this case, the aggregated treated unit (consisting nine local authorities) is much larger then individual control units. It is therefore much easier to reproduce the potential outcome for the treated aggregate unit than it is for individual control units. As a result, I apply the simulation method proposed in Kreif et al. (2016). They randomly select (with replacement) the same number of units from the control group as are in the aggregated treated group (nine in our case). They then treat this simulated group as the aggregate treated region and proceed with estimation in an identical fashion. They repeat the simulation process and use these simulations to create standard errors and confidence intervals. I follow this approach across different specifications for the synthetic controls.

4.4 Synthetic Control Analysis

This section presents the main results. It starts by describing the data sources and sample (4.4.1), creation of the synthetic controls (4.4.2) and then presents the main impact estimates (4.4.3). I test the robustness of the impact estimates to different ways to construct synthetic controls. In section 4.4.4, I then compare my preferred synthetic control estimates with those produced by other non-experimental estimators (OLS, matching, difference-in-differences) and relate these to the identification assumptions underlying the different methods presented in the previous section. I then analyse the potential mechanisms driving the results as part of sub-group analysis (4.4.5) and overall interpretation of the results (4.4.6).

My preferred methods and estimates largely conform to a pre-specified Statistical Analysis Plan that was compiled before the data became available and originally published in an independent evaluation report for the Education Endowment Foundation (EEF)[11]. There are two main exceptions. First, the preferred sample here is slightly different. In particular, I use the sample that delivers the lowest prediction errors for English as opposed to maths. The evaluation report used the preferred sample for maths instead. I focus on the preferred sample for English largely because the estimates are of greater interest for English. However, I present results for both samples in any case as part of robustness checks, and this has no material effect on the pattern of results. Second, I employ the Kreif et al. (2016) method for conducting inference in the context of multiple treatment units. This accounts for the fact that it is easier to construct synthetic controls for large regions than is is for small individual local authorities. The original evaluation report and analysis plan did not account for this and instead conducted inference by finding placebo treatment effects for individual local authorities. The main effect of this change is to increase the p-value on the main estimate for English from 0.07 to 0.1.

[11]https://educationendowmentfoundation.org.uk/scaling-up-evidence/campaigns/making-best-use-of-teaching-assistants/evaluation/

4.4.1 Data description and sample selection

All the analysis is based on pupil-level data from the National Pupil Database (NPD) for England. The NPD is an administrative census that includes pupil demographics and external examination results for all pupils in state-funded schools in England[12]. I use data from academic years 2002-03 to 2016-17. The period from 2002-03 to 2014-15 is classified as the pre-treatment phase and 2015-16 to 2016-17 as the intervention period, with 2016-17 representing the primary outcome date of interest. I focus on outcomes two years after the treatment began to allow for delays between the provision of guidance, changes in practice and impacts on test outcomes.

The main outcomes of interest are Key Stage 2 test results in maths and English taken at the end of primary school when pupils are aged 11. These are the only externally-assessed tests taken by all pupils in state-funded primary schools in England. They represent the main outcomes used for school accountability for primary schools, national targets and are a key consideration in school inspections. The analysis uses average points scored in the maths and English tests as the main outcomes (standardised at the national level). It also considers threshold outcomes: the proportion of pupils achieving the expected level in each subject over time (Level 4 or above up to 2014-15 and a score of 100 or more from 2015-16 onward).

There were major changes to Key Stage 2 tests for pupils taking them from 2015-16 onward (the first year of the treatment phase). These included reforms to the curriculum, assessments and the way tests are scored. For example, the proportion of pupils achieving Level 4 or above in maths in 2014-15 was 89%, whilst the proportion achieving the expected score of 100 or more in maths was 70% in 2015-16 and 75% in 2016-17. This timing is unfortunate from the perspective of this analysis. The approach will select the synthetic control group to best approximate the treatment group over the whole course of the pre-treatment period. One might therefore hypothesise that it will respond to the new tests in the same as the treatment. However, there were no changes to the tests of this magnitude during the pre-treatment phase. It is therefore impossible to predict with any degree of certainty the plausible effects of changes to the tests. The effect could be small or large, and any bias could be positive or negative.

I use a range of pupil and school characteristics as predictors and covariates. This includes pupil demographics from the NPD: % eligible for free school meals; % with English as an additional language; % of White-British ethnicity; % Black ethnicity; % Asian ethnicity; % of pupils with Statement of Special Educational Needs). In addition I calculate the proportion of pupils in Academies from 2010-11 onward. Although there were a small number of Academies up to 2010, after that point schools were able and encouraged to convert to Academy status. Such schools have more freedom over decision-making, including pay and conditions of staff. Given the focus of the campaign, I also include controls for the ratio of Teaching Assistants to pupils in each school over time. This data is only available from 2011 onward. I do not include data on funding as the fast pace of conversions to Academy status over time meant that many schools were temporarily missing from funding data during the period of conversion.

All data is aggregated at the local authority level over time because the focus of the treatment was at a local-authority level. This also allows for area-wide effects and ensures a balanced panel over time. This produces a panel of 152 local authorities over 16 years. A school-level panel would not be balanced due to school openings and closures over time. A region panel would not have sufficient sample size as there are

[12]Pupil in private schools who opt to take Key Stage 2 tests at age 11 are also included, but these pupils are dropped from the analysis as participation in these tests is voluntary for private schools and data on key demographics is also missing for such pupils.

only 9 regions in England. As described in the previous section, I allow for multiple treatment units in the analysis by aggregating the nine local authorities to a single treatment unit and then simulating groups of nine local authorities in the inference procedure.

In creating the local authority panel, I exclude any pupils from independent or special schools (which were not part of the campaign). As shown in Table 4.1, this amounts to about 5% of the original sample taking the test in 2002-03, falling to 2% by 2016-17. This fall over time is likely to be due to fewer pupils in special and independent schools sitting Key Stage 2 tests over time. I also exclude pupils with missing Key Stage 2 or pupil covariate data. This equates to about 20-30,000 or about 1-4% of pupils in most years. The one exception is 2010 where nearly 30% of pupils have missing data owing to the boycott of Key Stage 2 in that year. This has minimal effect on the analysis as 2010 is not used as one of the key predictor variables.

Excluding pupils with missing data is unlikely to be a source of bias in this context as these variables are only used to predict lagged outcomes and the synthetic control analysis seeks to minimise differences here in any case.

Table 4.1: Sample sizes under various sample restrictions

Outcome	**Original Sample**	**School Exclusions**	% of original sample	**Missing Pupil Data**	% of original sample	**Local Authority Exclusions**	% of original sample	**Final Sample**	% of original sample
2002-03	**637,675**	**29,528**	4.6%	**7,696**	1.2%	**27,625**	4.3%	**572,826**	89.8%
2003-04	**613,726**	**27,107**	4.4%	**6,170**	1.0%	**27,278**	4.4%	**553,171**	90.1%
2004-05	**610,918**	**25,630**	4.2%	**15,089**	2.5%	**26,722**	4.4%	**543,477**	89.0%
2005-06	**595,306**	**22,377**	3.8%	**11,655**	2.0%	**26,964**	4.5%	**534,310**	89.8%
2006-07	**588,122**	**22,239**	3.8%	**12,121**	2.1%	**27,208**	4.6%	**526,554**	89.5%
2007-08	**597,332**	**21,233**	3.6%	**13,217**	2.2%	**27,436**	4.6%	**535,446**	89.6%
2008-09	**580,033**	**18,596**	3.2%	**11,622**	2.0%	**26,657**	4.6%	**523,158**	90.2%
2009-10	**571,341**	**14,903**	2.6%	**157,010**	27.5%	**19,400**	3.4%	**380,028**	66.5%
2010-11	**554,888**	**13,122**	2.4%	**9,013**	1.6%	**27,221**	4.9%	**505,532**	91.1%
2011-12	**544,222**	**12,047**	2.2%	**6,508**	1.2%	**27,217**	5.0%	**498,450**	91.6%
2012-13	**540,197**	**11,963**	2.2%	**7,647**	1.4%	**27,413**	5.1%	**493,174**	91.3%
2013-14	**561,543**	**12,712**	2.3%	**7,675**	1.4%	**28,623**	5.1%	**512,533**	91.3%
2014-15	**579,263**	**12,888**	2.2%	**7,412**	1.3%	**29,509**	5.1%	**529,454**	91.4%
2015-16	**592,272**	**13,061**	2.2%	**22,477**	3.8%	**29,651**	5.0%	**527,083**	89.0%
2016-17	**604,575**	**12,101**	2.0%	**21,376**	3.5%	**30,009**	5.0%	**541,089**	89.5%

Notes and sources: Author's calculations using National Pupil Database 2002-03 to 2016-17. Original sample refers to all unique pupils in the data. School exclusions refers to pupils in special, independent or other non-maintained schools; missing pupil characteristics refers to pupils in Maintained schools, Academies or Free schools with missing pupil characteristics; local authority exclusions refers to pupils in Inner London or Isles of Scilly with non-missing characteristics

As per the analysis plan, I also exclude inner London and the Isles of Scilly from the analysis as the former is subject to very different time trends (Blanden et al. (2015)) and the latter is extremely small. These exclusions reduce the number of local authorities in the sample from 152 to 136. I also exclude local authorities where the average outcomes of interest are more than 0.15 standard deviations different to those in our treatment group of South and West Yorkshire. This reduces the number of local authorities in the panel further to 87 (9 in South and West Yorkshire and 78 in our control group of donor pool). These local authorities are excluded on the grounds that they are unlikely to represent good controls. A threshold of 0.15 was used as it delivered the lowest prediction error for English as compared with other thresholds. I show that the pattern of results is largely unchanged under different thresholds in the robustness checks.

4.4.2 Construction of Synthetic Controls

Table 4.2 compares the characteristics of primary schools in S&W Yorkshire with the donor pool (all other local authorities in England, excluding inner London and the Isles of Scilly) averaged over the whole pre-treatment phase (the pupil:teacher assistant ratio (pupil:TA ratio) is averaged over 2010-11 to 2014-15 and the proportion of primary schools that are Academies is averaged over 2010-11 to 2014-15).

A comparison between columns (1) and (2) shows how South and West Yorkshire differs from the rest of England (these are the same for both outcomes). It is more deprived, on average, with 19.4% of pupils eligible for free school meals (FSM) in South and West Yorkshire as compared with 15.3% in the rest of England. The ethnic mix is also slightly different from the rest of England, with a slightly smaller share of White-British ethnicity, a smaller share of Black ethnicity and larger share of Asian ethnicity. There is also a larger share of pupils with English as an Additional Language (EAL) in South and West Yorkshire (15.6%) as compared with the rest of England (11.9%). The proportion of pupils with a statement of special educational needs (SEN) is similar across groups. However, we see a slightly smaller pupil:TA ratio in South and West Yorkshire, suggesting a greater reliance on teaching assistants as compared with the rest of England. However, in both cases there is approximately one TA per class of just under 30. A slightly smaller share of primary schools, on average, were Academies in South and West Yorkshire (7.0%) than in the rest of England (8.1%).

Figures 4.1(a) and (b) show the evolution of standardised maths and English results over time in S&W Yorkshire (dashed line) and the donor pool (grey line). As can be seen, outcomes are lower on average in S&W Yorkshire as compared with the rest of England in the donor pool. This gap also increases slightly over time for both maths and English. For maths, the gap goes from about 0.08 to 0.11 standard deviations between 2003 and 2015. For English, the increase in the gap is even faster, going from 0.12 to 0.17 standard deviations between 2003 and 2015. These different time trends represent an important reason why simply matching on pre-treatment outcomes just before the point of treatment could be misleading, as outcomes for S&W Yorkshire were on a different trajectory.

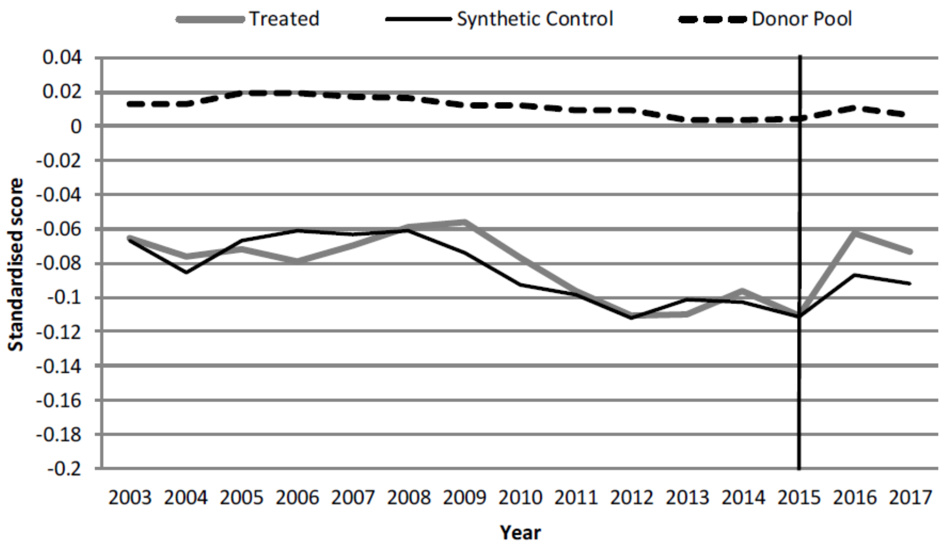
Table 4.2: Summary statistics of donor pool, treatment and synthetic controls

Outcome	**(1) Treatment**	**(2) Donor Pool**	**(3) Treatment - Donor Pool**	**(4) Maths Synthetic Control**	**(5) Treatment - Maths SC**	**(6) English Synthetic Control**	**(7) Treatment - English SC**
% Eligible for FSM (avg)	19.4	15.3	4.2	19.5	-0.1	25.1	-5.7
% White-British (avg)	77.5	79.1	-1.6	77.4	0.1	77.4	0.0
% Black (avg)	2.0	3.5	-1.5	1.6	0.4	1.6	0.4
% Asian (avg)	13.2	6.7	6.5	12.6	0.6	12.6	0.6
% with English as Additional Language (avg)	15.6	11.9	3.7	15.6	0.0	15.7	-0.1
% with Statement of SEN (avg)	2.0	1.9	0.1	2.0	0.0	2.3	-0.3
Teaching Assistant: Pupil Ratio (2011 to 2015)	28.1	33.3	-5.2	29.3	-1.2	28.7	-0.6
% Academy (2011 to 2015)	7.0	8.1	-1.0	7.4	-0.3	7.8	-0.7
Root Mean Squared Prediction Error (Maths, SDs)			0.096		0.010		-
Root Mean Squared Prediction Error (English SDs)			0.126		-		0.005

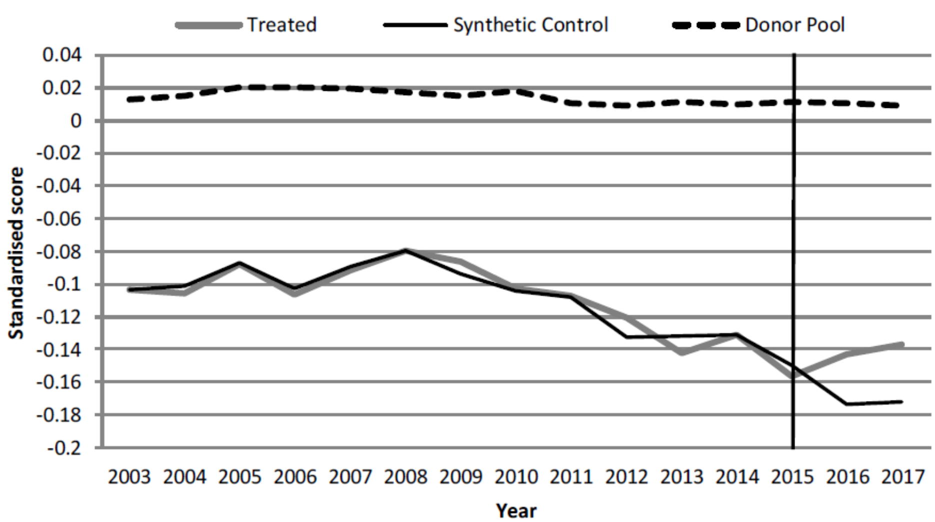
Notes and sources: Author's calculations using National Pupil Database (2002-03 to 2014-15) and School Workforce Census Statistics (2011 to 2015). Avg refers to the average of pupil characteristics in pre-treatment years 2002-03 to 2014-15. Treatment refers to all nine local authorities in South and West Yorkshire. Donor pool excludes inner London, Isles of Scilly and local authorities where average KS2 English scores are more than 0.15 standard deviations different from treatment in pre-treatment phase.

