Essays on the Recruitment and Retention of Teachers

Sam Sims
UCL

Submitted for the degree of PhD
July 2018

Declaration

I, Sam Sims, 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.

Abstract

Teachers are among the most important school inputs for pupil attainment (Hanushek, 2011). Despite this, economically-advanced countries experience recurring shortages of teachers, resulting in sub-optimal hiring and deployment of teachers and reduced pupil attainment. This thesis investigates the determinants of entry to and exit from the teaching profession in order to understand how these shortages can be reduced.

Very little is known about the correlates of entry to the teaching profession. Non-cognitive skills and personality-type have been shown to be important predictors of occupational choice in general (Cobb-Clark & Tan, 2011; Nieken & Stormer, 2010). However, these have not been used to model entry to the teaching profession. Chapter 2 of this thesis uses rich data from a household panel survey to model entry to the profession. The model identifies groups of people who are up to four times more likely to enter teaching than the typical graduate. This information can be used to target recruitment efforts.

Retaining teachers is also important for ensuring sufficient supply. Research using administrative data generally finds the proportion of disadvantaged pupils in a school to be the strongest correlate of turnover. However, recent literature suggests that working conditions are important omitted variables in such analysis. Chapter 3 uses data on teachers from thirty-five countries to develop a rich set of working conditions measures and uses these to model teacher job satisfaction and intention to quit. The results highlight the importance of school leadership and assigning teachers to subjects in which they have been trained. Chapter 4 builds on this analysis by evaluating the impact of a subject-specific professional development intervention for science teacher retention. Double- and triple-difference models suggest that participation in the programme improves retention in the profession, though not in the participant's original school.

Impact Statement

The discoveries made in this PhD can be put to beneficial use in several ways. The Department for Education regularly states that teacher retention has been broadly stable for the last twenty years. Contrary to this, Chapter 1 demonstrates that the decline in early career retention is a major contributor to the current teacher shortage. I have recently written about this in the *tes* (Sims, 2018a). The Department's policy should shift to reflect this fact by placing a stronger emphasis on retaining more teachers, rather than just trying to increase recruitment. Chapter 3 identifies robust school-level correlates of teacher job satisfaction and turnover intentions. I am currently working with the Association of School and College Leaders (ASCL) to pilot a shortened version of the questionnaire used in Chapter 3 to help 14 schools self-assess the quality of working conditions they provide for their teachers. The aim is to help schools identify and address areas in which they have weaknesses. If the pilot proves a success, then ASCL plan to offer this service to their full membership in over 18,000 schools. Chapter 4 reports on an evaluation of the National STEM Learning Centre professional development courses for science teachers. The results from this evaluation are now being used by STEM Learning as part of the evidence base in their bid to have public funding for the programme renewed. This research was covered in the *Times*, the *i*, *tes* and *Schools Week*. Finally, Chapter 2 established that just four variables are able to identify individuals up to four times more likely to enter the teaching profession than the average graduate. Given the current shortage of teachers, this information can be used by policymakers to improve the targeting of marketing aimed at increasing recruitment.

Contents

List of Tables

List of Figures

Acknowledgements

Chapter 1: Introduction

Teachers have an important influence on their pupils. A one standard deviation (SD) increase in teacher quality is associated with a 0.1 to 0.2 SD increase in pupil attainment (Burgess, 2015). This is a strong relationship, at least when compared to other school level inputs. Indeed, Hanushek claims that *"no other attribute of schools comes close to having this much influence on student achievement"* (2011, p. 467). Good teachers improve equity as well as efficiency, because they have a disproportionately large impact on the attainment of poor pupils (Hamre & Pianta, 2005; Slater et al., 2012). Teachers also have an important influence on their pupils' wider socio-emotional skills (Jackson, 2016; Kraft, 2018). It is perhaps not surprising then that the effects of good teachers can be detected in their pupils' earnings long after leaving school (Chetty et al., 2014). Ensuring a sufficient supply of high quality teachers should therefore be a priority for education policymakers.

Despite this, shortages of qualified teachers are a widespread and recurring problem in advanced economies (Dolton, 2006). The 2013 Teaching and Learning International Survey (TALIS) collected data from a representative sample of school principals in over thirty countries. When these school leaders were asked about the biggest constraint they faced in improving the quality of instruction, the most common response was the lack of appropriately qualified staff (OECD, 2014). Faced with such shortages, school leaders are forced to lower recruitment standards, increase their use of temporary teachers or increase class sizes (Smithers & Robinson, 2000). This has prompted organisations including the UN, World Bank, OECD and the EU to warn that teacher recruitment efforts need to be stepped up (Figazzolo, 2012; Ranguelov et al., 2012; Schleicher, 2011).

This thesis aims to address the problem of teacher shortages by providing new evidence on why people enter and exit the teaching profession. In Chapter 2, I use household panel data to analyse long terms trends in the types of people who become teachers and then utilise the rich data to conduct detailed modelling of the reasons why people choose to become teachers. In Chapter 3, I switch focus to investigate teacher retention, utilising the TALIS survey data to develop a set of working conditions measures which allow me to model the determinants of teacher job satisfaction and desire to move school. Then in Chapter 4, I focus on identifying the causal impact of one important aspect of working conditions, professional development, on teacher retention. Chapter 5 concludes by summarising the findings of the research.

This thesis is part of a wider programme of research which I have been conducting on the teacher labour market, including work on socio-economic inequalities in access to good teachers (Allen & Sims, 2018a; 2018b), the quality of teacher working conditions across countries (Zieger et al., 2018), the effects of targeted salary supplements on early-career teacher retention (Sims, 2018b), and the

effect of leadership training (Knibbs et al., 2017) and school inspections (Sims, 2016) on teacher retention.

1.1 Teacher Shortages

Teacher shortages are a cyclical phenomenon. Long run cycles occur as demographic bulges, such as the post-war Baby Boom, work their way through the age distribution (Dolton, 2006). As these demographic bulges reach school age the number of pupils increases, thus increasing the demand for teachers; as they reach adulthood and enter the labour market, the supply of teachers increases (see Chapter 2); and as they reach retirement there can be sharp contractions in the supply of teachers. The post-war baby boom generation (born 1945-64), for example, began attending school in the fifties, began entering the labour market in the sixties and are now entering retirement (Callanan & Greenhaus, 2008; Chevalier & Dolton, 2004a). England and Wales experience peaks in the number of live births approximately once every twenty to thirty years (ONS, 2015).

As well as being influenced by demographic cycles, teacher shortages are also related to economic cycles. When recessions occur, employment opportunities in teaching are relatively prevalent. In the two years immediately following the 2008 recession for example, there was a 30% increase in the number of graduates choosing to train as teachers (Howson & McNamara, 2012). In periods of economic expansion however, relative wages in other occupations rise, incentivising graduates to choose occupations other than teaching (Chevalier et al., 2007). Demographic and economic cycles therefore conspire to ensure that teacher shortages are a recurring public policy challenge.

England is currently experiencing something of a perfect storm in terms of the demographics and economics of teacher supply. First, a demographic bulge is currently working its way through the secondary school-age population (DfE, 2016), with the result that there is forecast to be a 13% increase in pupil numbers between 2015 and 2024 (Lynch et al., 2016). Second, Baby Boom teachers have now largely retired, increasing the need for additional recruitment (Chevalier & Dolton 2004a). Third, England has recently experienced an improvement in graduate (un)employment (DfE, 2017a) and an increasing gap in median earnings between teachers and other professions in the majority of regions in England (STRB, 2017), drawing graduates away from teaching.

It is perhaps not surprising then that there is a growing shortage of teachers in England. The Department for Education (DfE) uses a sophisticated model to estimate the number of teachers it will need to train each year in order to ensure a sufficient supply of classroom teachers. On the supply side, the model takes into account all flows into and out of the profession, which vary depending on the state of the economy and the age profile of the teaching workforce, among other factors. On the demand side, the model takes into account the number of pupils of school age, determinants of which include the number of births and net migration patterns. The difference between these two numbers can then be used to calculate the number of teachers that need to be recruited each year in order to

balance supply and demand, within recent (though not necessarily optimal) pupil teacher ratios. Further information on how sufficient levels of recruitment are calculated can be found in DfE (2013). Figure 1 shows recruitment to initial teacher training (ITT) since 2010 as a proportion of the government's target, for both primary and secondary. It shows that the shortages are largely confined to secondary teachers and that there has only been an overall shortage of secondary teachers since the 2013/14 academic year. Measured in absolute terms, the secondary recruitment shortage in 2017/18 was 3,731.

Figure 1: Annual post-graduate ITT recruitment against targets
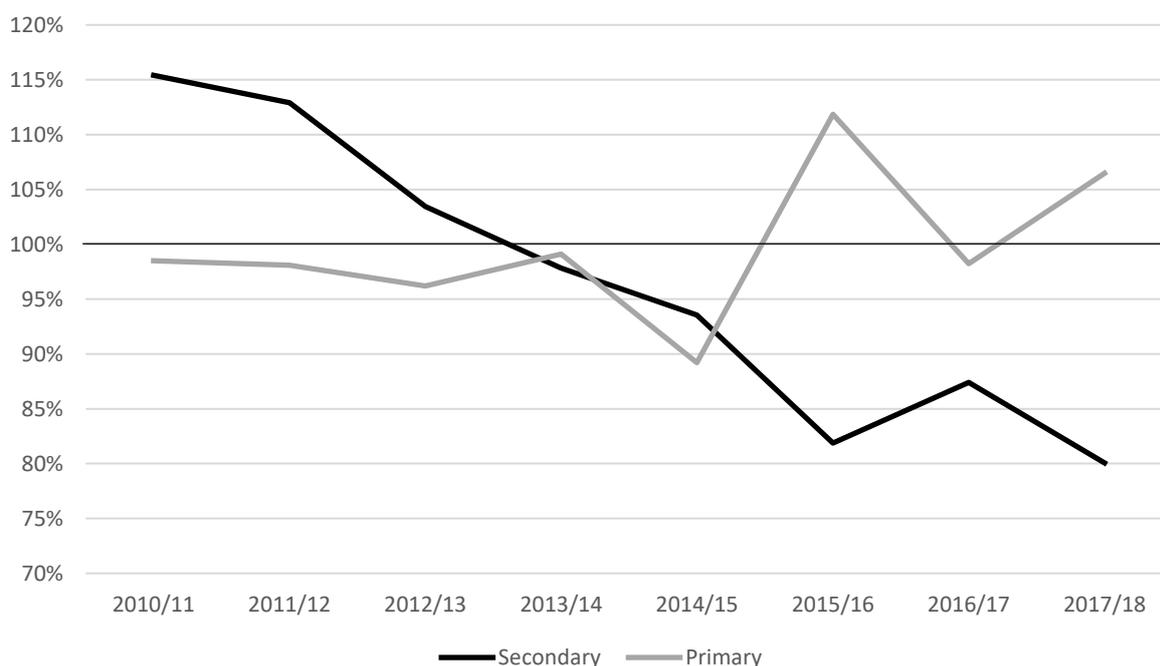
Notes: Source: Initial Teacher Training Census Main Tables. Post-grad ITT was 83% of ITT in 2017 (DfE, 2017d).