Figure 4.1: Key Stage 2 Test Scores for Treated, Synthetic Controls and Donor Pool Groups

a) Mathematics



b) English



Notes and Source: Author's calculations using National Pupil Database (2002-03 to 2014-15) and School Workforce Census Statistics (2011 to 2015). Year refers to academic year starting each September. Vertical line indicates when treatment began in South and West Yorkshire

In constructing the synthetic controls, a key consideration is which variables to include in the V matrix of predictor variables used to achieve balance. Abadie and Gardeazabal (2003) and Abadie et al. (2010) describe this as a partially subjective judgment informed by the likely factors driving potential outcomes. They discuss choosing the number of predictor values in the V matrix as a trade-off between obtaining convincing balance, efficiency and transparency. Adding more pre-treatment outcomes can improve balance, but it can also lead to a lack of transparency and over-fitting with sparse weights across a large number of local authorities.

To create the synthetic controls, I include average pupil and school characteristics over the pre-treatment phase as listed in Table 4.2, which seem likely to be major determinants of the difference in potential outcomes over time. I include English and maths outcomes for years 2002-03, 2007-08, 2012-13, 2014-15 to account for further unobservable determinants of potential outcomes, with the years chosen to represent key turning points in trends over time. I then seek to find the set of weights W associated with this V matrix of key predictor variables that delivers the lowest prediction error across lagged outcomes across all pre-treatment years for maths and English. Robustness checks examine the effects of changing the variables included in the V matrix.

Columns (4) and (6) show the average pupil characteristics of the resultant synthetic control groups for maths and English, respectively. For maths, all the pupil and school characteristics are within 1 percentage point of each other (often less) and pupil:TA ratio is also very similar across the two groups. The differences shown in column (5) are all small in value. For English, some slightly larger differences remain with the difference in the percentage of pupils eligible for FSM around 6 percentage points. Such differences are not necessarily problematic, however, as the aim of the synthetic control method is to minimise differences in pre-treatment outcomes rather than predictor variables.

The bottom of Table 4.2 shows that the synthetic control groups achieve a major reduction in the root mean squared prediction error (RMSPE) as compared with the donor pool. For maths, the RMSPE reduces from 0.096 to 0.01 standard deviations for maths and from 0.126 to 0.005 standard deviations for English. Figure 4.1 further shows that the synthetic control groups provide a good visual match for the pre-treatment trends over time for S&W Yorkshire in the pre-treatment phase.

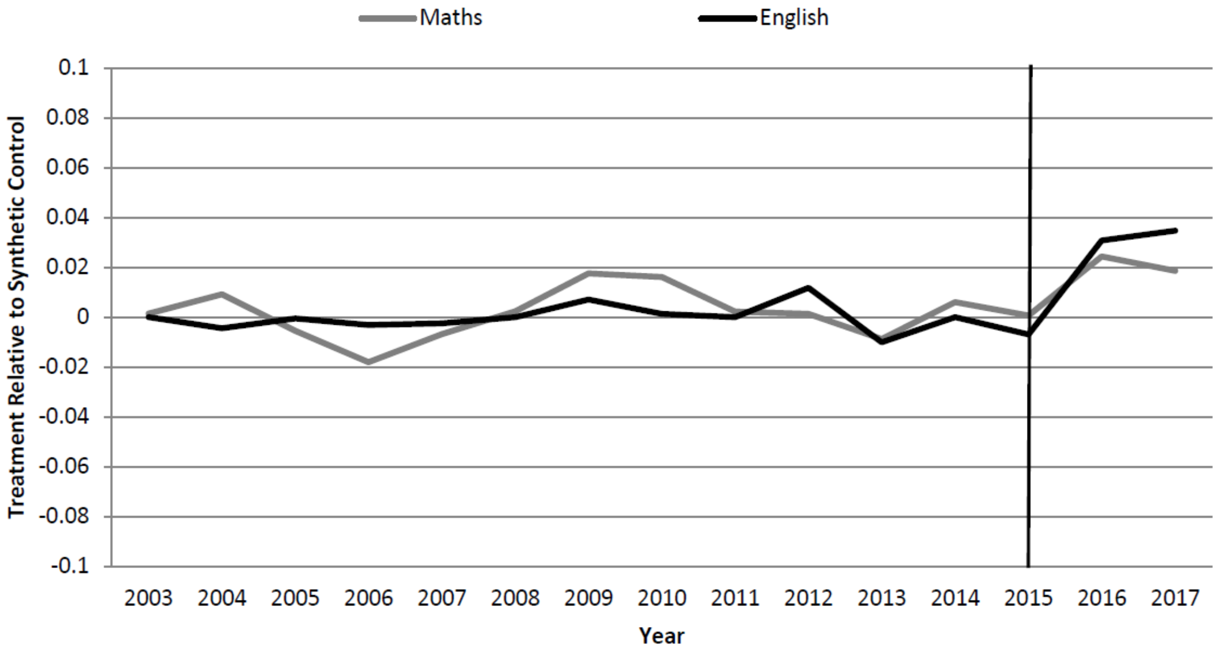
Whilst the main specification uses the same set of predictors (V) in all cases, different weights (W) are used for the construction of synthetic controls for different outcomes. The process driving potential outcomes is likely to be different across different outcomes. For example, pre-treatment maths outcomes will likely be better at predicting pre-treatment maths outcomes than English outcomes, and vice versa. This approach of using different weights for different outcomes is common in the literature (Kreif et al. (2016); Hinrichs (2012)).

Figure C1 shows the difference between S&W Yorkshire and the synthetic control for maths and English outcomes, and how this changes if the weights for the alternate outcomes are used to construct the synthetic controls instead (e.g. English scores with weights calculated for the maths synthetic controls). As can be seen, using the alternate weights clearly reduces the fit between treatment and synthetic controls during the pre-treatment phase. This demonstrates the value of using outcome–specific weights.

4.4.3 Main Results

The difference between S&W Yorkshire and the synthetic control groups in the treatment phase represents the main estimates of the impact of the TA campaign in S&W Yorkshire. Figure 4.2 shows an uptick in both maths and English in S&W Yorkshire relative to the synthetic control group in the post treatment phase in 2015-16 and 2016-17. For maths, the post-treatment gap is around 0.024 standard deviations in 2016 and 0.019 standard deviations in 2017. These differences are both higher than any seen over the pre-treatment phase. For English, we see more evidence of a growing impact of the treatment. In 2016, the implied impact estimate is 0.031 standard deviations, growing to over 0.035 standard deviations in 2017. The latter is about three times the size of any other absolute difference seen over the pre-treatment phase.

Figure 4.2: Difference between treatment and synthetic controls over time, Key Stage 2 standardised scores



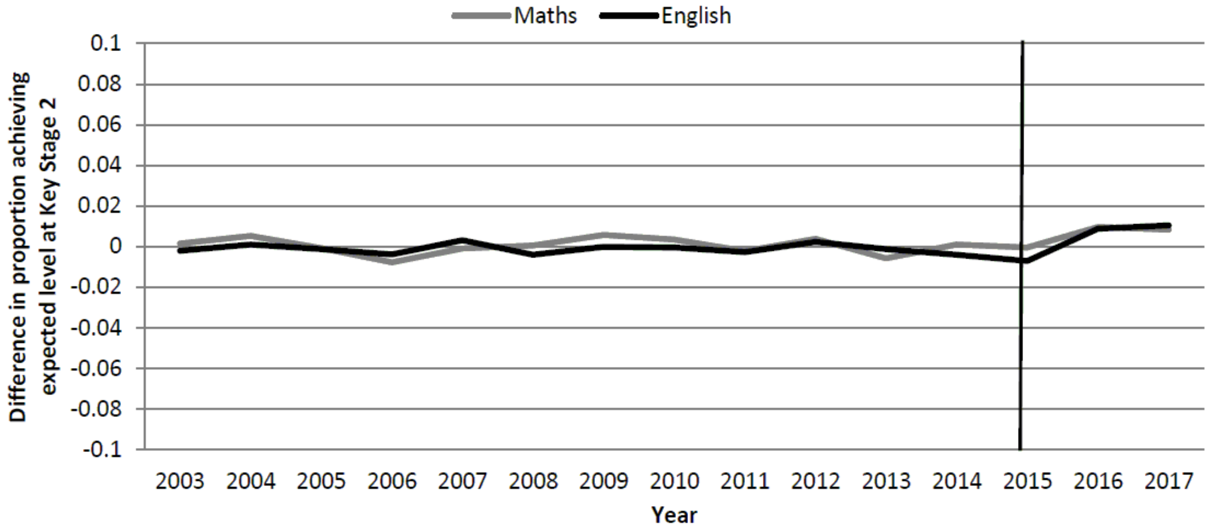
Author's calculations using National Pupil Database (2002-03 to 2014-15) and School Workforce Census Statistics (2011 to 2015). Year refers to academic year starting each September. Vertical line indicates when treatment began in South and West Yorkshire

Figure 4.3 shows the difference in the proportion of pupils achieving the expected standard in English and maths between S&W Yorkshire and a synthetic control group. This gives a very similar pattern to the main standardised outcomes, with good balance over the pre-treatment phase and a small uptick in S&W Yorkshire as compared with the synthetic control group following the treatment, though there is more similarity across maths and English for these threshold outcomes.

Table 4.3 details the main treatment estimates in 2016-17 for maths and English, and compares these against a range of alternative specifications for defining the synthetic control group. This also shows the ratio of the impact estimate relative to the RMSPE over the pre-treatment phase. and p-values based on simulations of placebo treatment estimates (using 250 simulations). Column (1) presents the main impact estimates for maths (upper panel) and English (lower panel). For English, I estimate an impact of 0.035 standard deviations, which is over 6 times larger than the RMSPE over the pre-treatment phase. The implied p-value from the placebo treatment estimates is 0.1, which puts the estimate on the boundary of statistical significance at the 10% level.

Figure 4.4 explores this in more detail by showing the full cumulative distribution of placebo treatment estimates (relative to the RMSPE) and the associated RMSPE. This makes clear that the estimate for S&W Yorkshire is unusual relative to other placebo estimates and that the distribution of placebo treatment estimates is relatively well-behaved. Unsurprisingly, placebo treatment estimates based on a high RMSPE tend to generate a relatively low ratio of the estimate to the RMSPE.

Figure 4.3: Difference between treatment and synthetic controls over time, % achieving expected level



Author's calculations using National Pupil Database (2002-03 to 2014-15) and School Workforce Census Statistics (2011 to 2015). Year refers to academic year starting each September. Vertical line indicates when treatment began in South and West Yorkshire

For maths, the estimated treatment effect of 0.019 standard deviations is only about 2 times larger than the RMSPE seen over the pre-treatment phase. This is not unusual relative to the pre-treatment phase, with an implied p-value of 0.35. This estimate remains statistically insignificant and small across all specifications listed in Table 4.3. I therefore focus on the results for English for the rest of this section.

Columns (2)-(4) vary the donor pool. Column (2) allows all LAs within the donor pool to form part of the synthetic controls. For English, this increases the estimated treatment effect to 0.05 standard deviations and generates a p-value of 0.02 and thus would be statistically significant at the 5% level. Column (3) narrows the donor pool to those with pre-treatment outcomes within 0.25 standard deviations of those for S&W Yorkshire (the specification delivering the lowest prediction errors for maths as opposed to English). This generates a very similar estimated treatment effect to the preferred estimate in column (1), though with a p-value of 0.15. Column (4) use the preferred donor pool, but with LAs neighbouring S&W Yorkshire removed in order to reduce any potential spillover effects. The magnitude of the treatment effect is slightly larger than the main estimate in column (1), with a p-value of 0.05.

Columns (5)-(7) of Table 4.3 maintain the preferred donor pool, but vary the set of predictor variables used in the V matrix. Column (5) includes more lagged outcomes within the V matrix, including all lagged outcomes for alternate years between 2003 and 2015. In this case, the estimated impact reduces to 0.015 and the result is clearly not statistically significant. However, in this specification, the weights become highly spread out across many local authorities and the RMSPE in some simulations can become implausibly low (less than 0.001 as compared with 0.005 in the main specification). Adding all lagged outcomes (not reported here) exacerbates this problem by producing near perfect fit in synthetic controls for many simulations. This demonstrates the trade-off described by Abadie et al. (2010) between balance, efficiency and transparency in selecting how many predictors to include in the V matrix.

Column (6) reduces the number of lagged outcomes included and column (7) excludes the school characteristics. The estimates are similar to the main specification, with the p-value varying from 0.05 to 0.16.

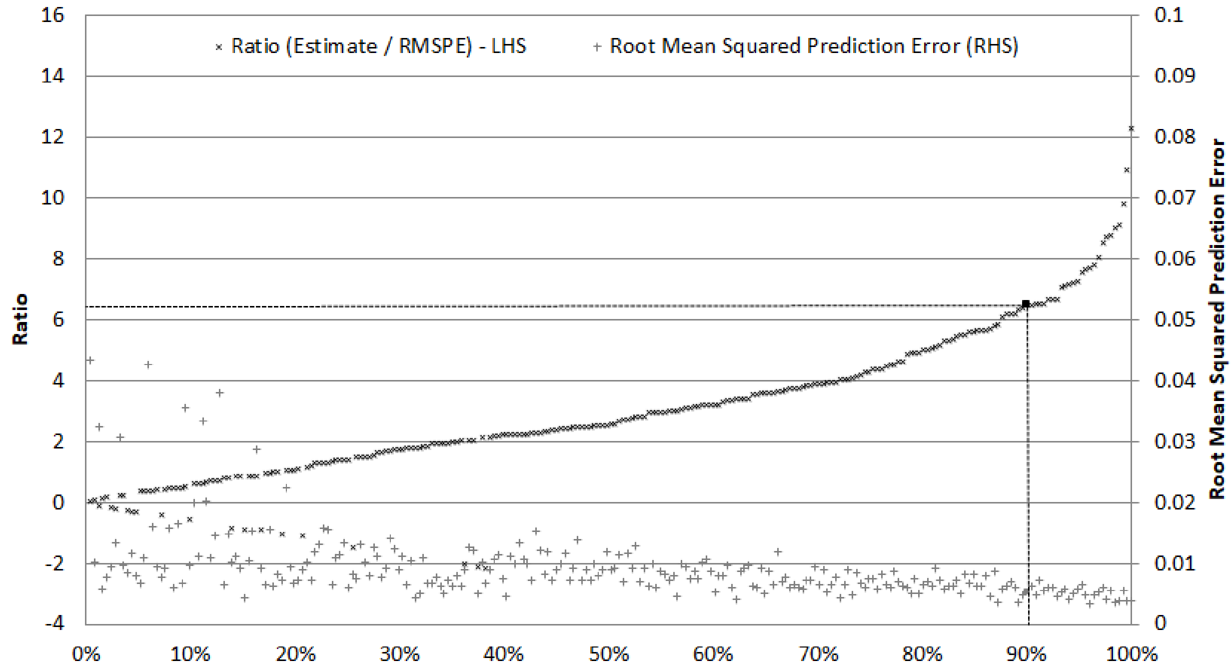
Finally, in column (8) I adopt the Acemoglu et al. (2016) approach by estimating synthetic controls for each LA in S&W Yorkshire and then aggregate the results. This produces a very different estimate, with near zero treatment effects. However, this estimator performs poorly, with a RMSPE of 0.03 as compared with less than 0.01 in other approaches (all following the Kreif et al. (2016) approach of aggregating the multiple treatment units first). Looking at the results for the nine individual local authorities (see Appendix Table C1), the treatment estimates range from about -0.13 to 0.08. The RMSPE is also relatively high across all individual local authorities, ranging from 0.02 to 0.05 standard deviations. This strongly suggests that synthetic control methods are inappropriate for finding a good match for individual local authorities. Given the poor performance of this estimator, I do not generate p-values for this estimate.