The picture is considerably more varied at subject level. Figure 2 shows the absolute teacher balance for the English Baccalaureate (EBACC) science, technology, engineering and mathematics (STEM) subjects. Figure 3 shows the same for EBACC Non-STEM subjects. The two are shown separately to improve visual clarity. The subject-specific estimates of sufficiency are calculated using the same method as those shown in Figure 1, but also taking into account pupil subject choices at GCSE. For further detail on how this is calculated see DfE (2013). The absolute teacher balance has been calculated by taking the subject-specific absolute recruitment shortage in each year and deflating it by the DfE's assumptions about the proportion of initial teacher trainees who take up jobs in state schools in England after qualification. A number above zero represents a surplus of teachers for a given subject and a number below zero indicates a shortage. Figure 2 and Figure 3 show that none of the EBACC subjects had a shortage of teachers in the 2010/11 or 2011 academic years. Mathematics, overall science and computing went into shortage in 2012; followed by MFL in 2013; geography,

biology and physics in 2014; chemistry in 2015; then English in 2016. By 2017, the only EBACC subject not in shortage was history. Among non-EBACC subjects (not shown here for space reasons) art, music, religious studies and business are also in shortage. Indeed, physical education is the only non-EBACC subject included in the data which was not in shortage in 2017.

Figure 2: Teacher balance by subject (STEM) 2010-17
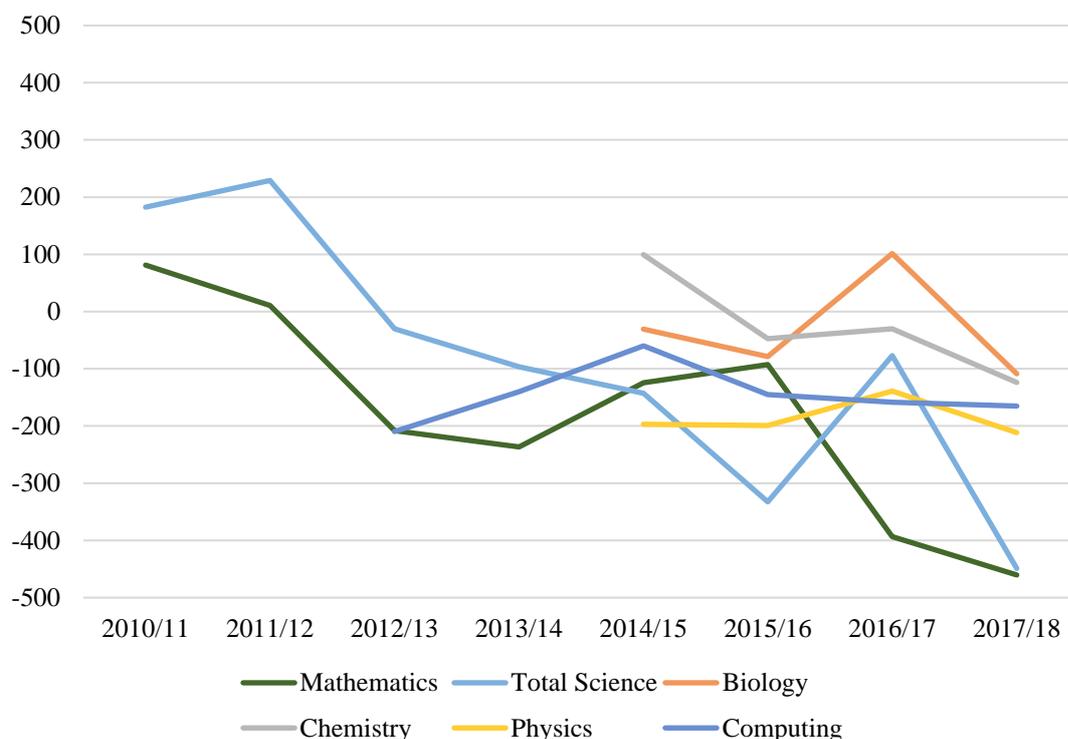
Notes: Source: Initial Teacher Training Census Main Tables and Teacher Supply Model. Figures for computing first reported for 2012/13. Figures for the three sciences first reported separately for 2014.

Despite being foreseeable based on demographic and economic trends, the current teacher shortage has an additional, less predictable component: declining early-career retention. Figure 4 shows retention of each cohort of newly qualified teachers (NQT) in England. The 2005 to 2009 cohorts follow a broadly similar trajectory. From 2010 onwards however, there is a clear decline in retention, with each new cohort having a steeper downward gradient than the last. Early-career retention is worse among shortage-subject teachers such as scientists (Kelly, 2004; Worth & De Lazzari, 2017), perhaps because they face higher outside pay ratios (MAC, 2016). Early-career retention is also worse among certain initial teacher training routes such as Teach First, which has expanded over this period (Allen et al., 2013). This decline in early-career retention is materially significant. As shown in Table 14 in Appendix A, if early-career retention rates had been frozen at 2009 levels, there would now be an additional 4,398 teachers working in England. To put this figure in context, the total shortfall of EBACC teachers is currently 2,080. Declining early-career teacher retention is therefore a major contributor to the current teacher shortage.

Figure 3: Teacher balance by subject (Non STEM) 2010-17
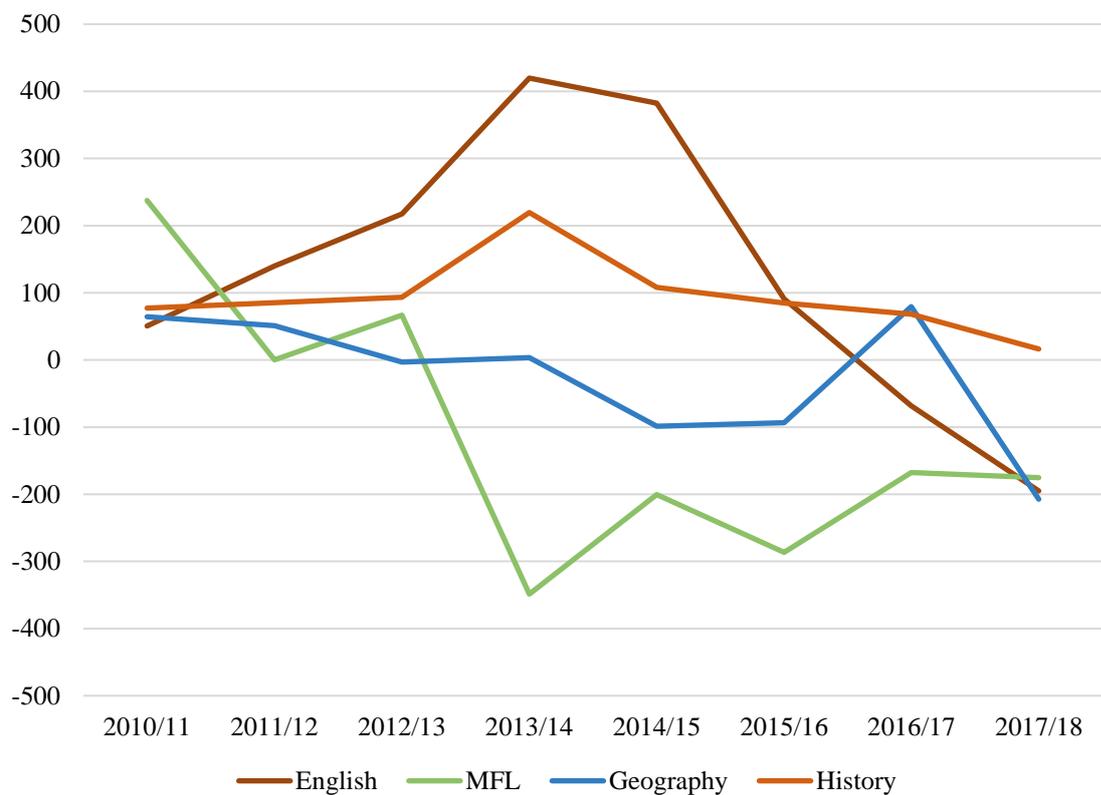
Notes: Source: Initial Teacher Training Census Main Tables (see: https://www.gov.uk/government/collections/statistics-teacher-training) and DfE(2013).