Table 4.3: Estimated treatment effects for range of synthetic control approaches

Outcome	**(1) Main**	**(2) All LAs**	**(3) Larger sample**	**(4) Excluding spillovers**	**(5) More Lagged Outcomes**	**(6) Fewer Lagged Outcomes**	**(7) Excluding school data**	**(8) Aggregate treatment effect**
A) Maths								
Estimated Treatment Effect	0.0187	0.0228	0.0154	-0.0136	-0.0039	0.0182	0.0052	-0.0188
RMSPE	0.0096	0.0072	0.0055	0.0102	0.0116	0.0067	0.0103	0.0305
Ratio	1.950	3.183	2.794	-1.330	-0.334	2.723	0.505	-0.617
P-value	0.3480	0.0960	0.2240	0.5360	0.8960	0.1320	0.8000	n/a
B) English								
Estimated Treatment Effect	0.0348	0.0511	0.0332	0.0412	0.0147	0.0451	0.0423	-0.0146
RMSPE	0.0054	0.0068	0.0072	0.0069	0.0052	0.0093	0.0104	0.0319
Ratio	6.478	7.526	4.596	6.014	2.802	4.849	4.055	-2.6787
P-value	0.1000	0.0200	0.1480	0.0520	0.6960	0.0520	0.1640	n/a
Donor Pool of LAs	$<$ 0.15 SDs	All	$<$ 0.25 SDs	Exclude neighbours	$<$ 0.15 SDs	$<$ 0.15 SDs	$<$ 0.15 SDs	$<$ 0.15 SDs
Predictor Variables	Main	Main	Main	Main	Main +	Main -	Main minus school chars	Main, individual LA SCs

Notes and sources: Author's calculations using National Pupil Database (2002-03 to 2014-15) and School Workforce Census Statistics (2011 to 2015). Main predictor variables include all pupil characteristics listed in Table 4.2 averaged over the pre-treatment phase (% eligible for FSM, % White-British, % Black, % Asian, % with EAL, % with SEN), school characteristics averaged over 2011 to 2015 (TA Ratio, % Academies) and lagged outcomes for 2003, 2008, 2013 and 2015. Column (2) includes all LAs in the donor pool. Column (3) includes all LAs with average pre-treatment outcomes within 0.25 standard deviations of those for South and West Yorkshire. Column (4) drops any local authorities bordering on South and West Yorkshire. Column (5) adds lagged outcomes for alternate years between 2003 and 2015. Column (6) drops lagged outcomes for 2008 and 2013 as compared with the main specification. Column (7) removes school characteristics as predictor variables. Column (8) creates synthetic controls for each LA within South and West Yorkshire and weights these according to population shares. RMSPE refers to the Root Mean Squared Prediction Error. The ratio represents the ratio between the main estimate and RMSPE over the pre-treatment phase. The p-values are calculated by creating 250 simulated versions of South and West Yorkshire from the donor pool and calculating placebo treatment effects. P-values based on distribution of ratio.

Figure 4.4: Cumulative distribution of simulations for ratio between estimate and RMSPE, KS2 English Scores



Notes and sources: Author's calculations using National Pupil Database (2002-03 to 2014-15) and School Workforce Census Statistics (2011 to 2015). Year refers to academic year starting each September. Dashed line indicates preferred actual estimate. 250 simulations in total

Appendix Table C2 performs further robustness checks by showing the estimated treatment effect if each local authority receiving a positive weight in the main analysis is excluded (and the results re-estimated). For maths, the treatment estimates vary between zero and 0.03 standard deviations, suggesting some weak dependence on exactly which local authorities are included in the donor pool. For English, there is a tighter range of estimates between about 0.02 and 0.04 standard deviations. The number of local authorities receiving positive weight English is also smaller, with all in relative close proximity to S&W Yorkshire.

The main results suggest a small uptick in English results in primary schools in S&W Yorkshire as compared with the synthetic control group, with no evidence of an impact on maths. The preferred estimate suggests an impact of 0.035 standard deviations, with a p-value of 0.1, so that the results are on the boundary of statistical significance. These results are generally robust to changing the V specification for predictor variables, the size of the donor pool, exclusion of neighbouring local authorities and are not dependent on single local authorities being part of the synthetic control group. The results are not always statistically significant, however. In most cases, the p-value varies from 0.02 to 0.16. When including many years of lagged outcomes, the estimate falls and is no longer on the boundaries of statistical significance, but the estimates and inference procedure seem to perform poorly when many years of lagged outcomes are added.

4.4.4 Comparison across methods

This section compares the preferred synthetic control estimates with those produced by other non-experimental estimators (OLS, matching, difference-in-differences) and relates these to the identification assumptions underlying the different methods presented in the previous section.

For maths, the results remain small and statistically insignificant across all methods, confirming a lack of evidence for any impact on maths scores. For English, the results are smaller when using both OLS and propensity score matching, and are statistically insignificant. In the case of difference-in-differences, the estimates turn negative, though again are statistically insignificant.

This illustrates two main advantages of the synthetic control approach relating to the identifying assumptions and magnitude of standard errors. First, the main difference in identifying assumptions between the synthetic control method and all others shown in Table 4.4 is the relaxation of the impact of unobservables. OLS and matching both assume no unobservable differences between treatment and control groups. The smaller estimated impacts under OLS and matching suggest a negative bias of unobservables. Difference-in-differences controls for time constant differences in unobservables between treatment and control groups. The main identifying assumption for difference-in-differences is parallel trends in unobservables, which is relaxed under synthetic control analysis. The difference between the synthetic control and difference-in-differences estimates suggests the parallel trends assumption is likely to be violated in this case, as was suggested by the time trends in raw outcomes. The ability to relax the parallel trends assumptions is a significant advantage of synthetic control analysis.

Second, the implied standard errors for synthetic control analysis are slightly smaller than those generated by other non-experimental methods (about half the value of those for DiD). Assuming the identification assumptions are convincing, synthetic control analysis is thus relatively powerful. In this case, one is able to detect impact estimates of about 0.035 standard deviations or higher (based on a 10% threshold) or 0.04 standard deviations (based on a 5% threshold).

Table 4.4: Comparison with OLS, matching and Difference-in-Differences

Outcome	**(1) Main Specification**	**(2) OLS**	**(3) Kernel PS Matching**	**(3) Difference-in-Differences**
A) Maths				
Estimated Treatment Effect	0.0187	0.0152	0.0049	0.0186
Standard Error	0.0174	0.0204	0.0204	0.0331
P-value	0.3480	0.4603	0.8101	0.5751
B) English				
Estimated Treatment Effect	0.0348	0.0254	0.0054	-0.0123
Standard Error	0.0136	0.0211	0.0240	0.0292
P-value	0.1000	0.2321	0.8233	0.6738
Donor Pool of LAs	$<$ 0.25 SDs	$<$ 0.25 SDs	$<$ 0.25 SDs	$<$ 0.25 SDs
Control Variables	Main	Main	Main	Main

Notes and sources: Author's calculations using National Pupil Database (2002-03 to 2014-15) and School Workforce Census Statistics (2011 to 2015). OLS and Kernel Matching includes the same controls as used in the main synthetic control specification. Difference-in-difference specification includes local authority and year fixed effects, but excludes school characteristics given these are only measured for a subset of years.

4.4.5 Sub-group analysis

To better understand the potential mechanisms driving the overall effects, I undertake a range of sub-group and on-treatment analysis. First, I repeat the analysis across various sub-groups for English in Figure 4.5 and maths in Figure 4.6. This includes an examination of outcomes by deprivation (whether pupils were eligible for free school meals), amongst children who speak English as an additional language, and pupils with and without special educational needs. Given that Teaching Assistants spend most of their time helping pupils with specific educational needs, one might expect a larger impact of the advocacy campaign on more challenging groups of pupils (those from more deprived backgrounds, who don't speak English as a first language and those with special educational needs). Instead, the results suggest that most of the impacts on English scores are driven by pupils from less deprived backgrounds and with no identified special educational needs. There is, however, a clear impact on pupils with English as an additional language.

Whilst counter-intuitive, these results are in line with reported changes in practice and the recommendations being promoted by advocacy providers. In particular, one important recommendation was to avoid exclusively using Teaching Assistants to support pupils with special or additional educational needs, instead using them as part of whole class teaching to support all pupils. A separate qualitative evaluation of the campaign in S&W Yorkshire showed that schools followed this advice, with fewer schools reporting that Teaching Assistants time was focused on pupils from deprived backgrounds or with special educational needs[13]. The faster improvements among less deprived pupils and those without any special educational are in line with how usage of Teaching Assistants changed across schools as a result of the campaign. These results are in line with a higher impact of TAs on pupils from higher socio-economic status backgrounds as reported in Penney (2018).

[13]https://educationendowmentfoundation.org.uk/scaling-up-evidence/campaigns/making-best-use-of-teaching-assistants/evaluation/

Figure 4.5: Sub-group difference between treatment and synthetic controls over time, English

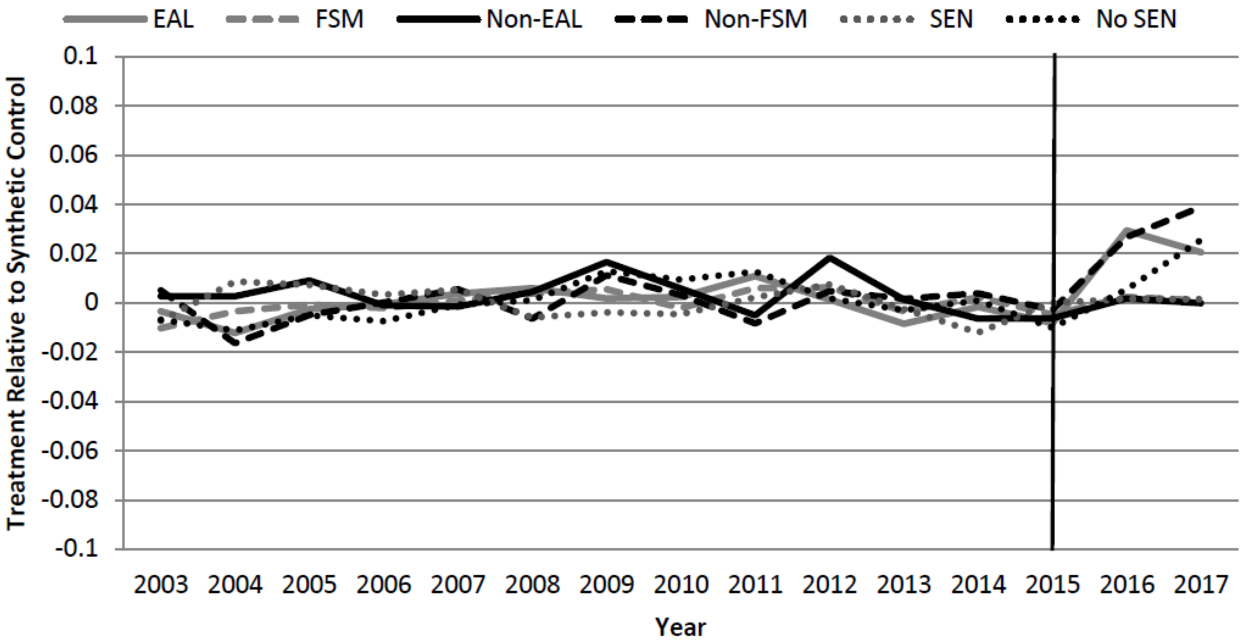
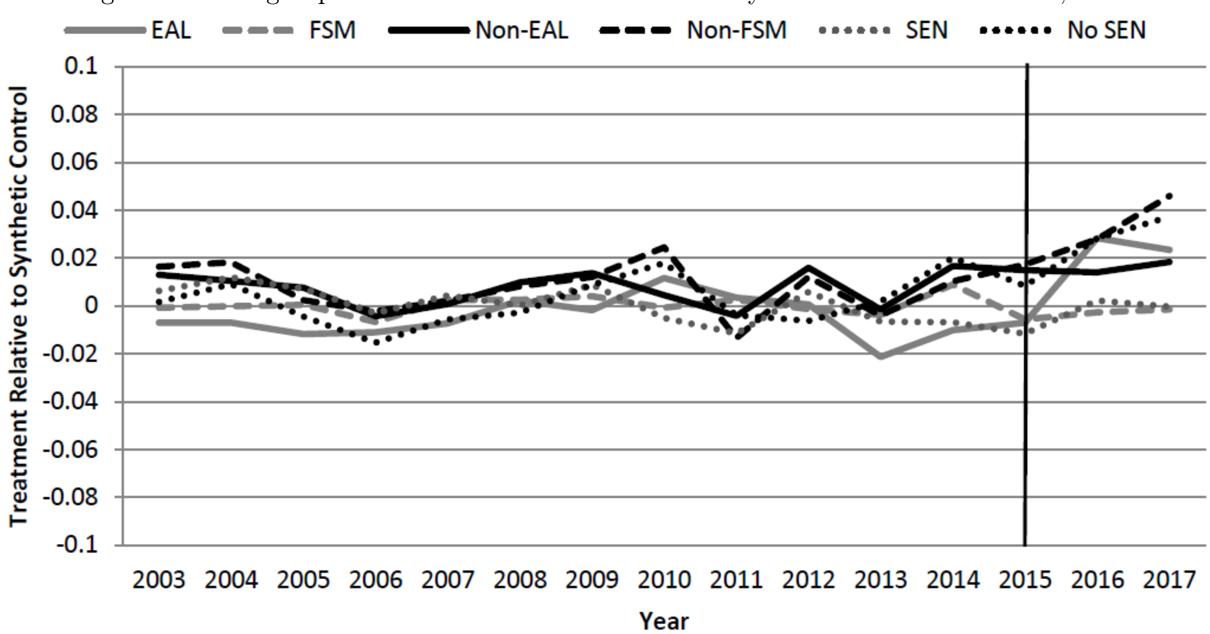


Figure 4.6: Sub-group difference between treatment and synthetic controls over time, Maths



Notes and sources: Author's calculations using National Pupil Database (2002-03 to 2014-15) and School Workforce Census Statistics (2011 to 2015). Year refers to academic year starting each September. Vertical line indicates when treatment began in South and West Yorkshire. "EAL" refers to English as an Additional Language. "FSM" refers to pupils eligible for Free School Meals

Two concrete ways in which schools in S&W Yorkshire could show engagement with the campaign was to sign up with one of the advocacy providers or participate in a trial of a TA-led intervention, which were heavily targeted at the area. About 480 or 42% of all primary schools in S&W Yorkshire signed up with one of the advocacy providers and around 20% of primary schools received additional TA-interventions as part of the TA campaign[14]. These two groups of schools are not mutually exclusive; schools could sign up with an advocacy provider and/or participate in a trial. Indeed, schools who signed up for advocacy provision were often targeted by advocacy providers for participation in trials. Table 4.5 estimates the difference in age 11 outcomes between schools that signed up to an advocacy provider or was involved with a trial as compared with other schools in S&W Yorkshire. I estimate these effects using OLS regression with a dummy variables for signing up to an advocacy provider and participating in a trial, and control for additional covariates (the same as those used in our main analysis). I do this for primary outcomes in 2017, and for the period just before the treatment began in 2015. This analysis is not necessarily causal as there could still be unobservable factors driving both outcomes and participation in trials.

Table 4.5: On-treatment analysis within South & West Yorkshire

Outcome	**(1) 2015**	**(2) 2017**
A) Maths		
Registered for Advocacy	0.0247	0.0178
Standard Error	0.0221	0.0244
Participation in trial	-0.0356	-0.0337
Standard Error	0.0261	0.0297
B) English		
Registered for Advocacy	0.0199	0.0120
Standard Error	0.0219	0.0220
Participation in trial	-0.0191	-0.0301
Standard Error	0.0259	0.0261
Control Variables	Main	Main

Notes and sources: Author's calculations using National Pupil Database (2002-03 to 2014-15) and School Workforce Census Statistics (2011 to 2015). Effects are estimated using OLS regression with dummy variables for each group of schools and the same set of covariates used as predictors in our main analysis. Only includes schools in South and West Yorkshire.

These results provide no clear empirical evidence that schools participating in advocacy or additional TA-led interventions had higher Key Stage 2 scores than schools that did not. This was the case in 2015 before advocacy provision began and two years later in 2017. These results are not necessarily causal and are also consistent with the small positive effects we estimate in our primary analysis. However, large effects of either advocacy or additional trials would only be consistent with the results in Table 4.5 if the latter are subject to large negative biases (e.g. schools that signed up to advocacy or trials were likely to have much worse outcomes, even after controlling for covariates).