Figure 4: Proportion of NQT cohort retained
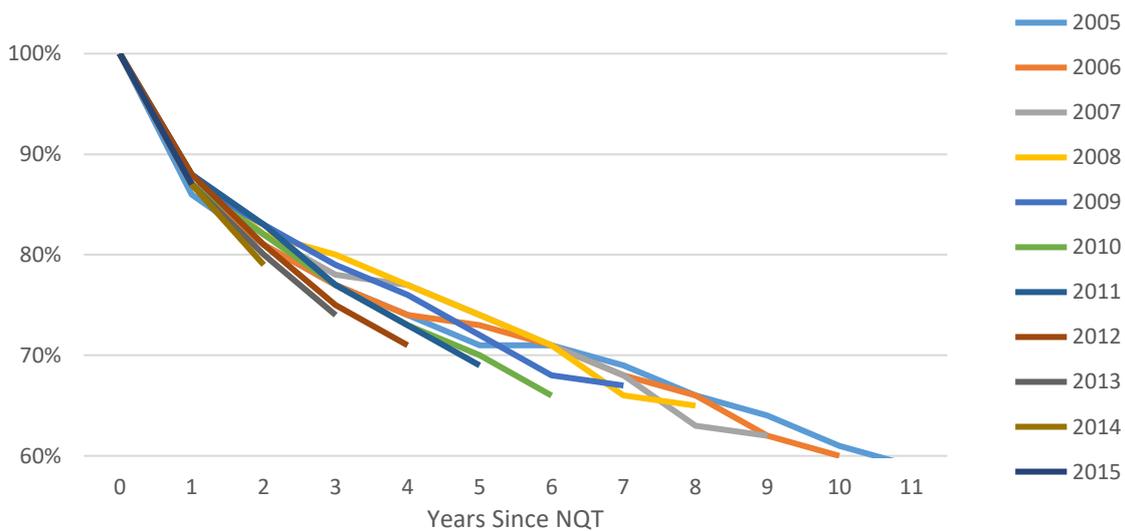
Notes: Source: School Workforce in England 2017 (see: https://www.gov.uk/government/statistics/school-workforce-in-england-november-2017)

1.2 Teacher Supply Policy

Policymakers have responded to these shortages by offering a range of financial incentives for teachers. Scholarships and bursaries have been introduced which award teachers a tax-free sum of money to support them through initial teacher training. The value of these incentives changes from year to year and depends on a trainee's degree class and degree subject. Among EBACC subjects in 2017 the value of these payments ranges from £4,000 for those with a 2:1 in history, up to £28,000 for those awarded a scholarship (requiring a 2:1) in physics, chemistry, languages, computing or geography.[1]

In recognition of the problem of decling early-career retention, policymakers have recently supplemented these recruitment incentives with retention incentives. In October 2017, the government announced two new policies in this area. The first involves reimbursing science and modern foreign language teachers' student loan repayments, which will be worth around £540 per year for a teacher in the fifth year of their career.[2] The second involves paying maths teachers who train in the 2018-19 academic year £5,000 in both the third and fifth years of their career, conditional on uninterrupted service in state funded schools in England after qualification. Those working in a list of priority areas are eligible for an additional £2,500 at each payment point (DfE, 2017b). Evidence from evaluations of three similar policies in the USA (Bueno & Sass, 2016; Clotfelter et al., 2008; Feng & Sass, 2016) suggests that teachers do respond to these sorts of incentives. Simulations applying the estimates from these evaluations to data on teachers in England suggest that the maths retention bonuses in England are large enough to substantially reduce, if not eliminate shortages in that subject (Sims, 2018b). Unfortunately, many of these policies were introduced too late to be evaluated in this PhD.

A second strand of government policy has focused on the non-pecuniary aspects of teachers' jobs, specifically their satisfaction with their working lives. This has primarily been motivated by widespread complaints about increasing workload in the profession. Workload surveys show that teachers' self-reported workload has increased by five hours per week over the past five years (Deakin et al., 2010; Higton et al., 2017). Full time secondary teachers in England now report spending an average of 55.3 hours per week working during term time, with 26% working 60 hours or more. On average, 17.4 hours of teachers' working time was reported as occurring during evenings and weekends, or otherwise outside of school time. Perhaps more important than the total number of hours spent at work is the composition of this workload. Of the 55.3 hours per week, only 20.9 hours were spent teaching – which is comparable to the average among TALIS countries (Higton et al., 2017; Micklewright et al., 2014). The additional workload in England is due instead to the 21 hours per week spent on planning, marking and admin – which is well above the TALIS average (Higton et al.,

---

[1] https://getintoteaching.education.gov.uk/funding-and-salary/overview accessed 06/12/2017
[2] https://www.gov.uk/guidance/teachers-student-loan-reimbursement-guidance-for-teachers-and-schools accessed 06/12/2017

2017; Micklewright et al., 2014). It is these sorts of administrative tasks and the overly detailed, bureaucratic nature in which they are conducted which teachers identify as being the aspects of workload that they like least (Gibson et al., 2015). This research has prompted policymakers to work with unions and school leaders to try to reduce unnecessary workload in schools (DfE, 2017c).

1.3 Working Conditions

Workload is however only one aspect of the working conditions which influence teachers' decisions about whether to remain in the profession. For years, researchers using administrative datasets found that pupil characteristics, particularly the deprivation of a school's intake, were the best predictor of high teacher turnover in schools (Boyd et al., 2005; Hanushek et al., 2004; Scafidi et al., 2007; Allen et al., 2018). However, following pioneering research by Eileen Weiss (1999) and Richard Ingersoll (2001), a new wave of research utilising survey data has demonstrated that the association between pupil deprivation and teacher retention is largely eliminated when measures of working conditions are included in the models (Simon & Johnson, 2015). The best articles in this literature utilise linked survey and administrative data and use factor analysis to identify latent variables measuring working conditions in schools (Boyd et al., 2010; Johnson et al., 2012; Ladd 2011). These studies generally find that the social aspects of school life, such as leadership/management and staff collaboration, are the best predictors of teacher retention (Simon & Johnson, 2015). Interestingly, two studies which analyse experienced teachers and early-career teachers separately found that working conditions have a stronger influence on the retention of those that have recently qualified (Boyd et al., 2011; Kukla-Acavedo, 2009), making working conditions a candidate for explaining the recent decline in early career retention.

Since 2015, two important extensions have been made to this literature. First, Kraft et al. (2016) have conducted the first analysis of working conditions using school panel data, allowing them to get closer to a causal analysis. Echoing previous research, they find that the nature of leadership and teacher relationships are strongly associated with retention. The robust relationship between leadership/management and retention across this literature begs questions about which specific leadership practices matter most. The second development in the literature has begun to address this question. Bloom et al. (2015) use data from the World Management Survey, which measures the use of evidence-based management practices in 1,800 schools across eight countries. Although they study pupil attainment as their outcome, their findings are largely consistent with those from the education literature in that "people management" practices - such as rewarding high performers and managing talent - are the strongest predictors of school performance. Fryer (2017) reports on a field trial in which school principals were given intensive training in a related set of practices – including teacher observation and coaching – and find a positive causal effect on attainment. The latest studies on

working conditions therefore suggest that working conditions really matter, both for teachers and their pupils.

1.4 This Thesis

In Chapter 2 of this thesis, I take a detailed look at entry into the teaching profession. Existing research in this area tends to be limited to interviews or surveys with either in-service teachers or trainees (Matthias, 2014; Richardson et al., 2006; Roness & Smith, 2009; Roness & Smith, 2010; Sinclair, 2008; Watt & Richardson, 2007; Watt et al., 2012). Research using objective measures and following cohorts from education into the labour market is comparatively rare, perhaps because of the demanding data requirements for conducting such a study. I am aware of only four papers that conduct such an analysis for teachers (Bacolod, 2007b; Chevalier et al., 2007; Goldhaber & Liu, 2003; Reback, 2004). All four use graduate cohort surveys, which have the advantage of following individuals into the labour market, allowing a comparison of the characteristics of those who do and do not choose to become teachers. However graduate cohort datasets are also limited by the narrow set of variables that they tend to include. In particular, they do not include a range of variables such as personality type, self-efficacy, social networks, and values which have been shown to predict occupational choice (Bentolila et al., 2010; Borghans et al., 2008; Cobb-Clark & Tan, 2010; Filer, 1986; Krueger & Schkade, 2008; Lyons et al., 2006; Nieken & Stormer, 2010; Quimby & Santis, 2006).

Chapter 2 provides new evidence on this by analysing rich data on entry to the teaching profession in the UK stretching back to 1938. It begins by documenting a number of new descriptive findings: the proportion of ethnic minority teachers has risen with each new generation of entrants, but the conditional odds of somebody from an ethnic minority entering the profession have fallen substantially over the same period; the proportion of teachers whose parents also taught has steadily increased across four generations; and the well-known twentieth century trend towards a less female, more graduate workforce has levelled off among the most recent generation of teachers. I also model entry to the profession and show for the first time that personality type, particularly openness to new experience, is a very strong predictor of entry to the teaching profession. Using these models, I am able to identify groups of people who are up to four times more likely to enter the teaching profession than the average graduate. The findings of this study can be used to target future teacher recruitment efforts.

In Chapter 3, I return to the issue of working conditions. The existing research using survey data to model teacher retention is limited by the breadth of the questionnaires used to measure working conditions, which often contain around 20 questions (Boyd et al., 2011; Ladd, 2011; Johnson et al. 2012; Kraft et al., 2016). Moreover, the data they use in these studies covers only a single US state in each case. I extend this literature by modelling teacher job satisfaction and desire to move school

across all 35 countries for which there is usable data in TALIS 2013. I also utilise the additional working conditions measures included in the England-specific survey (48 items in total) and use factor analysis to derive a more comprehensive set of working conditions measures. I am also able to link this survey data to school-level administrative data on pupil intake characteristics.

The findings from Chapter 3 corroborate those reviewed by Simon & Johnson (2015) in that, conditional on working conditions, pupil disadvantage is not predictive of teacher job satisfaction or desire to move school. In the international sample, teacher cooperation and professional development are predictive of both outcomes and these results hold in almost all of the 35 national sub-samples. In the England-only models however, when I add five more working conditions derived from the factor analysis, neither of these associations remain statistically significant. The nature of school leadership and preparation for teaching assignments (Donaldson & Johnson, 2010) are instead the best predictors of both job satisfaction and desire to move school. These findings are robust to checks for common method bias, including measuring the independent variable using responses from teachers' departmental colleagues. These findings have the potential to inform the practice of school leaders facing staffing difficulties in their schools.

In Chapter 4, I focus on one potentially important aspect of working conditions: professional development. Detailed qualitative research has been used to construct theoretical arguments that professional development can improve retention by enhancing teachers' knowledge, self-efficacy and motivation (Coldwell, 2017; Taylor et al., 2011; Wolstenholme et al., 2012). Some quantitative evidence exists for the links between these variables. Subject-specific professional development has been shown to improve teachers' knowledge of their subject (Goldschmidt & Phelps, 2010; Kutaka et al., 2017; Polly et al., 2015; van Driel et al., 2012) and increase the self-efficacy of both science (Lakshmanan et al., 2011) and maths teachers (Ross & Bruce, 2007). However evaluation research has found mixed results on retention, with one observational study finding that intensive CPD does help keep experienced teachers in the profession (Knibbs et al., 2017) but one randomised controlled trial finding that intensive professional development for new teachers had no effect on retention (Glazerman et al., 2010).