[14]This data was collected by Sheffield Hallam University for the purpose of conducting a mixed methods process evaluation of the TA campaign and they expressed potential for error in the reporting of participation in TA-led trials as schools weren't always clear whether this was definitely the case or not.

4.4.6 Interpretation and mechanisms

To summarise, the synthetic control approach is able to achieve good balance in pre-treatment outcomes in both maths and English, as well as pupil and school characteristics. There is then a post-treatment difference in English scores of about 0.035 standard deviations. This is large relative to pre-treatment differences in outcomes between the treatment and synthetic control and relative to a set of placebo treatments. For maths, there is little evidence of any impact. These figures represent estimates of implementing an advocacy campaign right across S&W Yorkshire to improve the ways in which Teaching Assistants are used. There are then a number of ways to interpret these results and the mechanisms driving them.

First, it could reflect a genuine causal impact of the campaign on English scores. If this is the case, a 0.035 impact need not necessarily be seen as small, given the large number of schools and pupils involved. The analysis covers about 43,000 taking KS2 tests in S&W Yorkshire, or about 350,000 pupils across all year groups. An impact of 0.035 would be sufficient to reverse almost half the 0.08 standard deviation decline in English results across S&W Yorkshire relative to the rest of England between 2008 and 2015. Furthermore, the control group would have been positively affected by the national campaign and guidance materials. Over 40% of head teachers reported reading the guidance document within six months of publication in 2015. This could have improved overall usage of teaching assistants in the control group over time. Whist this would have also affected S&W Yorkshire, it does make the search for a differential impact in South and West Yorkshire slightly more demanding, and would need to reflect more than just becoming aware of the guidance.

This said, the on-treatment analysis suggests that schools who took up the offer of advocacy had similar outcomes to those of other schools in S&W Yorkshire. Although one can't give this a causal interpretation, it would caution against interpreting the main results as potentially reflecting a large causal effect of advocacy. It is also not clear that the main results are statistically significant. The p-value for the main estimates is 0.1, though this varies from 0.02 to 0.15, depending on which local authorities form the set of potential control units, and from 0.05 to 0.70, depending on which observable characteristics are used as predictor variables to obtain the weights.

There is, however, evidence that improvements in English scores were fastest among less deprived pupils and those without special educational needs. This is in line with the recommendations of the campaign to use Teaching Assistants more to support whole class teaching and actual changes in practice, and previous evidence (Penney (2018)).

Second, a key issue in fully understanding the mechanisms is why English scores improved and maths scores did not. Existing evidence on the impact of teaching assistants has shown a larger negative effect on English scores. This could suggest a greater scope for improvement in English (Blatchford et al. (2011)). Empirical evidence also suggests that TA-led interventions tend to have more positive effects than for maths (Alborz et al. (2009); Andersen et al. (2020)). Indeed, six of the seven TA-led interventions recommended by the EEF in this campaign (based on experimental evidence) relate to language and communication as opposed to numeracy. This may suggest than English and literacy outcomes are more malleable to the roles played by teaching assistants than maths and numeracy outcomes, with more research needed as to why this might be the case.

Third, it may be that the treatment effect is driven by the additional RCT interventions in S&W Yorkshire (these additional interventions are distinct from the recommended practices by EEF). The evidence for this interpretation is relatively weak, however. A qualitative evaluation showed that around 20% of primary schools received additional TA-interventions as part of the TA campaign. Although one cannot establish a reliable benchmark for what was happening in the rest of England, it seems reasonable to assume the figure was a lot lower outside of S&W Yorkshire. However, three of the four interventions promoted by advocacy providers related to numeracy, where there is little impact. Only one focuses on literacy (Switch-on Reading and Writing) and the published EEF evaluation finds little evidence that the trialed version of the treatment had a positive impact on literacy outcomes (NatCen, 2017). Furthermore, there is little evidence that schools in S&W Yorkshire participating in these trials showed high test scores than schools that did not.

Fourth, it could be that South and West Yorkshire was unusual in its response to the new Key Stage 2 tests and curriculum, in being an area that was better able to improve its outcomes. By definition, this is hard to test. Any bias could be either positive or negative, large or small. However, the synthetic controls are explicitly created in order to mimic the effects of unobservable shocks.

This paper's view is that the improvement in English scores is likely to be a genuine causal impact of the advocacy campaign, though there are some clear doubts on the statistical significance of the results. It is also more likely to be driven by changes in how Teaching Assistants were deployed within the classroom, rather than greater use of structured interventions. The larger impacts on English scores are also consistent with existing evidence suggesting greater scope for improvement in English scores and that language outcomes might be more malleable to the roles played by teaching assistants. However, there are some important caveats to this conclusion, given the lack of difference within S&W Yorkshire by whether or not schools participated in the training and the impact of changed assessments.

4.5 Conclusions

This paper used synthetic control methods to estimate the impact of an area-wide campaign to improve the ways in which teaching assistants are used. The conclusions relate to both the impact of the area-wide campaign and the potential to use synthetic control methods in other education evaluations.

I find that the area-wide campaign improved English test scores across S&W Yorkshire by about 0.03-0.04 standard deviations, with no evidence of an impact on maths scores. This estimate for English is likely to be statistically significant, though this is far from guaranteed. For the main estimates, the implied p-value from permutation tests is 0.1, though this can vary from 0.02 to 0.15, depending on which local authorities form the set of potential control units, and from 0.05 to 0.70, depending on which observable characteristics are used as predictor variables to obtain the weights.

The larger impact on English scores seem consistent with existing evidence suggesting greater scope for improvement here and that language outcomes might be more malleable to the actions and roles played by teaching assistants. If there is a positive impact on English scores, then the mechanisms driving this result are unlikely to include greater participation in trials of TA-led interventions, which were more widely offered to schools in S&W Yorkshire during the course of the campaign. There is no evidence of any difference in pupil outcomes in schools in S&W Yorkshire that did and did not report participating in these trials. Unfortunately, there is also little evidence of any difference in pupil outcomes among schools in S&W Yorkshire that did and did not report participating in the additional training and advocacy. However, these comparisons are not fully robust and could be driven by unobservable determinants of participation.

What is clear from analysis of sub-groups is that any impact on age 11 test scores is largely driven by improvements among 'less challenging pupils' (pupils who aren't poor enough to be eligible for Free School Meals and those who don't have any reported special educational needs). This may seem counter-intuitive, given that these pupils are less likely to interact with teaching assistants. However, it is in line with qualitative analysis of the impact of the advocacy campaign in S&W Yorkshire and the recommended approaches. The guidance recommended using teaching assistants more to support the whole class, rather than exclusively focusing support on specific under-achieving groups. Schools participating in the campaign did indeed report using teaching assistants more to support the whole class and became less likely to focus their time on low-attaining pupils, those with special educational needs and pupils eligible for Free School Meals. The greater impacts on pupils less likely to interact with Teaching Assistants are therefore fully in line with changes in practice that led to more support for the whole class.

These results have also shown that synthetic control methods are a valuable methodological tool in undertaking non-experimental education evaluations in England. They are relatively easy to apply given the large amounts of administrative education data available in England over a long-period of time and achieve a good match in pre-treatment outcome trends. The results are also different to those implied by other non-experimental estimators, such as matching or DiD. This shows the value of being able to relax assumptions on balance in unobservables and parallel trends. However, the results also suggest that synthetic control methods must must be applied over relatively large areas. The approach largely failed to achieve good matches when applied to individual local authorities. Synthetic control methods should be used more and can expand our understanding of the impact of area-wide educational interventions.

Chapter 5

Bibliography

Abadie, A., A. Diamond, and J. Hainmueller (2010): "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Programs," *Journal of the American Statistical Association*, vol. 105(490), 493–505.

Abadie, A., A. Diamond, and J. Hainmueller (2015): "Comparative Politics and the Synthetic Control Method," *American Journal of Political Science*, vol. 59(2), 495–510.

Abadie, A. and J. Gardeazabal (2003): "The Economic Costs of Conflict: A Case Study of the Basque Country," *American Economic Review*, vol. 93(1), 113–132.

Acemoglu, D., S. Johnson, A. Kermani, J. Kwak, and T. Mitton (2016): "The value of connections in turbulent times: Evidence from the United States," *Journal of Financial Economics*, vol. 121(2), 368–391.

Alborz, A., D. Pearson, P. Farrell, and A. Howes (2009): "The impact of adult support staff on pupils and mainstream schools," Technical report, Department for Children Schools and Families, London.

Allen, R., C. Belfield, E. Greaves, C. Sharp, and M. Walker (2014): "The Costs and Benefits of Different Initial Teacher Training Routes," Research Report R100, Institute for Fiscal Studies.

Allen, R., C. Belfield, E. Greaves, C. Sharp, and M. Walker (2016): "The longer-term costs and benefits of different initial teacher training routes," Research Report R118, Institute for Fiscal Studies.

Allen, R. and S. Sims (2018): "Do pupils from low-income families get low-quality teachers? Indirect evidence from English schools," *Oxford Review of Education*, vol. 44(4), 441–458.

Andersen, S. C., L. Beuchert, H. S. Nielsen, and M. K. Thomsen (2020): "The Effect of Teacher's Aides in the Classroom: Evidence from a Randomized Trial," *Journal of the European Economic Association*, vol. 18(1), 469–505.

Ashraf, N., O. Bandiera, and B. K. Jack (2014): "No margin, no mission? A field experiment on incentives for public service delivery," *Journal of Public Economics*, vol. 120(C), 1–17.

Atkinson, A., S. Burgess, B. Croxson, P. Gregg, C. Propper, H. Slater, and D. Wilson (2009): "Evaluating the impact of performance related pay for teachers in England," *Labour Economics*, vol. 16, 251–261.

Bacolod, M. (2007): "Do Alternative Opportunities Matter? The Role of Female Labor Markets in the Decline of Teacher Quality," *Review of Eco*, vol. 89(4), 737–751.

Baker, R. B. (2019): "School resources and labor market outcomes: Evidence from early twentieth-century Georgia," *Economics of Education Review*, vol. 70, 35 – 47.

Bauhoff, S. (2014): "The effect of school district nutrition policies on dietary intake and overweight: A synthetic control approach," *Economics & Human Biology*, vol. 12(C), 45–55.

Becker, G. S. and B. R. Chiswick (1966): "Education and the Distribution of Earnings," *The American Economic Review*, vol. 56(1/2), 358–369.

Besley, T. and M. Ghatak (2005): "Competition and Incentives with Motivated Agents," *American Economic Review*, vol. 95(3), 616–636.

Besley, T. and M. Ghatak (2006): "Sorting with Motivated Agents: Implications for School Competition and Teacher Incentives," *Journal of the European Economic Association*, vol. 4(2-3), 404–414.

Billmeier, A. and T. Nannicini (2013): "Assessing Economic Liberalization Episodes: A Synthetic Control Approach," *The Review of Economics and Statistics*, vol. 95(3), 983–1001.

Blanden, J., L. Macmillan, and L. Sibieta (2015): "Understanding the improved performance of disadvantaged pupils in London," Technical report, CASE Social Policy in a Cold Climate.

Blatchford, P., P. Bassett, P. Brown, C. Martin, A. Russell, and R. Webster (2011): "The impact of support staff on pupilsâ âpositive approaches to learningâ and their academic progress," *British Educational Research Journal*, vol. 37(3), 443–464.

Blundell, R., L. Dearden, and B. Sianesi (2005): "Evaluating the effect of education on earnings: models, methods and results from the National Child Development Survey," *Journal of the Royal Statistical Society Series A*, vol. 168(3), 473–512.

Blundell, R. and M. C. Dias (2009): "Alternative Approaches to Evaluation in Empirical Microeconomics," *Journal of Human Resources*, vol. 44(3).

Bó, E. D., F. Finan, and M. A. Rossi (2013): "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service," *The Quarterly Journal of Economics*, vol. 128(3), 1169–1218.

Bohn, S., M. Lofstrom, and S. Raphael (2014): "Did the 2007 Legal Arizona Workers Act Reduce the State's Unauthorized Immigrant Population?" *The Review of Economics and Statistics*, vol. 96(2), 258–269.

Bonesronning, H., T. Falch, and B. Strom (2005): "Teacher sorting, teacher quality, and student composition," *European Economic Review*, vol. 49(2), 457–483.

Britton, J., L. Dearden, N. Shephard, and A. Vignoles (2016a): "How English domiciled graduate earnings vary with gender, institution attended, subject and socio-economic background," IFS Working Papers W16/06, Institute for Fiscal Studies.

Britton, J., L. Dearden, N. Shepherd, and A. Vignoles (2016b): "How English domiciled graduate earnings vary with gender, institution attended, subject and socio-economic background," IFS Working Paper W16/06.

Britton, J., C. Farquharson, and L. Sibieta (2019): "2019 Annual Report on Education Spending in England," IFS Report R162, Institute for Fiscal Studies.

Britton, J. and C. Propper (2016a): "Teacher pay and school productivity: Exploiting wage regulation," *Journal of Public Economics*, vol. 133, 75–89.

Britton, J. and C. Propper (2016b): "Teacher pay and school productivity: Exploiting wage regulation," *Journal of Public Economics*, vol. 133(C), 75–89.

Burgess, S. and M. Ratto (2003): "The Role of Incentives in the Public Sector: Issues and Evidence," *Oxford Review of Economic Policy*, vol. 19(2), 285–300.

Cabrera, J. M. and D. Webbink (2019): "Do higher salaries yield better teachers and better student outcomes?" *Journal of Human Resources*. http://jhr.uwpress.org/content/early/2019/04/04/jhr.55.4.0717-8911R3.abstract.

URL: *http://jhr.uwpress.org/content/early/2019/04/04/jhr.55.4.0717-8911R3.abstract*

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008): “Bootstrap-Based Improvements for Inference with Clustered Errors,” *The Review of Economics and Statistics*, vol. 90(3), 414–427.

Cameron, A. C. and D. L. Miller (2015): “A Practitioners Guide to Cluster-Robust Inference,” *Journal of Human Resources*, vol. 50(2), 317–372.

Card, D. and A. B. Krueger (1992): “Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States,” *Journal of Political Economy*, vol. 100(1), 1–40.

Cavallo, E., S. Galiani, I. Noy, and J. Pantano (2013): “Catastrophic Natural Disasters and Economic Growth,” *The Review of Economics and Statistics*, vol. 95(5), 1549–1561.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014): “Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood,” *American Economic Review*, vol. 104(9), 2633–2679.

Chetty, R. and N. Hendren (2018): “The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates,” *The Quarterly Journal of Economics*, vol. 133(3), 1163–1228.

Clotfelter, C., E. Glennie, H. Ladd, and J. Vigdor (2008): “Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina,” *Journal of Public Economics*, vol. 92(5-6), 1352–1370.

Cook, S. C., K. A. Mcduffie-Landrum, L. Oshita, and B. G. Cook (2011): *Co-Teaching for Students with Disabilities: A Critical and Updated Analysis of the Empirical Literature*, chapter 16 (Routledge New York), pp. 147–159.

Cowan, J. and D. Goldhaber (2016): “National Board Certification and teacher effectiveness: Evidence from Washington State,” *Journal of Research on Educational Effectiveness*, vol. 9(3), 233–258.

Crawford, C. and W. Jin (2014): “Payback time? Student debt and loan repayments: what will the 2012 reforms mean for graduates?” Technical report, Institute fo Fiscal Studies.

Cribb, J., C. Emmerson, and L. Sibieta (2014): “Public Sector Pay in the UK,” IFS Report R97, Insttitute for Fiscal Studies.

Cruz, T. (2018): “Teacher hiring decisions: How do governments react to an exogenous redistribution of education funds?” *Economics of Education Review*, vol. 67, 58 – 81. http://www.sciencedirect.com/science/article/pii/S0272775717306593.

URL: *http://www.sciencedirect.com/science/article/pii/S0272775717306593*

Cunha, F., J. J. Heckman, and S. M. Schennach (2010): “Estimating the Technology of Cognitive and Noncognitive Skill Formation,” *Econometrica*, vol. 78(3), 883–931.

de Ree, J., K. Muralidharan, M. Pradhan, and H. Rogers (2018): “Double for Nothing? Experimental Evidence on an Unconditional Teacher Salary Increase in Indonesia*,” *The Quarterly Journal of Economics*, vol. 133(2), 993–1039.

Dee, T. S. and J. Wyckoff (2015): “Incentives, Selection, and Teacher Performance: Evidence from IMPACT,” *Journal of Policy Analysis and Management*, vol. 34(2), 267–297.

Delfgaauw, J. and R. Dur (2007): “Signaling and screening of workers’ motivation,” *Journal of Economic Behavior & Organization*, vol. 62(4), 605–624.

Dixit, A. (2002): “Incentives and Organizations in the Public Sector: An Interpretative Review,” *Journal of Human Resources*, vol. 37(4)(4), 696–727.

Ehrenberg, R. G. and D. J. Brewer (1994): “Do school and teacher characteristics matter? Evidence from High School and Beyond,” *Economics of Education Review*, vol. 13(1), 1–17.

Eyles, A., S. Machin, and O. Silva (2018): "Academies 2 The New Batch: The Changing Nature of Academy Schools in England," *Fiscal Studies*, vol. 39(1), 121–158.

Farrell, P., A. Alborz, A. Howes, and D. Pearson (2010): "The impact of teaching assistants on improving pupilsâ academic achievement in mainstream schools: a review of the literature," *Educational Review*, vol. 62(4), 435–448.

Feng, L. and T. Sass (2018a): "The Impact of Incentives to Recruit and Retain Teachers in Hard to Staff Subjects," *Journal of Policy Analysis and Management*, vol. 37(1), 112–135.

Feng, L. and T. Sass (2018b): "The Impact of Incentives to Recruit and Retain Teachers in Hard to Staff Subjects," *Journal of Pol*, vol. 37(1), 112–135.

Figlio, D. N. and L. W. Kenny (2007): "Individual teacher incentives and student performance," *Journal of Public Economics*, vol. 91(5-6), 901–914.

Friend, M. P. and L. Cook (2016): *Interactions: Collaboration Skills for School Professionals* (Pearson Education, Upper Saddle River, United States).

Fryer, R. G. (2014): "Injecting Charter School Best Practices into Traditional Public Schools: Evidence from Field Experiments," *The Quarterly Journal of Economics*, vol. 129(3), 1355–1407.

Fryer, R. G. (2016): "The Production of Human Capital in Developed Countries: Evidence from 196 Randomized Field Experiments," NBER Working Papers 22130, National Bureau of Economic Research, Inc.

Fryer, R. G., S. D. Levitt, J. List, and S. Sadoff (2012): "Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment," NBER Working Papers 18237, National Bureau of Economic Research, Inc.

Gautier, P. A., A. Siegmann, and A. Van Vuuren (2009): "Terrorism and attitudes towards minorities: The effect of the Theo van Gogh murder on house prices in Amsterdam," *Journal of Urban Economics*, vol. 65(2), 113–126.

Gibbons, S., S. McNally, and M. Viarengo (2018): "Does Additional Spending Help Urban Schools? An Evaluation Using Boundary Discontinuities," *Journal of the European Economic Association*, vol. 16(5), 1618–1668.

Gilpin, G. A. (2012): "Teacher salaries and teacher aptitude: An analysis using quantile regressions," *Economics of Education Review*, vol. 31(3), 15 – 29.

Gorard, S., B. H. See, and N. Siddiqui (2014): "Switch on Reading," Technical report, Education Endowment Foundation.

Greaves, E. and L. Sibieta (2019): "Constrained optimisation? Teacher salaries, school resources and student achievement," *Economics of Education Review*, vol. 73.

Grier, K. and N. Maynard (2016): "The economic consequences of Hugo Chavez: A synthetic control analysis," *Journal of Economic Behavior & Organization*, vol. 125(C), 1–21.

Hanushek, E. A. (2003): "The Failure of Input-Based Schooling Policies," *Economic Journal*, vol. 113(485), 64–98.

Hanushek, E. A. (2006): "School Resources," in *Handbook of the Economics of Education*, vol. 2, ed. by E. Hanushek and F. Welch, chapter 14 (Elsevier), pp. 865–908.

Hanushek, E. A., E. F. Kain, and S. G. Rivkin (2004): "Why Public Schools Lose Teachers," *Journal of Human Resources*, vol. 39(2).

Hanushek, E. A., S. Link, and L. Woessmann (2013): "Does school autonomy make sense everywhere? Panel estimates from PISA," *Journal of Development Economics*, vol. 104(C), 212–232.

Hendricks, M. D. (2014): "Does it pay to pay teachers more? Evidence from Texas," *Journal of Public Economics*, vol. 109(C), 50–63.

Herrmann, M. A. and J. E. Rockoff (2012): "Worker Absence and Productivity: Evidence from Teaching," *Journal of Labor Economics*, vol. 30(4), 749–782.

Hinrichs, P. (2012): "The Effects of Affirmative Action Bans on College Enrollment, Educational Attainment, and the Demographic Composition of Universities," *The Review of Economics and Statistics*, vol. 94(3), 712–722.

Hoxby, C. and A. Leigh (2005): "Pulled Away or Pushed Out? Explaining the Decline of Teacher Aptitude in the United States," *American Economic Review*, vol. 94, 236–240.

Hutchings, M., C. Greenwood, S. Hollingworth, A. Mansaray, A. Rose, S. Minty, and K. Glass (2012): "Evaluation of the City Challenge programme," Technical report, Department for Education.

Hyman, J. (2017): "Does Money Matter in the Long Run? Effects of School Spending on Educational Attainment," *American Economic Journal: Economic Policy*, vol. 9(4), 256–80.

Jackson, C. K. (2018): "Does School Spending Matter? The New Literature on an Old Question," Working Paper 25368, National Bureau of Economic Research.

Jackson, C. K., R. C. Johnson, and C. Persico (2015): " The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms *," *The Quarterly Journal of Economics*, vol. 131(1), 157–218.

Jackson, C. K., R. C. Johnson, and C. Persico (2016): "The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms," *The Quarterly Journal of Economics*, vol. 131(1), 157–218.

Jackson, C. K., C. Wigger, and H. Xiong (2018): "Do School Spending Cuts Matter? Evidence from the Great Recession," Working Paper 24203, National Bureau of Economic Research.

Johnson, R. C. and C. K. Jackson (2019): "Reducing Inequality through Dynamic Complementarity: Evidence from Head Start and Public School Spending," *American Economic Journal: Economic Policy*, vol. 11(4), 310–49.

Jones, M. D. (2013): "Teacher behavior under performance pay incentives," *Economics of Education Review*, vol. 37(C), 148–164.

Kane, T. J., J. E. Rockoff, and D. Staiger (2008): "What Does Certification Tell Us about Teacher Effectiveness? Evidence from New York City," *Economics of Education Review*, vol. 27(6), 615–31.

Kirkeboen, L., E. Leuven, and M. Mogstad (2016): "Field of Study, Earnings, and Self-Selection," *Quarterly Journal of Econimcs*, vol. 131(3), 1057–1111.

Kosfeld, M. and S. Neckermann (2011): "Getting More Work for Nothing? Symbolic Awards and Worker Performance," *American Economic Journal: Microeconomics*, vol. 3(3), 86–99.

Kreif, N., R. Grieve, D. Hangartner, A. J. Turner, S. Nikolova, and M. Sutton (2016): "Examination of the Synthetic Control Method for Evaluating Health Policies with Multiple Treated Units," *Health Economics*, vol. 25(12), 1514–1528.

Lafortune, J., J. Rothstein, and D. W. Schanzenbach (2018): "School Finance Reform and the Distribution of Student Achievement," *American Economic Journal: Applied Economics*, vol. 10(2), 1–26.

Lavy, V. (2009): "Performance Pay and Teachers' Effort, Productivity, and Grading Ethics," *American Economic Review*, vol. 99(5), 1979–2011.

Lee, D. S. and T. Lemieux (2010): "Regression Discontinuity Designs in Economics," *Journal of Economic*

Literature, vol. 48(2), 281–355.

Leigh, A. (2012): "Teacher pay and teacher aptitude," *Economics of Education Review*, vol. 31(3), 41–53.

Loeb, S. and M. E. Page (2000): "Examining The Link Between Teacher Wages And Student Outcomes: The Importance Of Alternative Labor Market Opportunities And Non-Pecuniary Variation," *The Review of Economics and Statistics*, vol. 82(3), 393–408.

Machin, S. and S. McNally (2008): "The literacy hour," *Journal of Public Economics*, vol. 92(5-6), 1441–1462.

Machin, S., S. McNally, and C. Meghir (2010): "Resources and Standards in Urban Schools," *Journal of Human Capital*, vol. 4(4), 365–393.

Muijs, D. (2015): "Improving schools through collaboration: a mixed methods study of school-to-school partnerships in the primary sector," *Oxford Review of Education*, vol. 41(5), 563–586.

NAO (2016): "Training New Teachers," Technical report, National Audit Office.

NFER (2014): "Catch Up Numeracy," Technical report, Education Endowment Foundation.

Nickell, S. and G. Quintini (2002): "The Consequences of The Decline in Public Sector Pay in Britain: A Little Bit of Evidence," *Economic Journal*, vol. 112(477), F107–F118.

Nicoletti, C. and B. Rabe (2018): "The effect of school spending on student achievement: addressing biases in valueâ added models," *Journal of the Royal Statistical Society Series A*, vol. 181(2), 487–515.

Papay, J. P. and M. A. Kraft (2015): "Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement," *Journal of Public Economics*, vol. 130(C), 105–119.

Penney, J. (2018): "Dynamic Treatment Effects Of Teachers' Aides In An Experiment With Multiple Randomizations," *Economic Inquiry*, vol. 56(2), 1244–1260.

Peri, G. and V. Yasenov (2019): "The Labor Market Effects of a Refugee Wave: Synthetic Control Method Meets the Mariel Boatlift," *Journal of Human Resources*, vol. 54(2), 267–309.

Psacharopoulos, G. and H. A. Patrinos (2018): "Returns to investment in education: a decennial review of the global literature," *Education Economics*, vol. 26(5), 445–458.

Ransom, M. R. and D. P. Sims (2010): "Estimating the Firm's Labor Supply Curve in a "New Monopsony" Framework: Schoolteachers in Missouri," *Journal of Labor Economics*, vol. 28(2), 331–355.

Roy, A. (1951): "Some thoughts on the distribution of earnings," *Oxford Economics Papers (New Series)*, vol. 3, 135–146.

Shapiro, C. and J. E. Stiglitz (1984): "Equilibrium Unemployment as a Worker Discipline Device," *American Economic Review*, vol. 74(3), 433–44.

Sibieta, L. (2015a): "The distribution of school funding and inputs in England: 1993-2013," Working paper w15/10, Institute for Fiscal Studies.

Sibieta, L. (2015b): "School Spending in England 2010-2015," *Fiscal Studies*, vol. 36(3), 283–302.

Sibieta, L. (2018): "The teacher labour market in England: shortages, subject expertise and incentives," Technical report, Education Policy Institute.

Springer, M. G., W. A. Swain, and L. A. Rodriguez (2016): "Effective teacher retention bonuses: evidence from Tennessee," *Educational Evaluation and Policy Analysis*, vol. 38(2).

Steele, J. L., R. J. Murnane, and J. B. Willett (2010): "Do financial incentives help low-performing schools attract and keep academically talented teachers? Evidence from California," *Journal of Policy Analysis and Management*, vol. 29(3), 451–478.

Swain, W. A., L. A. Rodriguez, and M. G. Springer (2019): "Selective retention bonuses for highly effective teachers in high poverty schools: Evidence from Tennessee," *Economics of Education Review*, vol. 68(C), 148–160.

Tirole, J. and R. Bénabou (2006): "Incentives and Prosocial Behavior," *American Economic Review*, vol. 96(5), 1652–1678.

Todd, P. E. and K. I. Wolpin (2003): "On The Specification and Estimation of The Production Function for Cognitive Achievement," *Economic Journal*, vol. 113(485), F3–F33.

Tran, H. (2017): "Does the Pay Stance of South Carolina Public School Districts Influence their Math and Science Achievement Scores?" *Journal of Education Finance*, vol. 43(2), 105–122.

Wayne, A. J. and P. Youngs (2003): "Teacher Characteristics and Student Achievement Gains: A Review," *Review of Educational Research*, vol. 73(1), 89–122.

Wiswall, M. (2013): "The Dynamics of Teacher Quality," *Journal of Public Economics*, vol. 100, 61–78.

Zabalza, A., T.-P. and G. Williams (1979): *The Economics of Teacher Supply* (Cambridge University Press).

Appendix A - Appendix Tables and Figures for Chapter 2

Table A1: Balance of pupil characteristics and summary statistics across Fringe Boundary (1km)

	Within 1km		
	Inside	**Outside**	**Difference**
A) Pupil Characteristics			
Prop. FSM	0.07 (0.07)	0.08 (0.08)	-0.01
Prop. SEN (no statement)	0.21 (0.10)	0.21 (0.12)	0
Prop. SEN (statement)	0.02 (0.02)	0.01 (0.02)	0.01***
Prop. EAL	0.06 (0.09)	0.06 (0.10)	0
Prop. non-white	0.15 (0.11)	0.15 (0.13)	0
Number of Pupils	246.3 (124.5)	231.6 (100.4)	14.77
IMD Rank	0.75 (0.15)	0.72 (0.17)	0.03*
IDACI Rank	0.68 (0.16)	0.66 (0.18)	0.02
Pseudo R-Squared			*0.08*
Likelihood Ratio Test (p-value)			*<0.01*
B) Funding and Expenditure			
Total grant funding per pupil (log)	8.218 (0.177)	8.204 (0.194)	0.014
Total expenditure per pupil (log)	8.209 (0.176)	8.198 (0.194)	0.012
C) KS2 Outcomes			
English fine points score (std)	0.139 (0.332)	0.097 (0.374)	0.042
Maths fine points score (std)	0.108 (0.319)	0.098 (0.331)	0.011
Number of observations	280	319	
Number of Schools	56	64	

Notes: *** denotes where difference between schools on inside and outside of boundary are significant at the 1%, ** at 5%, and * at 10% level. Standard deviations are in parentheses. The unit of analysis is the school. The sample includes all primary schools that are present in the National Pupil Database in all of the academic years 2005/2006, 2006/2007, 2007/2008,2008/2009, 2010/2011 and where the school is within 1km of the fringe London pay boundary. The likelihood ratio tests the null hypothesis that the differences in school characteristics are jointly zero. The Pseudo R-squared is taken from a probit regression of an indicator of whether schools are in the high-pay area on the set of school characteristics reported in panel A.

Table A2: Balance of pupil characteristics and summary statistics across Fringe Boundary (3km)

	Within 3km		
	Inside	**Outside**	**Difference**
A) Pupil Characteristics			
Prop. FSM	0.083 (0.084)	0.087 (0.095)	-0.003
Prop. SEN (no statement)	0.209 (0.106)	0.208 (0.114)	0.001
Prop. SEN (statement)	0.020 (0.022)	0.017 (0.020)	0.003**
Prop. EAL	0.067 (0.097)	0.072 (0.106)	-0.005
Prop. non-white	0.169 (0.134)	0.167 (0.146)	0.002
Number of Pupils	266.73 (125.6)	246.19 (108.7)	20.534***
IMD Rank	0.717 (0.185)	0.692 (0.186)	0.025**
IDACI Rank	0.639 (0.186)	0.637 (0.193)	0.002
Pseudo R-Squared			*0.051*
Likelihood Ratio Test (p-value)			*<0.01*
B) Funding and Expenditure			
Total grant funding per pupil (log)	8.204 (0.180)	8.214 (0.189)	-0.01
Total expenditure per pupil (log)	8.198 (0.180)	8.207 (0.191)	-0.01
C) KS2 Outcomes			
English fine points score (std)	0.105 (0.434)	0.102 (0.361)	0.003
Maths fine points score (std)	0.070 (0.449)	0.093 (0.349)	-0.023
Number of observations	865	966	
Number of Schools	173	195	

Notes: *** denotes where difference between schools on inside and outside of boundary are significant at the 1%, ** at 5%, and * at 10% level. Standard deviations are in parentheses. The unit of analysis is the school. The sample includes all primary schools that are present in the National Pupil Database in all of the academic years 2005/2006, 2006/2007, 2007/2008,2008/2009, 2010/2011 and where the school is within 3km of the fringe London pay boundary. The likelihood ratio tests the null hypothesis that the differences in school characteristics are jointly zero. The Pseudo R-squared is taken from a probit regression of an indicator of whether schools are in the high-pay area on the set of school characteristics reported in panel A.