Chapter 4 provides new evidence on this point by evaluating the effect of science teachers attending National STEM Learning Network (NSLN) courses in England. The NSLN CPD is distinctive in that it is aimed specifically at science teachers and focuses on providing participants with both enhanced subject knowledge and closely linked pedagogical knowledge about how to teach this knowledge in the classroom. I use double- and triple-difference models to try and identify the causal impact of the programme on retention in participants' original school and in the profession overall. I find that participation is associated with a three percentage point reduction in departmental wastage rates two years after first observed participation, which is equivalent to around a third of average wastage in science departments. This finding is robust across several model specifications and survives a placebo test. Despite increasing retention in the profession, participation is not associated with retention in

participants' original school. This research represents some of the first credible findings that subject-specific CPD has a positive causal impact on teacher retention.

Chapter 2: Entry to the Teaching Profession

2.1 Introduction

Teacher shortages are prevalent in public school systems around the world (Dolton, 2006). In the thirty-three countries and regions included in the main report on the Teaching and Learning International Survey 2013, over a third of teachers work in schools in which the head teacher reported significant difficulties with recruitment (OECD, 2014). Indeed, staff shortages were on average cited by head teachers as the most important constraint on good quality teaching. In recent years, demographic trends and policy changes have contributed to an increase in the severity of these shortages in many countries, prompting organisations including the UN, World Bank, OECD and the EU to warn that recruitment efforts need to be stepped up (Figazzolo, 2012; World Bank, 2013; Schleicher, 2011; Ranguelov et al., 2012).

In England, the context for this study, shortages are also a growing problem. The Department for Education (DfE) uses its Teacher Supply Model to estimate the number of teachers required in the English school system each year. On the supply side, the model takes into account all flows into and out of the profession, which vary depending on the state of the economy and the age profile of the teaching workforce, among other factors. On the demand side, the model takes into account the number of pupils of school age, determinants of which include the number of births and net migration patterns. The difference between these two numbers can then be used to calculate the number of teachers who need to be recruited each year in order to balance supply and need, within current class size limits. As the pupil population has begun to grow in recent years however, the DfE has been missing these recruitment targets by increasing amounts: 1% in 2012/13, 5% in 2013/2014 and 9% in 2014/2015 (NAO, 2016).

School leaders tend to respond to teacher shortages by either lowering recruitment standards, increasing use of temporary teachers or increasing class sizes (Smithers & Robinson, 2000), all of which have been linked with reduced pupil attainment (Fredriksson et al., 2012; Mocetti, 2012; Schanzenbach, 2006). There is also some evidence to suggest that pupils from poorer families are most likely to suffer from such sub-optimal staffing arrangements (Allen & Sims, 2018a; Clotfelter et al., 2006). Governments have responded to these shortages with policies designed to attract and retain teachers. In England, for example, the Department for Education currently spends £167m per year on bursaries to try and incentivise people to train as teachers and a further £16.6m on marketing campaigns (NAO, 2016). However, a lack of knowledge about which types of people have the highest propensity to enter teaching (Goldhaber et al., 2014a) has hampered policymakers' efforts to target such policies.

Existing research on the determinants of entering the profession largely uses interviews (Matthias, 2014) or surveys with existing teachers or trainees (Müller et al., 2009; Richardson et al., 2006; Roness & Smith, 2009; Roness & Smith, 2010; Sinclair, 2008; Watt et al., 2012; Watt & Richardson, 2007). However, such studies provide limited information about why people begin teaching, since they lack any comparison group capable of providing a counterfactual and likely suffer from social-desirability (Krumpal, 2013) and attrition bias. The analysis of objective information about teachers before they enter the profession has been hindered by the fact that administrative datasets only contain information on teachers from the point at which they joined the profession. To my knowledge, only four studies manage to overcome these problems. Goldhaber and Lui (2003) study the correlates of applying for a teacher training course and find that being female, from a lower income family, having a mother who is a teacher, having a non-STEM major, and having a higher undergraduate GPA are all associated with an increased chance of becoming a teacher. Reback (2004) adds to this the findings that those with parents without post-graduate degrees are also more likely to enter teaching and Bacolod (2007b) finds that being married, living in a state with lower proportions of poor pupils, lower proportions of minority pupils, lower minority/majority pupil ratios, or higher relative teacher wages all increase the chances of going into teaching. Chevalier et al. (2007) replicate many of these findings using data from the UK.

Goldhaber & Liu (2003), Reback (2004) and Bacolod (2007b) use the Baccalaureate and Beyond graduate cohort study data from the US, while Chevalier et al. (2007) use the UK Graduate Cohort Studies. The major benefit of graduate cohort studies is that they collect information both before and after entry into the labour market and, by extension, teaching. This allows prospective data collection from before the point that many people decide to enter teaching and enables comparison of those that do and do not go into teaching. However, graduate cohort studies also have important limitations, in that they tend to collect only a narrow range of variables. Crucially, neither Baccalaureate and Beyond nor the Graduate Cohort Studies contain information on variables that have been linked with job choice in the occupational choice literature, including: personality type (Nieken & Stormer, 2010; Cobb-Clark & Tan, 2010); self-efficacy or locus of control (Cobb-Clarke & Tan, 2010); social interactions (Borghans et al., 2008; Krueger & Schkade, 2008) and job values/preferences (Filer, 1986; Lyons et al., 2006). The studies reviewed above are therefore likely to suffer from omitted variable bias.

I use exceptionally rich data on entry to the teaching profession stretching back to 1938 to analyse the changing patterns of entry to the teaching profession in the UK. This allows me to document a number of new stylised facts. First, the proportion of ethnic minority teachers has risen with each new generation of entrants, reflecting the increase in the proportion of ethnic minorities in the population as a whole. At the same time however, the odds of ethnic minorities entering teaching have fallen, conditional on other characteristics. Second, the proportion of teachers whose parents also taught has

steadily increased across four generations. Third, the well-known twentieth century trend towards a more gender-balanced workforce has halted among the most recent generation of teachers. The current stock of teachers have relatively high job- and life-satisfaction relative to comparable occupational groups and graduates in general, but are also relatively dissatisfied with their leisure time compared to other graduates. I model entry to teaching and provide what is, to my knowledge, the first evidence on how personality types affect entry to the teaching profession. In particular, the openness to new experience dimension of personality type has a strong and robust association with entry to the profession. Using these models, I am able to identify groups of people who are up to four times more likely to enter the teaching profession than the average graduate. The findings of this paper can help policymakers refine their approach to teacher recruitment by targeting those with a high propensity to enter the profession.

2.2 Data

This study utilises the UK Household Longitudinal Study, also known as Understanding Society, or USoc (University of Essex. Institute for Social and Economic Research, NatCen Social Research and Kantar Public, 2009-2015). USoc is a panel of 40,000 UK households, beginning in 2009. Households are sampled using a clustered, stratified sampling design. Cross-sectional and longitudinal weights are supplied with the data in adjust for this complex survey design, as well as non-response, order to make the sample data representative of the UK population at the time of data collection. The analysis in this chapter is cross-sectional, in the sense that I never analyse more than one observation per individual. I also want to avoid dropping individuals who failed to respond in just one or two of the six waves (longitudinal weights are non-zero for those who have responded in all six waves of data collection). I therefore use the relevant cross-sectional weights at each stage of the analysis. Unfortunately, weights are not provided for combinations of waves due to the very high number of possible combinations of non-response (Lynn & Kaminska, 2010). When I am using data from across waves in my model I therefore use the Wave 1 weights. This is not ideal in the sense that adjustments for non-response should be wave specific. This will mean that data drawn from Waves 2-6, such as self-efficacy, will not be entirely representative of the target population. Due to the complexity of the survey design however, there is little alternative in this case.

Data is collected through face to face interviews with all adult members of a household, conducted in the home once every twenty-four months. As of 2017, six waves of data were available for analysis and household response rates for each wave are around 60%. The survey collects very rich information on all participants including a range of demographic, economic and psychological measures. Retrospective information is also collected on all adults in the original sample, which

allows me to observe data before labour market entry for both those who become teachers and those who do not. I also utilise data on UK over-sixteen unemployment rates stretching back to 1971 taken from Office for National Statistics.

The dependent variable for this study, entry to the teaching profession, is measured in USoc through the collection of partial employment histories. All participants are asked about their first job after leaving full time education and their current job at each data collection wave. Participants who left their first job before the first data collection wave (2009/10) therefore have an interval in which their occupation is censored. For example, somebody who started their first job after fulltime education in 2000 would have this job recorded in the data (because it is their first job) and would have their job in 2009/10 recorded in the data (because this is the year in which data collection began). If their job in 2009/10 was their second job after fulltime education, they would have a full employment history recorded in the data. However, if they had a second job which finished before 2009/10 then this would be censored, since it was neither their first job after fulltime education nor during the data collection period.

Information about the specific type of job is recorded as free text and then converted to ONS Standard Occupational Codes (SOC) using a computer assisted package. I define a teacher as anyone who has the SOC code 2314 (Secondary education teaching professionals), 2315 (Primary and nursery education teaching professionals), or 2316 (Special needs education teaching professionals). This definition includes assistant or deputy heads but does not include head teachers, executive heads, local authority or academy chain staff or inspectors. Table 15 (Appendix B) contains counts of people who fit this definition across different waves of the dataset. The USoc data contains 2,458 people who can be identified as having worked in teaching at some point during their careers, evenly split between secondary and primary/nursery teachers. Of these, 1,761 have worked as teachers during the data collection period (2009-2016) and 1,123 worked as teachers in their first job after full time education. Only 234 teachers taught for their first job after full time education and were still teaching in the most recent available wave of the data (collected in 2014-15).

Figure 5 shows the historical reach of the dataset. The left panel of the graph shows the frequency distribution of the year in which USoc panel members can be first identified as entering the teaching profession. The first panel member to enter teaching did so in 1938, there is then a peak around 1970, as the so-called Baby Boomer generation enter the labour market, and a second peak around the turn of the century. Table 15 (Appendix B) shows that 56 teachers can be identified as entering the profession between 1935 and 1965, 439 between 1955 and 1974, 358 between 1975 and 1994, and another 263 since 1995. The right panel of the graph shows the same information but restricted to those who are observed teaching between Wave 1 (2009/10) and Wave 6 (2014/15) of data collection.

I refer to these as W1-W6 teachers. The first W1-W6 teacher doesn't enter the profession until the 1960s.

Figure 5: Histograms showing year of entry to the teaching profession
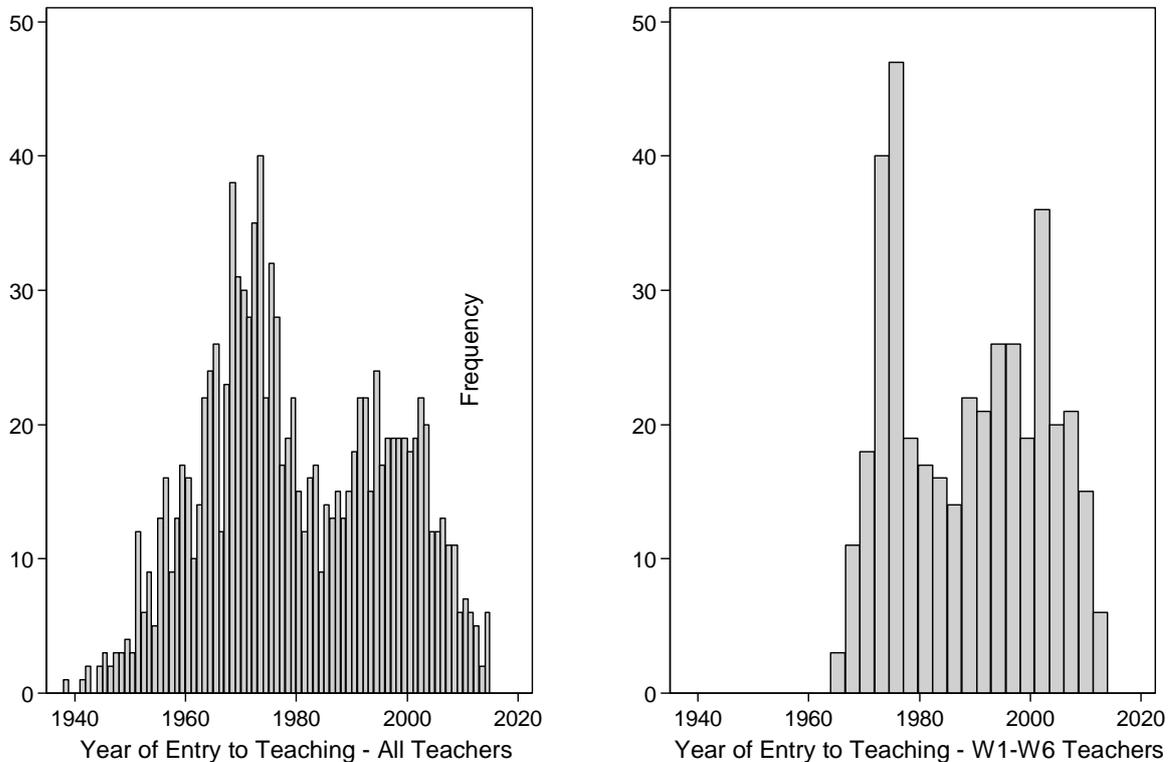
Notes: The left-hand side histogram shows the year in which each teacher can first be identified as entering the teaching profession. The right-hand side diagram shows the same but only for those who have taught between Wave 1 and Wave 6 (W1-W6) of the data (2009-14).

2.3 Comparing teachers with other occupational groups

This section presents detailed descriptive statistics on teachers compared to other professions, in order to place entry to the profession into comparative perspective. Table 1 compares time-invariant characteristics of all those who have ever been teachers in the USoc data to all those that have ever been nurses[3], ever been a government/administrative officer[4], or ever been a protective officer[5]. These three occupations were chosen as comparison groups because global surveys show that they are considered to be of comparable social status by the general public (Dolton & Marcenaro-Gutierrez, 2013) and because they are relatively large occupational groups, which enables more precise

---

[3] Standard Occupational Code (2010) 2231.
[4] Civil servants below grade 5 and their private/charitable equivalents. Standard Occupational Codes (2010) 4112, 4113, and 4114.
[5] NCOs, police officers (sergeant and below), fire service officers (watch manager and below), prison service officers (below principal officer). Standard Occupational Codes (2010) 3311, 3312, 3313 and 3314.

estimates of population characteristics. For completeness, I also show descriptive statistics for all other occupations in the final column. The teachers are 71% female, which is lower than nurses (91%) but higher than government/administrative (60%), protective officers (13%) and other occupational groups (50%). The teachers are 16% ethnic minority and 10% born overseas, which is lower than nurses (22%, 18%) and other occupations (20%, 12%) but higher than government/administrative workers (13%, 7%). Teachers are far more likely to have a degree (72%) than nurses (31%), government/administrative occupations (32%), protective officers (19%) and other occupations (19%).

Table 1: Comparing all teachers with other occupational groups

	Ever a Teacher	Ever a Nurse	Ever a Govt/ Admin	Ever a Protective Officer	All Others
Female (% Cross Wave)	70.6	90.7**	59.9**	12.8**	49.8**
Ethnic Minority (% Wave 1)	15.9	21.6**	12.6*	15.2	20.3**
Born Overseas (% Wave 1)	10.3	18.0**	7.2*	8.5	12.2*
Has a Degree (% Wave 1)	71.9	31.1**	31.8**	18.5**	18.9**
Father has a Degree (% Cross Wave)	24.5	23.1	23.5	20.7	15.7**
Number in Occupational Group	2,485	1,623	913	725	109,614

Notes: N=50,994 (At Wave 1). All Others includes those who have never been a teacher, a nurse, or a government official. Government officials does not include senior civil servants (Grade 5 and above). ** = p< 0.01, ** = p< 0.05.

Many of the variables collected in USoc, such as income, are time-varying and are therefore best compared amongst those teaching (or in comparable professions) during the period the data was collected. Table 2 shows how these characteristics compare between teachers, the three comparison occupation groups, and all other graduates, at Wave 1 (2009/10). Looking across the five groups, teachers are less likely to have permanent contracts than the three comparison occupation groups, but more likely to have a permanent contract than all other graduates. Teachers have higher net income than nurses and govt/admins, but not protective officers or all other graduates. Teachers spend less time commuting to work than the other groups, except for nurses. USoc measures job, income, leisure and overall life satisfaction on a seven-point scale stretching from "Completely dissatisfied" to "Completely satisfied". Teachers have higher job and income satisfaction than nurses, govt/admins and other graduates, but not protective officers. Teachers have comparable satisfaction with leisure time to nurses and govt/admins but worse than protective officers and other graduates. Finally, teachers have higher overall satisfaction than govt/admins and all other graduates, but similar to nurses and protective officers. Because these satisfaction measures are Z transformed, the results depend on the assumption that the seven responses are evenly spaced. In Table 16 (Appendix B) I

report the proportions of each professional group reporting that they are at least "Mostly Satisfied" on each measure, which does not rely on this assumption. The relative satisfaction of teachers, nurses, govt/admins and protective officers are qualitatively very similar.

Table 2: Comparing W1 teachers with other occupational groups

	Teachers	Nurses	Govt/ Admins	Protective Officers	Other Graduates
Permanent Contract (%)	88%	95%**	96%**	99%**	66%**
Net Personal Income (£/Month)	1,950	1,667**	1,389**	1,963	1,939
Travel to Work Time (Mins)	23	24	27**	30**	34**
Satisfaction Job (Z Score)	0.13	0.00*	-0.13**	0.10	0.01**
Satisfaction Income (Z Score)	0.33	0.17**	0.01**	0.28	0.22**
Satisfaction Leisure Time (Z Score)	-0.17	-0.12	-0.20	-0.03*	-0.04**
Satisfaction Overall (Z Score)	0.18	0.15	-0.01**	0.15	0.10*
No. in Occupation Group (W1)	725	604	313	278	10,118

Notes: N=50,994 (At Wave 1). All Others includes those who have never been a teacher, a nurse, or a government official. Government officials does not include senior civil servants (Grade 5 and above). ** = p< 0.01, * = p< 0.05. Standard errors are not clustered by workplace, since this is not observable in the data. However the number of, e.g. teachers is very small compared to the number of schools, making it unlikely that any USoc teachers work in the same school.

As well as providing information about entry to the teaching profession over the long run, USoc also includes four sets of variables which have been shown to predict occupational choice. The first of these is personality type. Aspects of personality type, particularly openness to new experience, have previously been shown to be associated with being a teacher (Nieken & Stormer, 2010; Cobb-Clark & Tan, 2010). USoc includes measures of personality type collected in Wave 3 (2011-2012) using an adapted version of the Big Five Inventory (John & Srivastara, 1999) measured on a seven-point scale. This instrument has been developed to yield distinct dimensions of personality and I confirm this by showing the low pairwise correlations between the five dimensions in Table 16B (Appendix B). Personality type is not fixed over time (Hampson & Goldberg, 2006) meaning that measures of personality collected in 2011/12 cannot confidently be used to explain entry to the teaching profession in, e.g. the 1980s. This measure can therefore only be used for those who are observed entering the teaching profession during or immediately before the data collection period (2009-2015).

The second set of USoc variables related to occupational choice measure self-efficacy. This is defined as a person's belief in their ability to succeed and has previously been linked to entering the teaching profession (Cobb-Clarke & Tan, 2010). USoc collected measures of self-efficacy in Wave 5 (2010-2011) using the 10-item version of the Generalized Self-Efficacy Scale (Schwarzer & Jerusalem,

1995). Self-efficacy also changes over time, meaning that it can only be used to model entry to the profession during or shortly before the data collection period.

The third set of variables included in Usoc relevant to occupational choice relate to social networks. People who are more sociable in their private lives are more likely to choose jobs which involve dealing with people, such as teaching, counselling, advising or caring for others (Borghans et al., 2008; Krueger & Schkade, 2008). USoc includes data on social networks collected in Wave 3 (2011-2012), in particular a question asking people how many close friends they have.