Table A3: Estimated difference in student achievement in English across the Fringe London boundary

	Raw gap	OLS	FILM	Kernel Matching
Within 1km				
2006	0.038	0.040	0.038	0.052
	[0.077]	[0.048]	[0.032]	[0.078]
2007	0.133	0.134	0.124	0.054
	[0.101]	[0.088]	[0.053]*	[0.100]
2008	0.009	-0.038	-0.045	-0.089
	[0.079]	[0.040]	[0.037]	[0.077]
2009	0.075	0.036	0.038	0.011
	[0.113]	[0.051]	[0.040]	[0.076]
2011	-0.019	0.000	-0.034	-0.041
	[0.118]	[0.065]	[0.031]	[-0.070]
Within 2km				
2006	0.027	0.027	0.036	0.007
	[0.077]	[0.023]	[0.021]	[0.052]
2007	0.037	0.033	0.029	0.034
	[0.067]	[0.028]	[0.016]	[0.058]
2008	0.033	0.005	0.003	0.005
	[0.086]	[0.025]	[0.025]	[0.051]
2009	0.051	0.020	0.007	-0.003
	[0.083]	[0.032]	[0.016]	[0.050]
2011	0.022	-0.001	-0.011	-0.036
	[0.099]	[0.039]	[0.029]	[0.050]
Within 3km				
2006	0.039	0.037	0.045	0.022
	[0.076]	[0.026]	[0.015]*	[0.043]
2007	0.017	0.017	0.013	0.012
	[0.077]	[0.023]	[0.018]	[0.048]
2008	0.030	0.013	0.009	0.007
	[0.098]	[0.026]	[0.011]	[0.041]
2009	0.010	0.023	-0.008	-0.008
	[0.100]	[0.025]	[0.017]	[0.039]
2011	-0.002	0.012	-0.002	-0.018
	[0.109]	[0.031]	[0.026]	[0.042]
School and Year Controls		Yes		

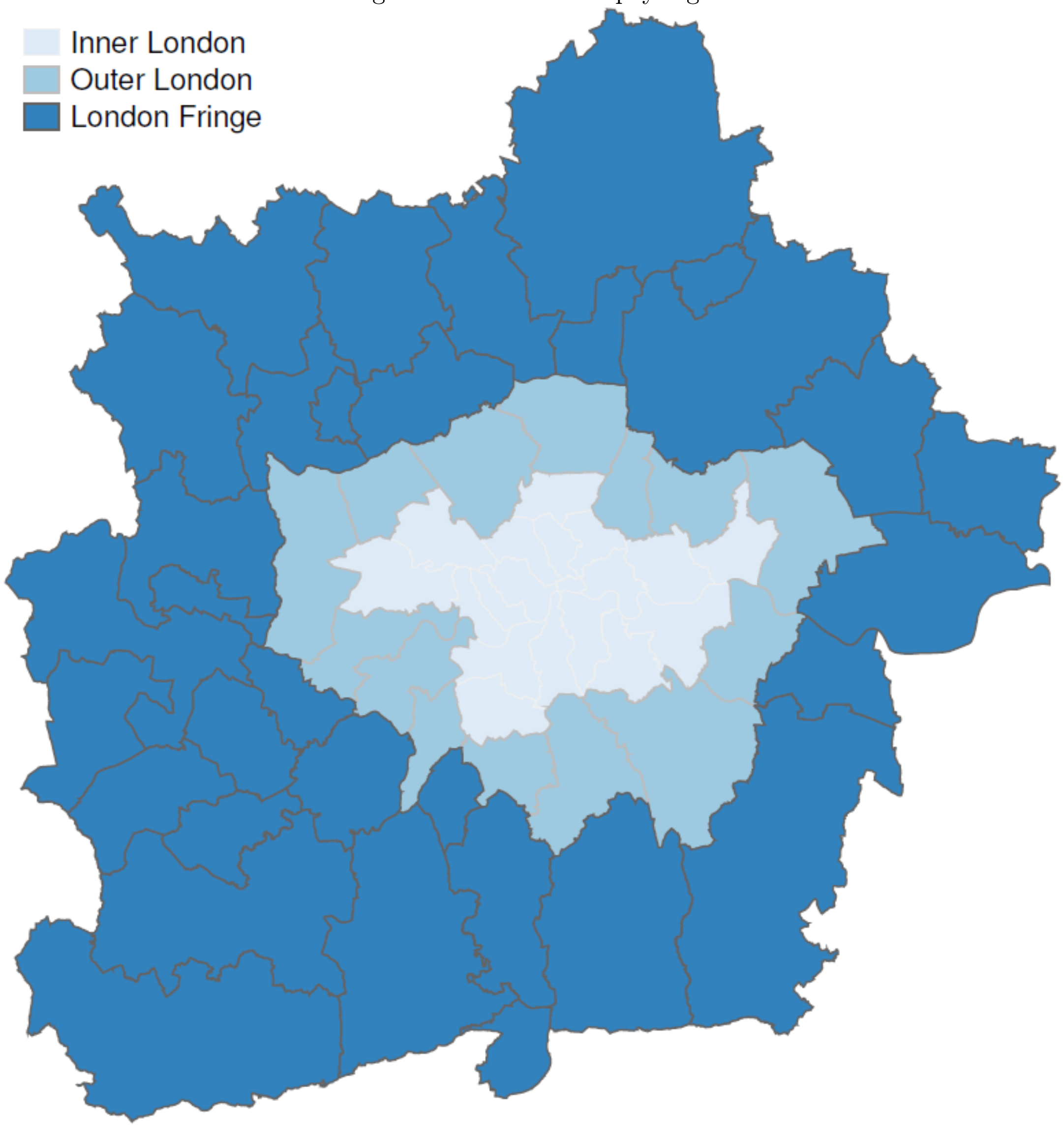
Notes: *** denotes significance at 1%, ** at 5%, and * at 10% level. Standard errors in brackets clustered at the local authority level. FILM refers to Fully Interacted Linear Matching. The unit of analysis is the school. The sample includes all primary schools that are present in the National Pupil Database in all of the academic years 2005/2006, 2006/2007, 2007/2008,2008/2009, 2010/2011 and where each dependent variable is observed, within 2km of the fringe London pay boundary. The coefficients reported represents the difference in KS2 fine point score in English (standardised at the national level) associated with the high-pay side of the boundary.

Table A4: Estimated difference in student achievement in Maths across the Fringe London boundary

	Raw gap	OLS	FILM	Kernel Matching
Within 1km of the Fringe Boundary				
2006	0.002	-0.002	-0.003	-0.003
	[0.075]	[0.053]	[0.045]	[0.075]
2007	0.066	0.076	0.073	0.038
	[0.101]	[0.074]	[0.048]	[0.087]
2008	-0.017	-0.035	-0.033	-0.083
	[0.082]	[0.036]	[0.040]	[0.073]
2009	0.053	0.018	0.002	-0.013
	[0.100]	[0.026]	[0.034]	[0.064]
2011	-0.017	0.012	-0.014	-0.009
	[0.094]	[0.053]	[0.029]	[0.064]
Within 2km of the Fringe Boundary				
2006	-0.021	-0.030	-0.016	-0.045
	[0.065]	[0.029]	[0.023]	[0.051]
2007	0.012	0.000	-0.001	-0.003
	[0.079]	[0.031]	[0.017]	[0.055]
2008	0.031	0.032	0.027	0.032
	[0.071]	[0.015]	[0.018]	[0.047]
2009	-0.007	-0.040	-0.046	-0.050
	[0.076]	[0.031]	[0.019]*	[0.047]
2011	0.021	-0.001	-0.014	-0.031
	[0.088]	[0.036]	[0.032]	[0.046]
Within 3km of the Fringe Boundary				
2006	0.011	0.007	0.014	-0.012
	[0.074]	[0.018]	[0.018]	[0.042]
2007	0.013	0.020	0.017	0.006
	[0.079]	[0.020]	[0.014]	[0.046]
2008	0.004	0.009	-0.007	-0.002
	[0.098]	[0.028]	[0.024]	[0.049]
2009	-0.033	-0.031	-0.045	-0.045
	[0.092]	[0.021]	[0.012]**	[0.038]
2011	-0.017	0.000	-0.016	-0.028
	[0.104]	[0.034]	[0.021]	[0.040]
School and Year Controls		Yes		

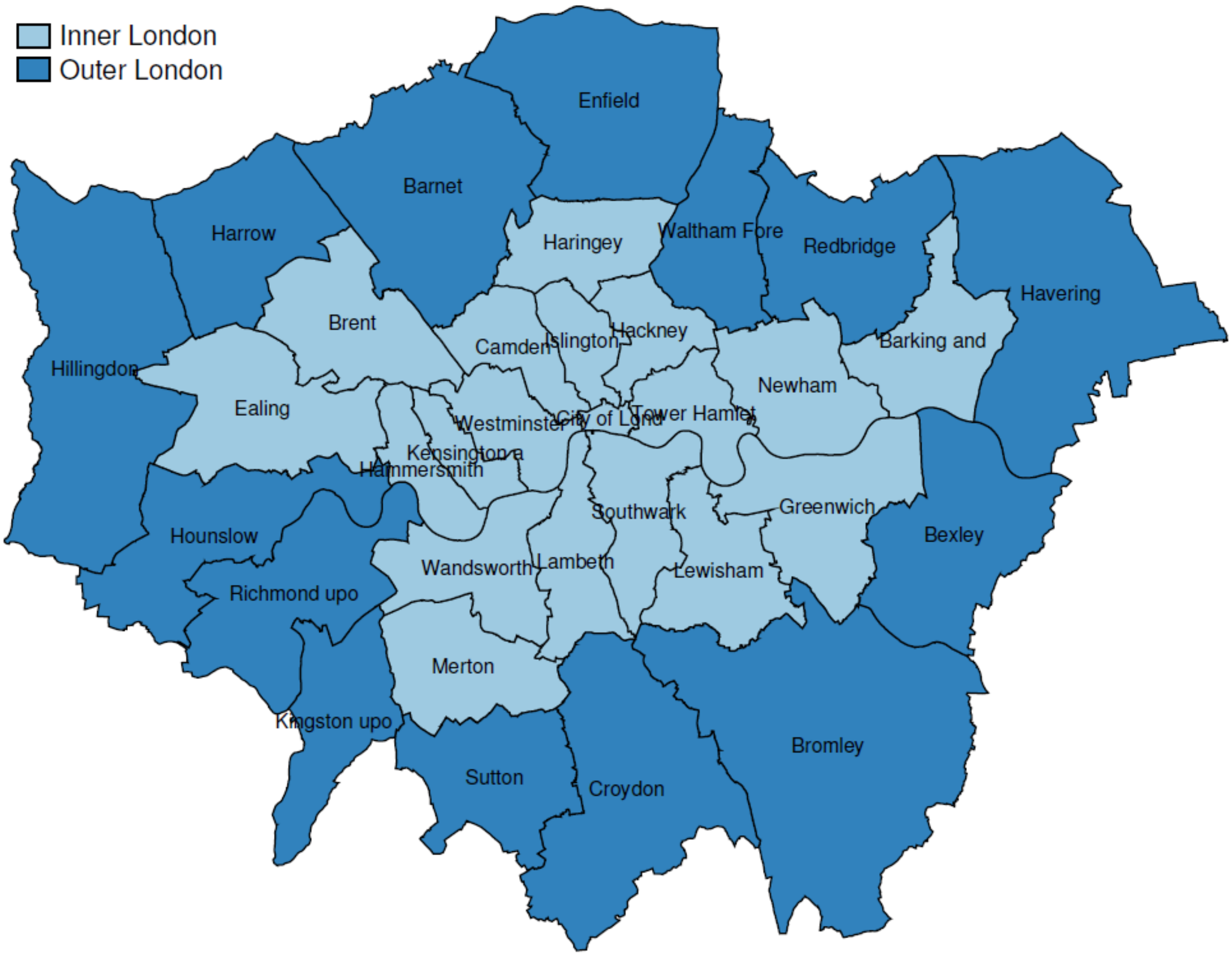
Notes: *** denotes significance at 1%, ** at 5%, and * at 10% level. Standard errors in brackets clustered at the local authority level. FILM refers to Fully Interacted Linear Matching. The unit of analysis is the school. The sample includes all primary schools that are present in the National Pupil Database in all of the academic years 2005/2006, 2006/2007, 2007/2008,2008/2009, 2010/2011 and where each dependent variable is observed, within 2km of the fringe London pay boundary. The coefficient reported represents the difference in KS2 fine point score in Maths (standardised at the national level) associated with the high-pay side of the boundary.

Figure A1: All teacher pay regions



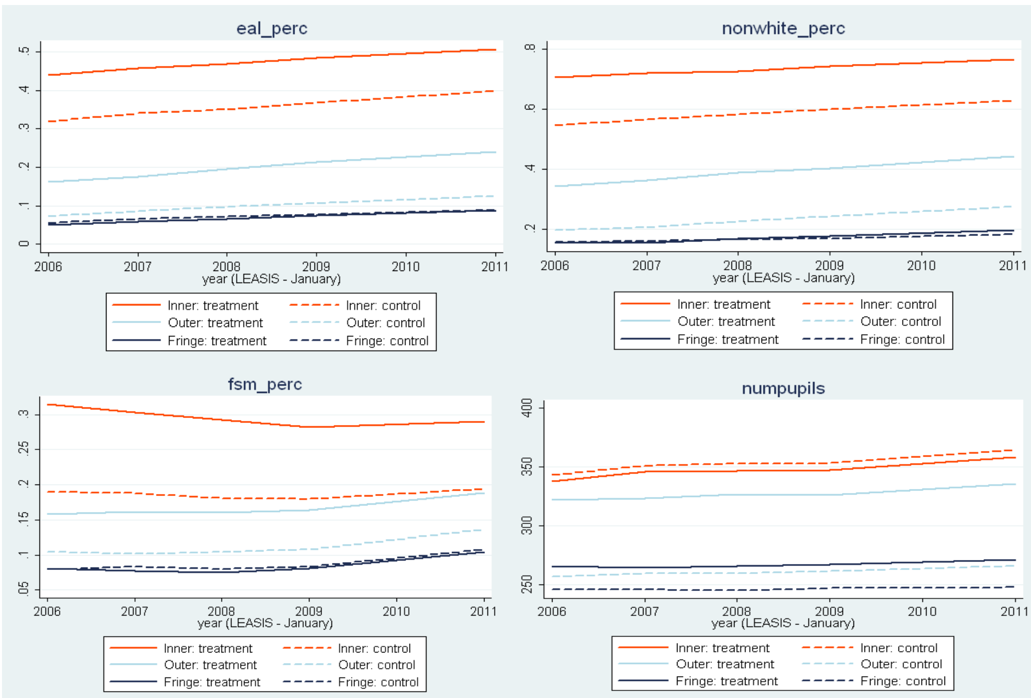
Source: School Teachers Pay and Conditions Document.