Finally, USoc collects information on respondents' preferences for job attributes, which have unsurprisingly been shown to predict occupational choice (Filer, 1986; Lyons et al., 2006). This was measured through a set of questions which asked respondents to rate the importance of seven job attributes – job security, income, leisure time, interesting tasks, contribution to society, time for family, and opportunities to help others – measured on a four-point scale from "Not at all important" to "Very important". These questions were asked to a subsample of 10% of respondents in each of waves 2 (2010-11), 3 (2011-12) and 5 (2013-14).

Table 3 shows the mean characteristics across the same five groups used in Table 2 and across graduates in general. It replicates the findings from existing literature that those currently working in teaching tend to have personality types that display greater openness to new experience – which manifests in curiosity and the pursuit of variety - than other occupational groups. The table also shows that teachers have personality types that display greater neuroticism - which manifests in anxious and self-conscious behaviour - than nurses, protective officers or graduates in general. Teachers also report having more close friends than protective officer and other graduates. The job preference characteristics were, by design, only asked to a sub-sample of 10% of USoc panel members, meaning there are larger standard errors. Despite this, teachers express: stronger preferences for a job which contributes to society than govt/admins; and a stronger preference for helping others than both govt/admins and graduates in general. Teachers report higher levels of self-efficacy than both nurses and govt/admins.

Table 3: Comparing covariates of W1-W6 teachers with those in other occupational groups

	Teachers	Nurses	Govt/ Admins	Protective Officers	Other Graduates
Personality (Z Score)					
Agreeableness	0.24	0.16	0.24	0.18	0.10**
Conscientiousness	0.23	0.19	0.21	0.25	0.13**
Extraversion	0.25	0.15*	0.19	0.26	0.11**
Neuroticism	0.25	0.13**	0.26	0.12*	0.11**
Openness	0.34	0.14**	0.23*	0.24*	0.23**
Social networks (Z Score)					
Number close friends	0.19	0.10	0.08	-0.15**	0.09**
Job preferences (Z Score)					
Job security	0.20	0.29	0.41	0.12	0.18
High income	-0.48	-0.30	-0.20	-0.12	-0.27
High leisure time	-0.02	-0.16	0.15	-0.44	0.22
Interesting tasks	0.36	0.53	0.17	-0.34	0.10
Contribution to society	0.81	0.51	-0.46**	0.54	0.57
Time for family	0.21	-0.09	0.12	-0.08	0.40
Help others	0.89	0.63	-0.53**	-0.06	0.31*
Self-Efficacy *(Z Score)*	0.21	0.05**	-0.07**	0.23	0.20
N	1,716	1,162	656	519	11,701

Notes: Teachers includes anyone who has taught between W1 and W6. Nurses, Govt/Admin and Protective officers include all those who have not been a teacher in W1 to W6 but have worked in the relevant occupation between W1 and W6. Other Graduates includes all those not in one of the other four occupation between W1 and W6 who have an undergraduate or postgraduate degree. Government officials does not include senior civil servants (Grade 5 and above). Job preferences were only asked from a 10% subsample. ** = p< 0.01, * = p< 0.05.

2.4 Long Run Changes in Teacher Characteristics

In this section, I present statistics describing long-run change in the characteristics of those entering teaching. Ideally, I would analyse change in the characteristics of teachers at the point they begin teaching. However, USoc only includes information on whether respondents were teachers in their first job after leaving full time education and in each year during the data collection period (2009-2015). Information between first job and 2009 is therefore censored, making it hard to ascertain exactly when somebody entered teaching. In order to analyse long run change in entry to teaching, I therefore look at all those in the dataset who I can identify as teaching at any time and analyse change based on the period in which the teacher was born. I refer to these as birth cohorts to distinguish them from either trainee cohorts or familial generations. I argue that looking at entry in this way captures the influence of formative social and cultural experiences that would have had a distinctive, common

influence on each birth-cohorts' attitudes towards work and occupational choice (Mannheim, 1970). There is an ongoing debate about the best way to define the different birth-cohorts (Parry & Urwin, 2011) but I follow the standard approach in dividing all teachers in the dataset into four commonly-used birth cohorts (Strauss & Howe, 1991; How & Strauss, 2009). These are: those born up to 1945 (Pre 1945, N=13,865); those born 1946-1964, who entered the labour market no earlier than 1961 (Baby Boomers; N=23,457); those born 1965-76, who entered the labour market no earlier than 1980 (Generation X; N=18,380); and those born 1977-1995, who entered the labour market no earlier than 1993 (Millennials; N=31,066).

Figure 6 shows changes in the characteristics of teachers based on which of these four birth-cohorts they belong to. Because data collection only began in 2009, I compare teachers only in terms of time-invariant characteristics. The proportion of female teachers dropped from 80% in the pre-1945 birth cohort to 66% in the Gen X birth cohort. It then increased slightly among Millenials, but the difference with Gen X is not statistically significant at the 5% level. This long run decline in the proportion of female teachers is consistent with evidence from the US and likely reflects the increase in employment opportunities for women in other professions (Bacolod, 2007a). Figure 6 also reveals a steady increase in the proportion of ethnic minority teachers, from 8% among the Pre 1945 birth-cohort to 21% in the Gen X and Millenial birth-cohorts. Table 17 in Appendix B shows that the odds of those from an ethnic minority teaching, conditional on other characteristics, has actually fallen across cohorts. This suggests that the increase in ethnic minority teachers is being driven by the increase in the proportion of ethnic minority people in the population as a whole, rather than an underlying increase in the propensity to enter the profession. The next set of bars shows the proportion of teachers who had either a mother or father who was a teacher. The point estimates increase steadily across the four birth-cohorts from around 4% among the Pre 1945 birth cohort of teachers to around 12% among Millenials. The difference between the Pre 1945 and Millenial birth-cohorts is statistically significant at the 5% level. The final set of columns shows the proportion of teachers in each birth cohort whose father was a university graduate. The point estimates among Pre 1945, Baby Boomers and Millenials are very similar, varying between 21% and 24%. The point estimate for Gen X is closer to 23% but this difference is not statistically significant at the 5% level.

Figure 6: Generational change in characteristics of entrants to the teaching profession.
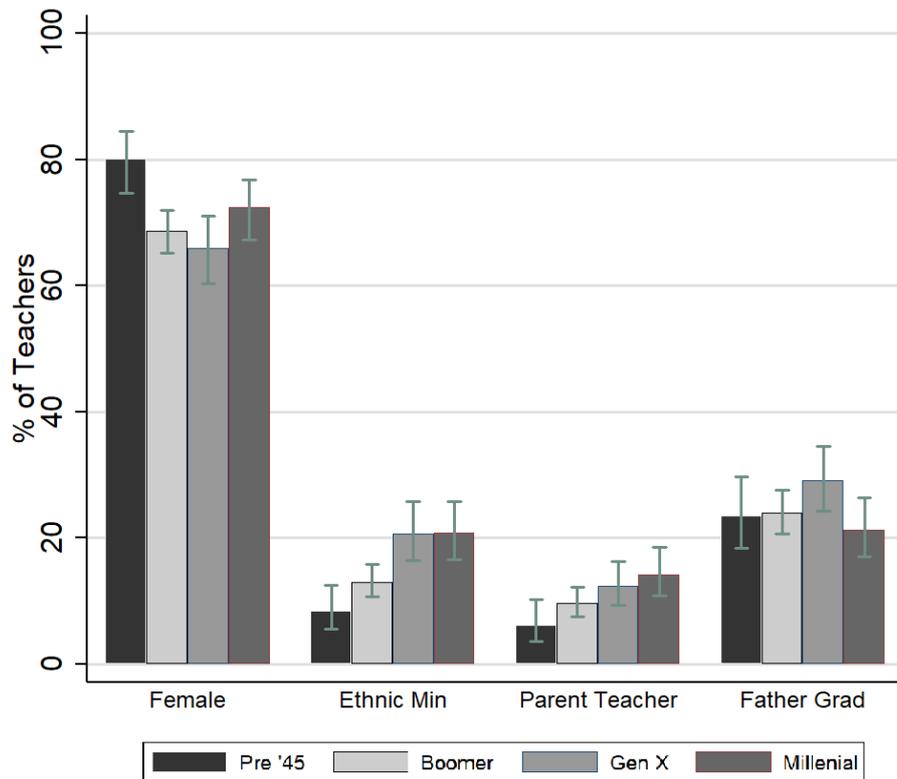
Notes: Figure 6 shows the characteristics of teachers from four different birth cohorts. Teachers are defined as anyone who can be identified as teaching at any point in their career. N (number of teachers) = 2,485. Pre '45 includes all those born before 1945. Boomer includes all those born 1946-1964. Gen X includes all those born 1965-76. Millennials include all those born 1977-1995. Error bars show 95% confidence intervals.

2.5 Modelling Entry to Teaching

In this section I use the full breadth of variables included in Understanding Society to model the choice to become a teacher. Many of the variables relevant to occupational choice are only measured during the data collection period, rather than retrospectively. For example, it would be very hard to accurately measure in 2009, which is the first year of data collection, what somebody's personality was like in 2001. This means that, for many of the teachers in the dataset, these variables are measured long after they made the decision to enter the teaching profession and are therefore potentially endogenous to that choice. This makes them of little use for predicting entry. I take two approaches to dealing with this problem. First, I follow Chevalier et al. (2007) and model teaching status. That is, whether or not an individual is working as a teacher in any year since the dataset was first collected. This has the advantage of maintaining sample size. However, strictly speaking, this is not a measure of entry to the teaching profession, since being in service requires not just entry to the profession, but also a lack of exit from the profession. The results from this analysis are shown in Table 4 below. My second approach is to run the models with those who I can genuinely observe

entering the teaching profession during the period of data collection and use their characteristics measured at or after 2009. This approach will reduce sample size and the precision of my estimates but will reduce measurement error due to both lags in measurement and endogeneity in my estimates. The results from these models are shown in Table 5 below.

I use logistic regression to model teaching status and entry into the teaching profession using variants of the following model:

$$log \left(\frac{p(entry_i)}{1 - p(entry_i)} \right) = \beta_0 + \beta_1 D_i + \beta_2 E_i + \beta_3 F_i + \beta_4 P_i + \beta_5 S_i + \beta_6 SE_i + \beta_6 Z_i + u_i$$

Where:

- $p(entry_i)$ is the probability that individual i enters the teaching profession
- D_i is a vector of individual demographic characteristics including gender, ethnic minority status, whether they are a university graduate, whether they are a first generation immigrant, and whether they are a second (or higher) generation immigrant
- E_i is a measure of the unemployment rate among over 16s at the time an individual was 21
- F_i is a vector of individual i's family characteristics including whether their father has a degree, whether their mother was a teacher and whether their father was a teacher
- P_i is a vector of personality measures for individual i covering the Big Five dimension of personality: openness to new experience, conscientiousness, extraversion, agreeableness and neuroticism
- S_i is a vector of individual i's social network characteristics including number of close friends and an interaction term which takes the value one if somebody is a graduate and has a majority of friends who are graduates
- SE_i is a single variable containing the average score of the ten item measures in the short Generalized Self-Efficacy Scale
- Z_i are country dummies for the four nations of the United Kingdom

Descriptive statistics for all the independent variables are available in Table 18. Job preferences variables are not included in the model because they were only collected for a small subset of USoc panel members.

I begin by modelling whether or not somebody taught at any point during the data collection period (2009-2014) among all those who were of working age during this period (N=18,022). Column 1 of Table 4 reports results from a model which includes only demographic and family background variables. Column 2 adds to this measures of the big five personality traits. Personality traits were only measured in Wave 3 of the survey and are therefore only available for a subset of my

respondents, meaning the number of observations drops to 14,760. It should be noted that only 16 of the units that drop out have missing data on specific personality traits, with the rest being attributable to the question not being administered. Column 3 adds a measures of the size of an individual's social networks. The number of observations in this model drops by 93, which is accounted for by the questions about number of close friends only being asked in Wave 3 and Wave 6, not due to specific item non-response. Finally, column 4 adds a measure of self-efficacy. The number of observations drops by 1,474 because this question was only asked in Wave 5. Despite these small changes in the sample across models, the coefficients and statistical significance of the coefficients are very stable across the different models.

Table 4 replicates a number of findings from previous research. The odds of females teaching are around 125% higher than men, conditional on the other covariates. Graduates are also significantly more likely to teach than non-graduates, which is not surprising given that teachers in the UK have been required to have a graduate level qualification in order to achieve qualified teacher status since the abolition of the Licensed Teacher Scheme in 1998 (Furlong et al., 2000). I also replicate the finding from Cobb-Clarke & Tan (2011) that people with a personality type characterised by more openness to experience are more likely to teach. A one standard deviation increase in openness is associated with a 28% increase in the odds of teaching.

Table 19 in Appendix B shows results from the same analysis using mean imputation and missing dummies. This findings remain virtually unchanged in this alternative specification, and the missing dummies for the different dimensions of personality type do not indicate any systematic association between missing values and the outcome, conditional on other covariates. Table 20 in Appendix B shows a third version of Table 4, this time with the Column 4 sample (N=13,649) used across all four columns. The results are again virtually unchanged. These two appendix tables provides reassurance that the sub-sample of individuals who had these questions administered to them constitute a representative sample of the full USoc sample.

Table 4: Logistic regression models of current teaching status

	(1) Demographics	(2) Demographics & Personality	(3) Demographics & Personality & Networks	(4) Demographics & Personality & Networks & Efficacy
Female	2.251**	2.231**	2.235**	2.267**
	(0.189)	(0.212)	(0.214)	(0.224)
Ethnic minority	0.890	0.858	0.859	0.861
	(0.138)	(0.154)	(0.154)	(0.163)
Graduate	10.89**	10.28**	10.15**	9.431**
	(1.114)	(1.211)	(1.201)	(1.168)
First gen immigrant	0.399**	0.444*	0.443*	0.541
	(0.0981)	(0.130)	(0.130)	(0.166)
Second plus gen immigrant	0.918	0.977	0.975	1.110
	(0.161)	(0.205)	(0.205)	(0.253)
Father is a graduate	0.985	0.993	0.990	0.980
	(0.0941)	(0.101)	(0.101)	(0.103)
Mother/father taught when 14	1.219	1.193	1.188	1.254
	(0.171)	(0.177)	(0.177)	(0.190)
Personality (z scored):				
Agreeableness		1.171*	1.167*	1.161*
		(0.0588)	(0.0586)	(0.0612)
Conscientiousness		0.958	0.961	0.966
		(0.0481)	(0.0483)	(0.0520)
Extraversion		1.018	1.009	0.983
		(0.0477)	(0.0480)	(0.0487)
Neuroticism		1.063	1.065	1.104
		(0.0521)	(0.0522)	(0.0605)
Openness		1.277**	1.277**	1.289**
		(0.0699)	(0.0702)	(0.0746)
No. of close friends (z scored)			1.056	1.056
			(0.0411)	(0.0420)
Self-efficacy (z scored)				1.144
				(0.0626)
Pseudo R-squared	0.055	0.117	0.121	0.122
Number of Observations	18,022	14,760	14,667	13,649

Notes: Each column is a separate logistic regression on a separate group of teachers. Coefficients are odd ratios. ** = p< 0.01, * = p< 0.05. Numbers in parentheses are standard errors. Models also contains country dummies for the four nations of the UK. Number of observation drops due to personality, social network and self-efficacy questions only be asked in certain waves. Pseudo R Squared is McFadden's Pseudo R Squared.

Table 4 also includes a number of new findings. Contrary to existing literature such as Goldhaber and Liu (2003), I do not find a relationship between having a parent who taught and being a teacher. The coefficients do show a stable and positive association but are not statistically significant at

conventional levels, conditional on other characteristics. Being from an ethnic minority is not associated with the odds of teaching between 2009 and 2015, conditional on other characteristics. However, being a first generation immigrant does reduce the odds of teaching by around 50%. One other personality characteristic, besides openness, is also associated with teaching. A one standard deviation increase in agreeableness – which manifests in cooperative and considerate behaviour - is associated with a 16-17% increase in the odds of teaching. Conscientiousness, extraversion and neuroticism do not have any statistically significant relationship with teaching status. Contrary to Borghans et al. (2008) and Krueger and Schkade (2008), I do not find any statistically significant relationship between how sociable somebody is in their private life, as measured by number of close friends, and whether they teach. Finally, self-efficacy also has a positive relationship with teaching, with a one standard deviation increase being associated with a 14% increase in the odds of teaching. However, this is not statistically significant at the 5% level. Again, the imputation dummies in Table 19 do not indicate any association between missing values and the outcome variable, conditional on other covariates. In Table 21 in Appendix B, I enter each of the sets of variables into the model separately in order to check that the associations are not being driven by correlations between the variables. All statistically significant relationships from the prior analysis are reproduced using this alternative approach.

Because of the timing of measurement, the characteristics of teachers identified in Table 4 may be the result of people choosing to become teachers, rather than the reason they chose teaching in the first place. Because of this, the results in Table 4 are best interpreted as rich descriptions of people who are in teaching. In the next steps of the analysis, I use the same set of variables to model entry to the profession. Each column in Table 5 uses a slightly different operationalisation of entry to the profession. Column 5 defines the entry outcome as one for anybody who is not teaching in Wave 1 of the dataset (2009-10) but is teaching in any subsequent year and zero for everybody else of working age. Column 6 defines the entry outcome as one for anybody who does not work as a teacher in their first job after leaving fulltime education *and* is not working as a teacher in Wave 1, but is working as a teacher in a subsequent year; and zero for everybody else of working age. Finally, column 7 defines the entry outcome as one for anybody who is not working as a teacher in Wave 1-6 *and* is less than 30 years of age at Wave 1, and zero for anybody else who is of working age but is less than thirty years of age at Wave 1. Models 6 and 7 are intended to exclude those who made the decision to enter teaching long before the covariates were measured. Column 6 excludes those who taught at the beginning of their career but then left and returned after Wave 1. Column 7 excludes all those who are not in the early stage of their career. Column 6 and Column 7 are therefore more likely to be measuring the covariates at or near the point of entry to the profession. As a result, there should be less measurement error and less endogeneity present in the estimates. However, as the falling number

of observations and rising standard errors in Table 5 show, this is achieved at the cost of reduced precision in the estimates. Models 5, 6 and 7 also include the rate of unemployment (measured as a z score) at the time each person in the data was 21, in order to account for the effect of macro-economic cycles on graduate occupational choice.

Being female and being a graduate has a strong positive association with the odds of entry to teaching across all three models. Being a first generation immigrant again has a strong negative association with entry to the teaching profession, although this not statistically significant at the 5% level. The self-efficacy and social network measures show consistently positive associations but these are also not statistically significant at conventional levels. The variable measuring unemployment rate at 21 enters with the expected sign, but is larger and only statistically significant at conventional levels, in column 7. This is the model that only includes those under 30 in 2009 and may therefore be picking up the effect of the financial crisis, which had a particularly large effect on graduate unemployment.

The five personality variables appear in the middle of Table 5. As in Table 4, openness to experience shows a positive relationship with the odds of entry to teaching. In Models 5 and 6, a one SD increase in openness is associated with a 71-77% increase in the odds of entry to teaching, statistically significant at the 5% level. This is around three times stronger than the associations with current teaching status in Table 4, suggesting that openness plays a particularly important role in attracting people into teaching in the first place, as opposed to keeping them in the profession. In Model 7, the association remains positive, but is weaker and no longer statistically significant at conventional levels. While agreeableness showed a robust association with current teaching status in Table 4, it shows no relationship with entry to teaching in Table 5. Conversely to openness, this suggests that agreeableness plays a particularly important role in keeping people in teaching, rather than attracting them in the first place. As in Table 4, extraversion and neuroticism show no relationship with entry to teaching, at conventional levels of statistical significance. Indeed, the only dimension of personality type, other than openness, which shows a relationship with entry to teaching in Table 5 is conscientiousness. In particular, a one standard deviation increase in conscientiousness is associated with an approximately 25% reduction in the odds of entry to teaching. As with agreeableness, this association is statistically significant in Models 5 and 6, but not in Model 7. The coefficient remains stable in Model 7 but the smaller sample size increases the standard errors.