Figure A2: Inner and Outer London pay regions



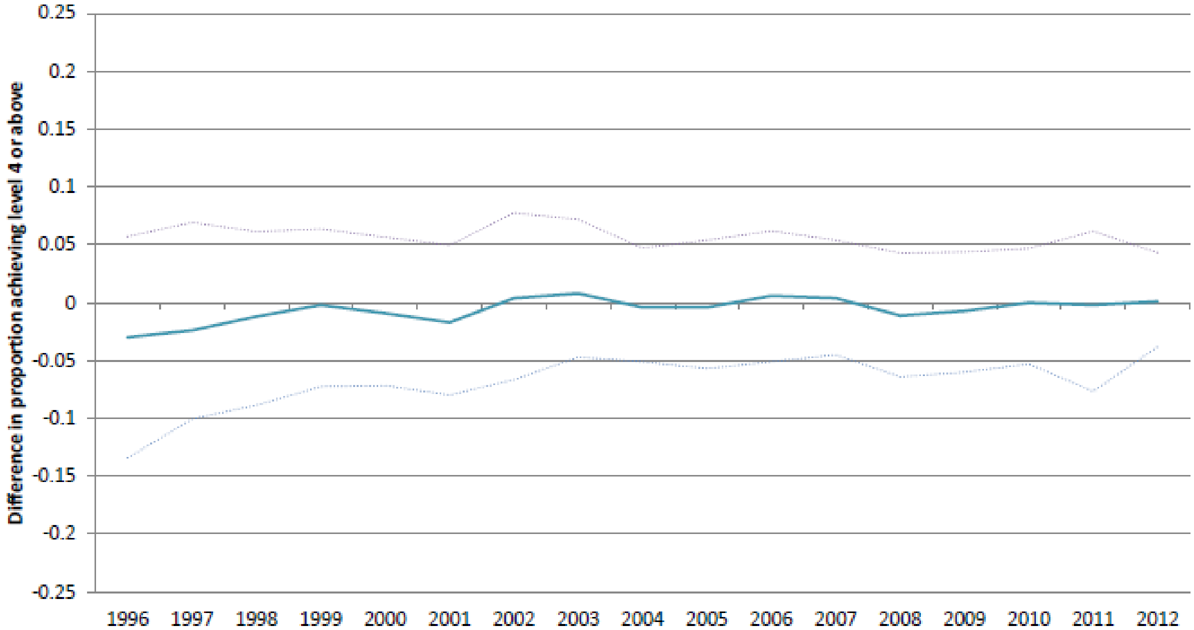
Source: School Teachers Pay and Conditions Document.

Figure A3: Changes in covariates over time for schools inside and outer pay boundaries



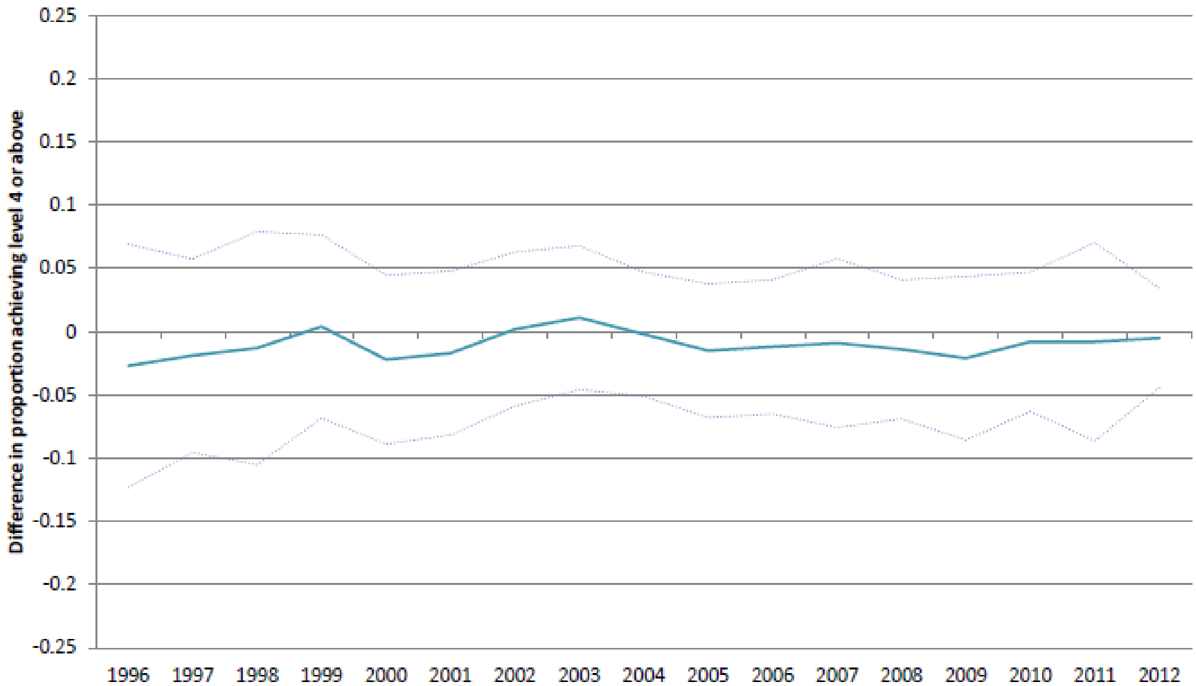
Notes Sample includes primary schools within 2km of each boundary. 'Treatment' refers to schools on the high-pay side of the boundary. Source: National Pupil Database and LEASIS.

Figure A4: Difference in proportion achieving level 4 or above in KS2 English at Fringe boundary for schools within 2km (with 95% CIs)



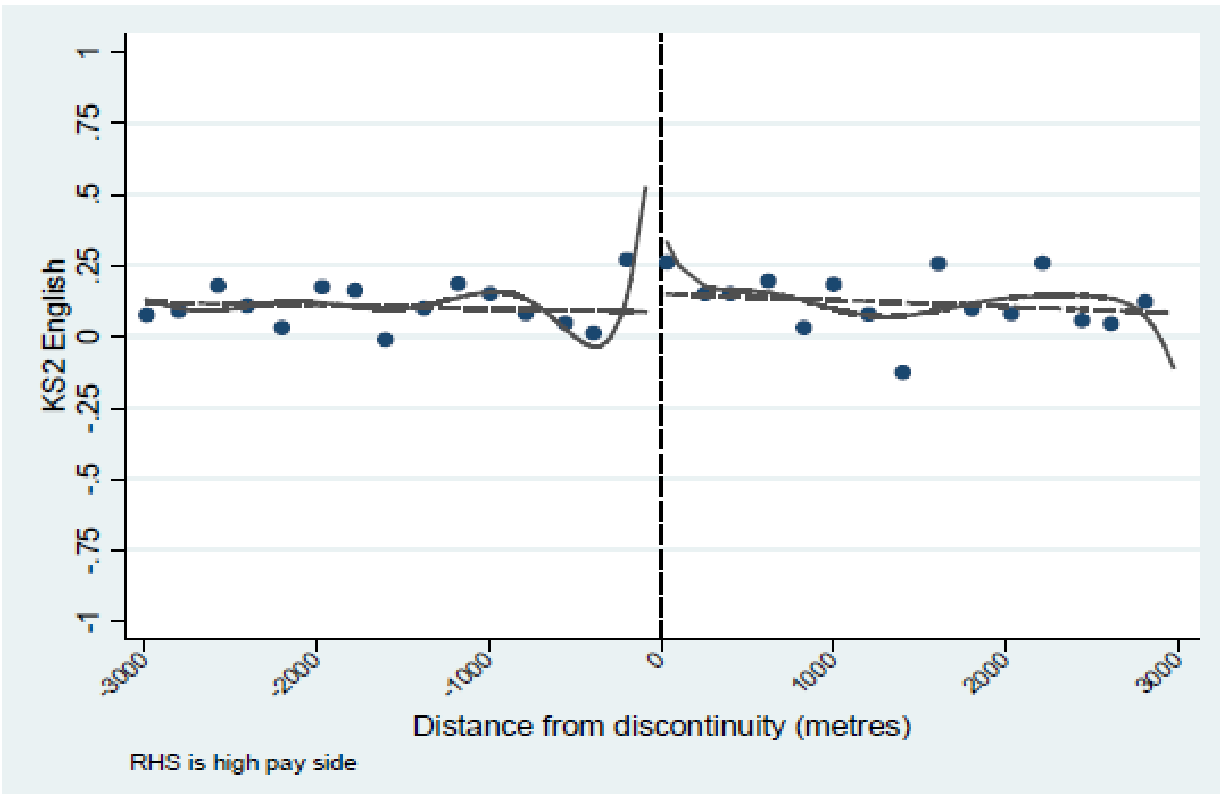
Notes Solid lines represents raw difference in proportion of pupils achieving level 4 in English at schools within 2km and either side of the fringe London pay boundary. Dashed lines show 95% confidence intervals.

Figure A5: Difference in proportion achieving level 4 or above in KS2 Maths at Fringe boundary for schools within 2km (with 95% CIs)



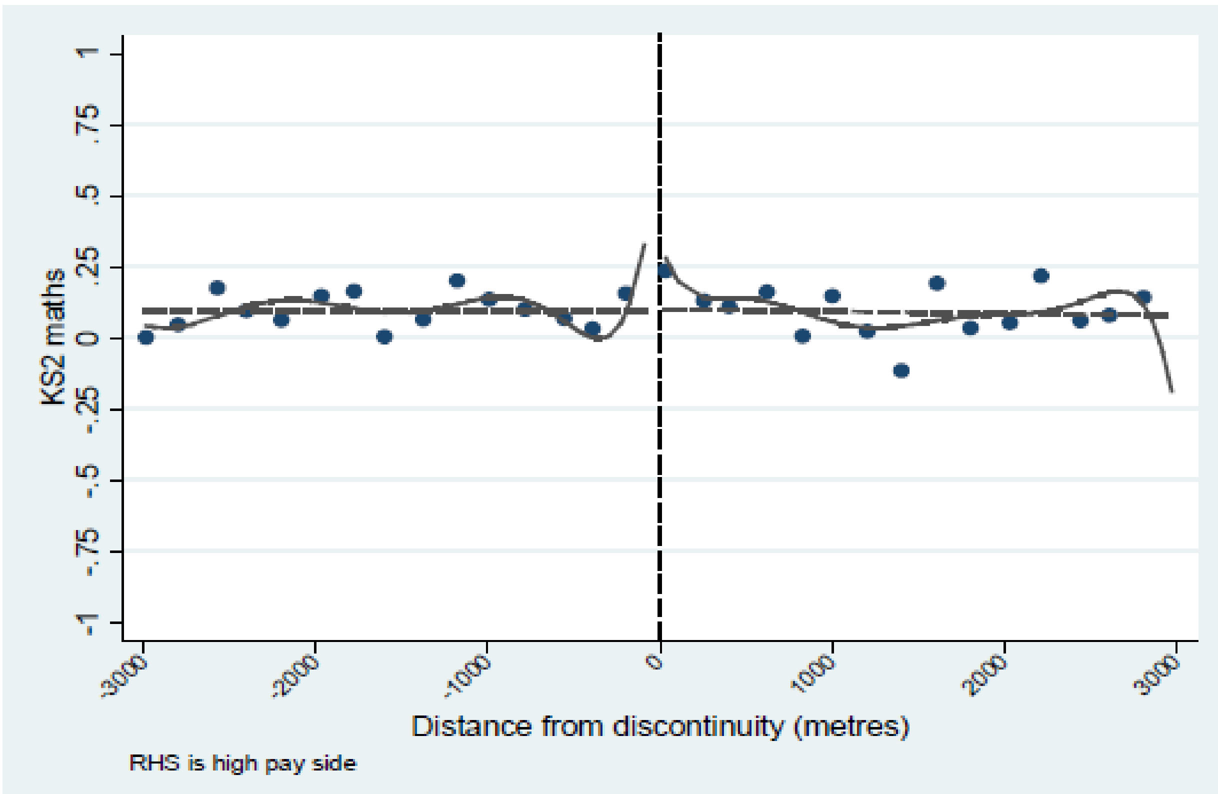
Notes Solid lines represents raw difference in proportion of pupils achieving level 4 in Maths at schools within 2km and either side of the fringe London pay boundary. Dashed lines show 95% confidence intervals.

Figure A6: Relationship between distance to the Fringe boundary and English test scores



Notes Dashed line represents linear specification estimated either side of boundary, solid line is based on a 7th order polynomial and dots are local averages by 200m from the pay zone boundary. KS2 English test scores are standardised at the national level.

Figure A7: Relationship between distance to the Fringe boundary and Maths test scores



Notes Dashed line represents linear specification estimated either side of boundary, solid line is based on a 7th order polynomial and dots are local averages by 200m from the pay zone boundary. KS2 Maths test scores are standardised at the national level.

Appendix B - Appendix Tables and Figures for Chapter 3

Table B1: UCAS Tariff Scores

Qualification	**UCAS Tariff Score**
A-level marks	
A* – A Level	140
A – A Level	120
B – A Level	100
C – A Level	80
D – A Level	60
E – A Level	40
Other example qualifications	
Distinction – BTEC (Group C)	40
45 IB Diploma Points	720
30 IB Diploma Points	392

Notes: UCAS tariff points pre-September 2016 measure, available at
https://graduates.teachfirst.org.uk/sites/graduates.teachfirst.org.uk/files/ucas-points-tariff-pre-17.pdf

Table B2: Subject Groupings

Qualification / Mark	UCAS Tariff Score	JACS Codes
High Priority Subjects		
Physics	Physical Sciences, Materials Science, Astronomy, Engineering	F0, F2, F3, F5, H
Maths	Maths, Operational Research, Statistics	G01, G1-3, G7, G91
Computing	Computer Science, Software Engineering	G02, G4-6, G92, I
Chemistry	Chemistry, Chemical Engineering	F1, H8
Languages	Ancient and Modern Foreign Languages	Q4, Q6, Q7, R, T
Other priority subjects		
Biology	Biological Sciences	C
English	English and Literary Studies, Linguistics	Q1, Q2, Q3
History	History, Classics, Philosophy, Archaeology	Q5, Q8, Q9, V
Geography	Geography, Geology, Oceanography	L7, F6, F7, F8
Design and Technology	All types of technology courses	J
Non-priority subjects		
Social Science and Law	Social Sciences, Law, Business, Management Studies	L (ex. L7), M, N
Medical Sciences	Clinical medicine, Clinical dentistry and Nursing	A, B
Arts	Art, Music, Design, Drama	W
Education	Teacher Training, Education Studies	X
Other	Veterinary Sciences, Architecture and Design, Media Studies	D, K, P, Q0,Y0,

Notes: Authors Calculations using JACS coding

Table B4: Estimated effect of subject priority status on average aptitude of teachers in training by sample

	(1) All FT Undergraduates		**(2) Under 25**		**(3) All undergraduates**		**(4) All HE Leavers**		**(5) All UK HE Leavers**	
Degree Class & Subject Group	**2007/08 to 2011/12**	**Change in effect after 2012/13**	**2007/08 to 2011/12**	**Change in effect after 2012/13**	**2007/08 to 2011/12**	**Change in effect after 2012/13**	**2007/08 to 2011/12**	**Change in effect after 2012/13**	**2007/08 to 2011/12**	**Change in effect after 2012/13**
Subject Grouping										
Non Priority	4.8	-4.7	5.7	-4.7	4.1	-5.6	3.7	-7.3	0.8	-3.9
	[-30.1;23.9]	*[-14.2;25.1]*	*[-30.2;24.9]*	*[-13.4;24.1]*	*[-39.1;23.7]*	*[-15.1;23.0]*	*[-31.3;23.1]*	*[-16.3;22.2]*	*[-40.7;23.1]*	*[-13.5;31.9]*
Other Priority	-2.2	-4.4	-2.2	-4.8	-1.7	-3.8	-1.5	-2.2	-0.4	-2.3
	[-26.9;13.8]	*[-16.7; 8.1]*	*[-25.6;15.1]*	*[-16.6; 7.3]*	*[-26.7;13.9]*	*[-15.1; 8.7]*	*[-27.0;13.8]*	*[-13.3; 8.9]*	*[-27.6;17.4]*	*[-11.9; 8.7]*
High Priority	-18.3	-2.1	-18.7	-2.5	-17.0	-0.9	-16.4	0.8	-14.4	1.2
	[-42.0; 6.8]	*[-15.2; 9.3]*	*[-42.6; 5.8]*	*[-15.3; 8.4]*	*[-41.9; 7.2]*	*[-12.7;10.5]*	*[-43.3; 6.2]*	*[-10.2;10.9]*	*[-42.8;10.9]*	*[-9.9;11.7]*
Individual Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,202,397		1,161,766		1,239,440		1,256,048		1,494,405	
Pseudo R-squared	0.159		0.135		0.176		0.173		0.168	

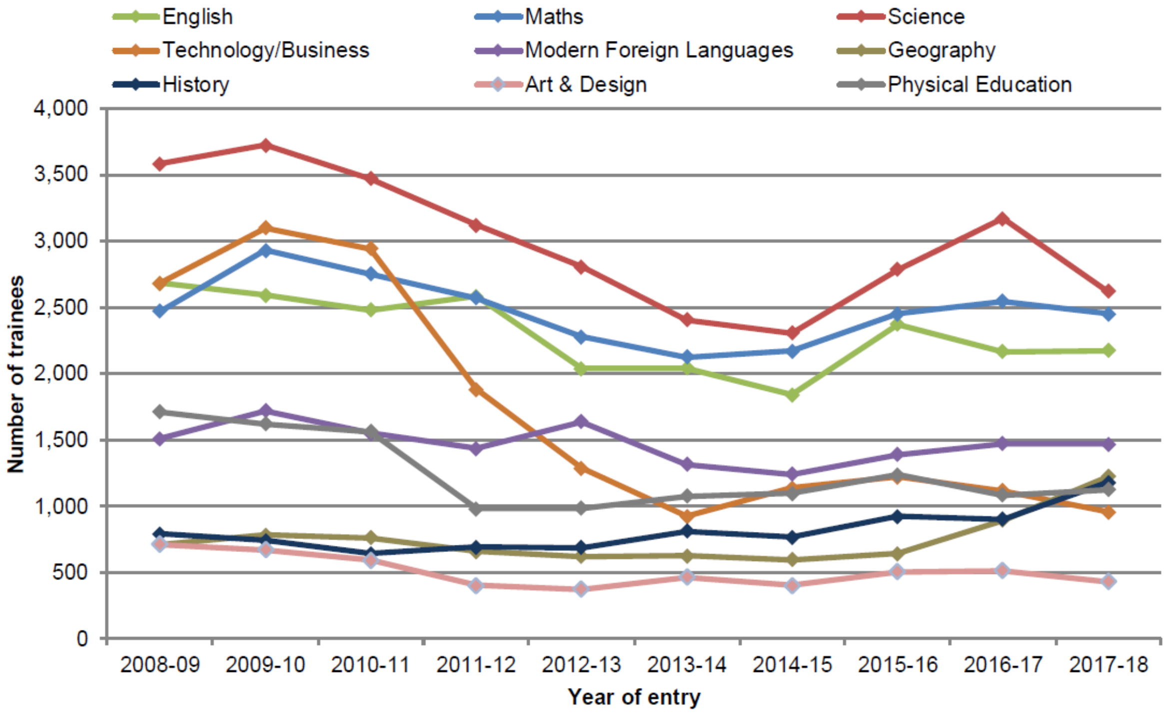
Notes: Author's calculations using HESA Destinations of Leavers from Higher Education. Teachers in training are defined as those undertaking a postgraduate teacher training course at the survey date. Sample only includes full-time undergraduate leavers from higher education institutions in England. Individual covariates include age, gender, region and subject-by-year fixed effects. 95% confidence intervals are shown in brackets and estimated using a wild cluster bootstrap at the subject level with Webb weights.

Table B3: Effect of degree class by subject priority on likelihood to train as teacher across samples

	(1) FT Undergraduates		**(2) Under 25**		**(3) All undergraduates**		**(4) All HE Leavers**		**(5) All UK HE Leavers**	
Degree Class & Subject Group	2007/08 to 2011/12	Change in effect after 2012/13	2007/08 to 2011/12	Change in effect after 2012/13	2007/08 to 2011/12	Change in effect after 2012/13	2007/08 to 2011/12	Change in effect after 2012/13	2007/08 to 2011/12	Change in effect after 2012/13
Non-priority (Baseline)										
First Class	0.055	0.005*	0.048	0.006*	0.035	0.005*	0.008	0.005**	0.004	0.006**
	[-0.007;0.117]	[-0.011;0.021]	[0.002;0.093]	[-0.007;0.019]	[-0.027;0.097]	[-0.011;0.021]	[-0.010;0.026]	[0.002;0.008]	[-0.013;0.021]	[0.001;0.010]
Upper Second	0.037**	0.003	0.034**	0.003	0.026*	0.003	0.020*	0.002	0.016*	0.003
	[-0.013;0.086]	[-0.007;0.012]	[-0.004;0.073]	[-0.007;0.012]	[-0.028;0.080]	[-0.009;0.014]	[-0.022;0.063]	[-0.008;0.013]	[-0.017;0.049]	[-0.009;0.015]
Lower Second	0.045*	0.000	0.042**	-0.001	0.027	0.000	0.019*	0.001	0.014	0.001
	[0.010;0.080]	[-0.006;0.005]	[0.021;0.063]	[-0.008;0.006]	[-0.008;0.062]	[-0.003;0.004]	[-0.010;0.048]	[-0.003;0.004]	[-0.011;0.039]	[-0.002;0.004]
Other Classification	[omitted]	-0.001	[omitted]	-0.002	[omitted]	-0.002	[omitted]	-0.001	[omitted]	-0.001
		[-0.012;0.010]		[-0.012;0.008]		[-0.008;0.005]		[-0.008;0.005]		[-0.008;0.006]
Other Priority (Relative to Baseline)										
First Class	0.010	-0.001	0.012	-0.002	0.009	-0.001	0.012	-0.001	0.013	-0.002
	[-0.008;0.028]	[-0.011;0.008]	[-0.002;0.027]	[-0.010;0.006]	[-0.014;0.031]	[-0.011;0.009]	[-0.024;0.048]	[-0.006;0.004]	[-0.024;0.051]	[-0.007;0.004]
Upper Second	0.011*	-0.001	0.011*	-0.001	0.010	-0.002	0.009*	-0.001	0.011*	-0.002
	[-0.010;0.032]	[-0.008;0.006]	[-0.005;0.028]	[-0.009;0.006]	[-0.014;0.035]	[-0.009;0.006]	[-0.014;0.033]	[-0.009;0.006]	[-0.013;0.035]	[-0.010;0.006]
Lower Second	0.012	0.000	0.012	0.001	0.011	0.000	0.010	0.000	0.012	-0.001
	[-0.013;0.037]	[-0.013;0.013]	[-0.009;0.033]	[-0.013;0.014]	[-0.017;0.040]	[-0.012;0.012]	[-0.016;0.037]	[-0.012;0.011]	[-0.013;0.038]	[-0.011;0.010]
Other Classification	0.028	-0.005	0.027	-0.004	0.011	0.000	0.007	0.000	0.015	-0.004
	[-0.006;0.062]	[-0.017;0.007]	[0.002;0.053]	[-0.016;0.007]	[-0.037;0.059]	[-0.009;0.008]	[-0.036;0.050]	[-0.007;0.007]	[-0.026;0.056]	[-0.008;0.001]
High Priority (Relative to Baseline)										
First Class	0.004	-0.002	0.005	-0.001	0.003	-0.002	0.009	-0.00175***	-0.002	-0.002
	[-0.016;0.024]	[-0.010;0.006]	[-0.017;0.027]	[-0.010;0.008]	[-0.015;0.020]	[-0.010;0.006]	[-0.026;0.044]	[0.000533]	[-0.007;0.004]	[-0.008;0.003]
Upper Second	0.010**	-0.003**	0.011**	-0.003**	0.009**	-0.003**	0.008*	-0.00253***	-0.003**	-0.003***
	[-0.011;0.032]	[-0.007;0.001]	[-0.008;0.030]	[-0.006;0.001]	[-0.011;0.030]	[-0.007;0.001]	[-0.017;0.033]	[0.000367]	[-0.007;0.002]	[-0.007;0.002]
Lower Second	0.028**	-0.001	0.031**	-0.001	0.025**	-0.002	0.021**	-0.00168***	-0.002	-0.002*
	[-0.003;0.059]	[-0.007;0.005]	[0.004;0.058]	[-0.007;0.005]	[-0.000;0.050]	[-0.008;0.004]	[-0.006;0.048]	[0.000577]	[-0.007;0.003]	[-0.007;0.003]
Other Classification	0.095*	-0.005	0.094*	-0.005	0.046	-0.005	0.030	-0.00388***	-0.004	-0.004
	[0.054;0.136]	[-0.026;0.015]	[0.060;0.127]	[-0.025;0.015]	[-0.008;0.100]	[-0.024;0.014]	[-0.016;0.076]	[0.000743]	[-0.021;0.013]	[-0.023;0.014]
Individual Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,628,292		1,369,660		1,945,547		2,391,809		2,843,090	
Pseudo R-squared	0.136		0.129		0.135		0.134		0.130	

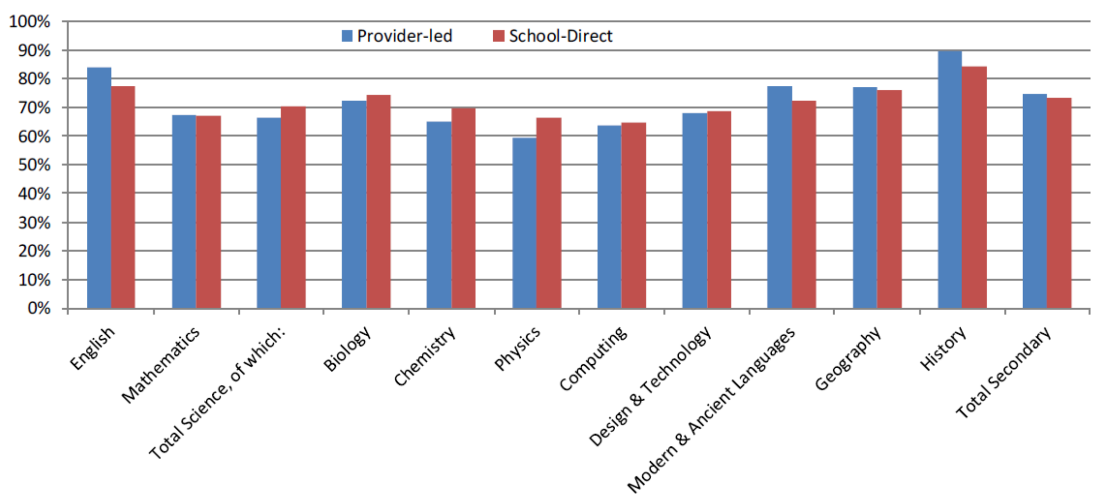
Notes: Author's calculations using HESA Destinations of Leavers from Higher Education. Trainee teachers defined as those undertaking postgraduate teacher training course at the survey date. Individual covariates include age, gender, region and subject fixed effects. 95% confidence intervals are shown in brackets and estimated using a wild cluster bootstrap at the subject level with Webb weights.

Figure B1: Numbers of teacher trainees by subject over time



Sources and Notes: Initial Teacher Training Statistics (2017-18, 2015-16 and 2014-15); Technology and Business includes Design and Technology, ICT, Computing and Business Studies; English includes Drama up to 2012-13

Figure B2: Proportion of trainees with a first or upper second class degree by subject and training route in 2015-16



Initial Teacher Training Statistics 2015-16

Appendix C - Appendix Tables and Figures for Chapter 4

Table C1: Estimates for individual local authorities in South and West Yorkshire

Local Authority	**Maths Estimate**	**Maths RMSPE**	**English Estimate**	**English RMSPE**
Barnsley	-0.044	0.038	-0.005	0.044
Doncaster	-0.112	0.035	-0.127	0.030
Rotherham	0.041	0.035	0.006	0.022
Sheffield	0.048	0.025	0.076	0.019
Bradford	0.019	0.033	0.026	0.047
Calderdale	-0.126	0.030	-0.057	0.029
Kirklees	0.022	0.031	0.008	0.032
Leeds	0.051	0.013	-0.021	0.019
Wakefield	-0.068	0.036	-0.036	0.045
Average	**-0.019**	**0.031**	**-0.015**	**0.032**

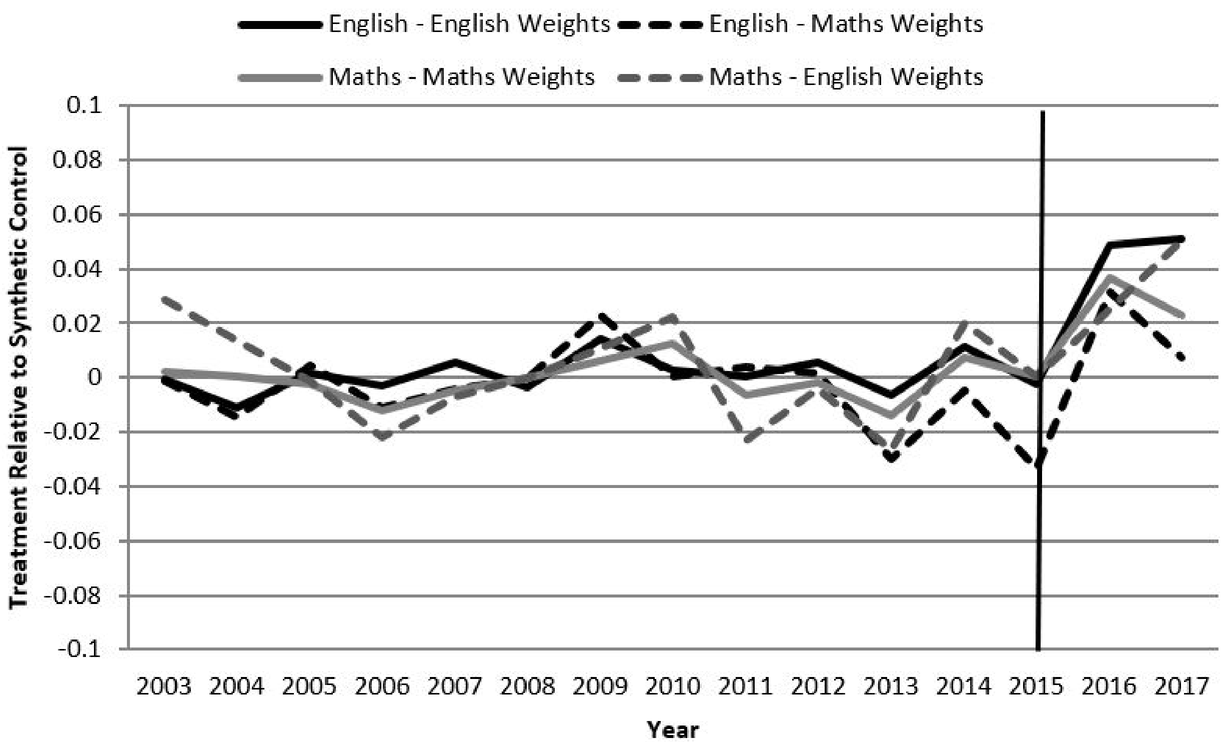
Notes and sources: Author's calculations using National Pupil Database 2002-03 to 2016-17. Synthetic control estimates use the same approach as the main estimates applied to individual local authorities within South and West Yorkshire.

Table C2: Leave one out estimated treatment effects and weights

Local Authority	**Weight**	**Estimated LOO Treatment Effect**
A) Maths		
Medway	0.271	0.027
Poole	0.255	0.008
Peterborough	0.108	0.011
Nottinghamshire	0.099	0.022
Oldham	0.091	0.013
Kingston Upon Hull, City of	0.064	0.020
Middlesbrough	0.055	0.015
Nottingham	0.038	0.027
Walsall	0.010	0.018
Dudley	0.006	0.020
Tameside	0.002	0.002
Knowsley	0.001	0.028
All	**1.00**	**0.019**
B) English		
Peterborough	0.227	0.035
Tameside	0.219	0.034
North East Lincolnshire	0.184	0.031
Knowsley	0.135	0.020
Middlesbrough	0.106	0.029
Rochdale	0.054	0.026
Nottingham	0.053	0.029
Walsall	0.023	0.041
All	**1.00**	**0.035**

Notes and sources: Author's calculations using National Pupil Database 2002-03 to 2016-17. Weight refers to the weight used for each local authority in constructing the main synthetic control estimates. LOO treatment effect refers to the estimated treatment effect if the respective local authority is excluded from the donor pool and the synthetic controls re-calculated.

Figure C1: Difference between treatment and synthetic controls, applying different weights



Notes and sources: Author's calculations using National Pupil Database (2002-03 to 2014-15) and School Workforce Census Statistics (2011 to 2015). Year refers to academic year starting each September. Vertical line indicates when treatment began in South and West Yorkshire. Solid lines show the main estimates, with dashed lines showing the difference between treatment and synthetic controls when using weights applied to the other primary outcomes (e.g. English outcome with weights estimated for Maths).