This raises the question of how and why openness, agreeableness and conscientiousness affect people's decision to teach. Psychologists define personality as the pattern of dispositional tendencies, motivations and values, and self-narratives (ways of interpreting or attributing meaning) specific to each individual (McAdams & Pals, 2006). Openness to new experience is associated with both current teaching status and entry to teaching in Table 4 and Table 5. People with personality types characterised by high openness to new experience are: disposed to seek out new experiences and

make connections between ideas, value and are motived by variety and novelty, and see themselves as learners (Sutin, 2017). Empirical studies show that individuals with high openness to new experience seek out and acquire jobs that provide autonomy and opportunities for growth (Bipp, 2010; Sutin & Costa, 2010). Interestingly, educationalists have highlighted both autonomy and opportunities for growth as notable and distinctive characteristics of teaching (Labaree, 2000; Loewenberg, Ball & Forzani, 2009). This may explain the strong associations found between openness, current teaching status and entry to teaching in this study.

I also find that agreeableness is positively related to current teaching status (Table 4) though not entry to teaching (Table 5). People with personality types characterised by high agreeableness are: disposed to being considerate and flexible, value and are motived by maintaining positive relations with others, and see themselves as social types (Graziano & Tobin, 2017). There is currently little research on the links between agreeableness and job characteristics that help shed light on this link. However, the need for teachers to maintain positive working relationships with very high numbers of pupils are well known (Labaree, 2000) and this may be harder for individuals who do not display agreeableness. This interpretation is also consistent with my finding that agreeableness is associated with current teaching status, rather than entry. Agreeableness therefore seems to have protective effect on teaching status.

The third and final personality dimension which is associated with teaching in my data is Conscientiousness, which I find to be negatively associated with entry to teaching (Table 5) but not related to current teaching status (Table 4). People with personality types characterised by high conscientiousness tend to be on the one hand goal oriented and hardworking and on the other hand self-controlled, orderly and rule abiding (Jackson & Roberts, 2017). Unfortunately, there is precious little research available to help explain why conscientious individuals would be less likely to enter teaching. It is plausible that highly goal-oriented individuals have already adopted specific career plans prior to reaching the end of their undergraduate degree, which is when most teachers begin their training. This is however only a hypothesis and needs to be tested empirically using other data.

Table 5: Modelling entry to the profession using three definitions of entry

	(5)	(6)	(7)
	Not teaching W1 & teaching W2-W6	Not teaching W1 & not teaching 1st job & teaching W2-W6	Teaching W1-W6 & under 30 at W1
Female	2.722**	2.695**	2.309*
	(0.575)	(0.619)	(0.744)
Ethnic minority	1.362	1.306	1.034
	(0.378)	(0.390)	(0.362)
Graduate	3.880**	3.716**	6.705**
	(0.806)	(0.817)	(2.276)
First gen immigrant	0.274	0.239	0.131
	(0.150)	(0.137)	(0.125)
Second plus gen immigrant	1.079	0.904	0.881
	(0.423)	(0.347)	(0.402)
Unemployment rate at 21	1.045	1.091	4.291**
	(0.0746)	(0.0833)	(1.058)
Father is a graduate	0.855	0.786	0.884
	(0.184)	(0.189)	(0.289)
Mother/father taught when 14	1.913	1.956	1.376
	(0.547)	(0.600)	(0.601)
Personality (z scored):			
Agreeableness	1.032	1.025	1.247
	(0.103)	(0.107)	(0.177)
Conscientiousness	0.754*	0.762*	0.868
	(0.0705)	(0.0752)	(0.133)
Extraversion	1.063	1.007	0.975
	(0.0930)	(0.0907)	(0.133)
Neuroticism	1.164	1.063	1.122
	(0.132)	(0.123)	(0.188)
Openness	1.712**	1.771**	1.376
	(0.183)	(0.212)	(0.212)
No. of close friends (z scored)	1.097	1.108	1.034
	(0.0590)	(0.0617)	(0.132)
Self-efficacy (z scored)	1.166	1.187	1.187
	(0.128)	(0.139)	(0.217)
Pseudo R-squared	0.103	0.097	0.113
Number of Observations	15,239	15,269	3,513

Notes: Each column is a separate logistic regression. Column 5 defines entry as not teaching in Wave 1 of the dataset but teaching subsequently, column 6 as not teaching in first job after fulltime education or in Wave 1 but teaching subsequently, column 7 as working as a teacher in Wave 1-6 and being <30 years old at Wave 1. ** = p< 0.01, * = p< 0.05. Numbers in parentheses are standard errors. Models also contain country dummies for the four nations of the UK. Pseudo R Squared is McFadden's Pseudo R Squared.

2.6 Predictive Margins

I now use the result from Section 2.5 to compare the predicted probability of different groups entering the teaching profession. Figure 7 shows the predictive margins for different values of three of the variables that showed consistently strong associations with teaching in Table 5: gender, openness and conscientiousness, when all other variables are evaluated at their mean (continuous) or mode (categorical) values. The vertical axis shows the predicted probability of entry to teaching from column 6 with 0.01 representing a 1% probability, 0.1 representing a 10% probability, and so on. The left hand panel plots this against the seven point openness scale (7 being the most open) and the right hand side does the same for conscientiousness. The black circles show the predicted probability of entry for women and the grey circles for men. In the left hand panel conscientiousness is being held constant at the mean; in the right hand panel openness is being held constant at the mean. The error bars around these points show the 95% confidence interval on these estimates. The chart shows that the predicted probability of entry to the teaching profession varies from close to 0%, to just under 1% (men) and just over 2% (women). To put this in context, the predicted probability of the average person entering teaching based on column 6 is around 0.5%. The findings are consistent with existing research which suggests that personality type has a stronger association with occupational choice for women than for men (Ham et al., 2009).

Figure 7: Predictive margins for entry to the teaching profession
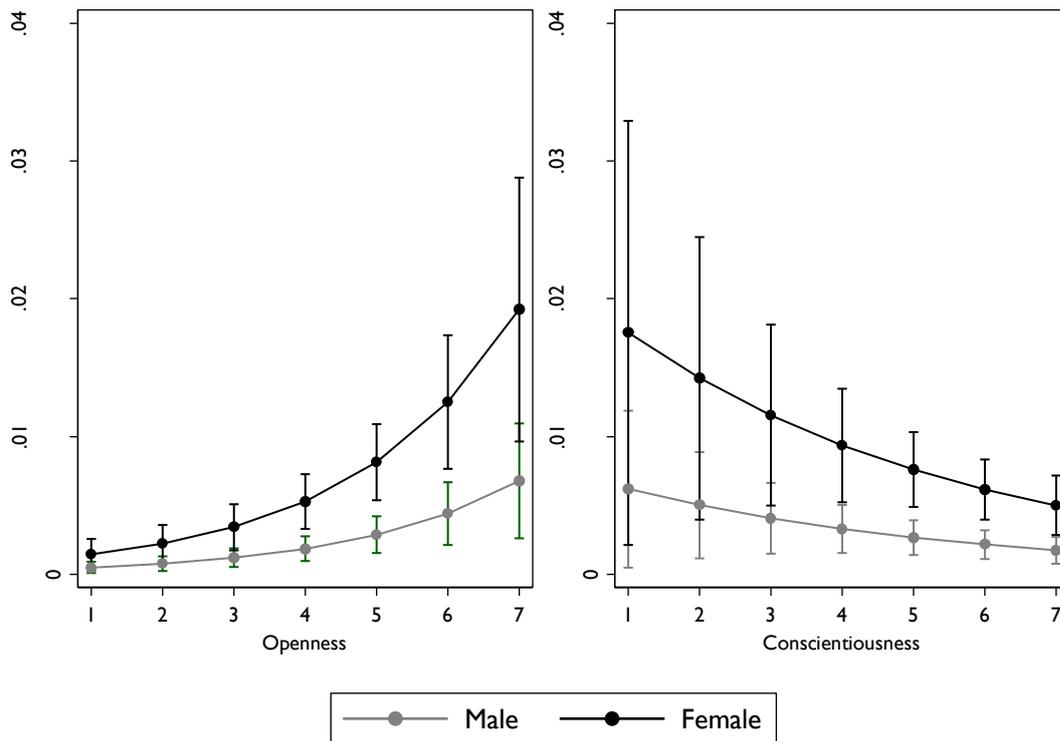
Notes: Figure 7 shows the predicted probability of entering the teaching profession for men and women for different values of openness and conscientiousness, when all other variables are set to their mean (continuous variables) or mode (categorical variables) values. The predicted probabilities are calculated using column 6 from Table 5. The Y axes shows probability measured on a scale of zero to one. The X axes shows personality traits measured on a seven point scale with 7 representing a highly open/conscientious person.

Figure 8 provides estimates of the predicted probability of entry for groups of graduates with particular combinations of four of the characteristics that showed consistently strong associations with entry to teaching in Table 5: Female, Parent teacher, Conscientiousness and Openness. The sixteen groups defined by these four variables are shown in Table 6. The sample is restricted to graduates, since all those entering teaching in column 5 were required to have an undergraduate qualification. The vertical axis of Figure 8 again shows predicted probability of entry to the teaching profession. The error bars show the 90% interval estimate of the predicted probability of entry for each of the sixteen groups, using the coefficients from Model 6. The group labelled Typical Grad shows the predicted probability of entry for the average graduate. Figure 8 reveals that there is wide variation in the predicted probability of entry to the teaching profession: from 0.8% for group 1 to 8% for Group 16. However, the estimates are also quite imprecise, with the average interval estimate being 3.3 percentage points wide. The only group with an interval estimate that is clearly non-overlapping with that for the typical graduate is group 16. This is the group of female graduates who had at least one parent who taught, have low levels of conscientiousness and high levels of openness to new experience. The predicted probability of entering teaching for this group is 8%, which is around four times higher than that for the typical graduate. Despite having overlapping confidence intervals,

groups 14 and 15 - females with parent teachers with only one of the two desirable personality characteristics – also have point estimates two to three times higher than the typical graduate.

Figure 8: Predicted probability of entry to the teaching profession
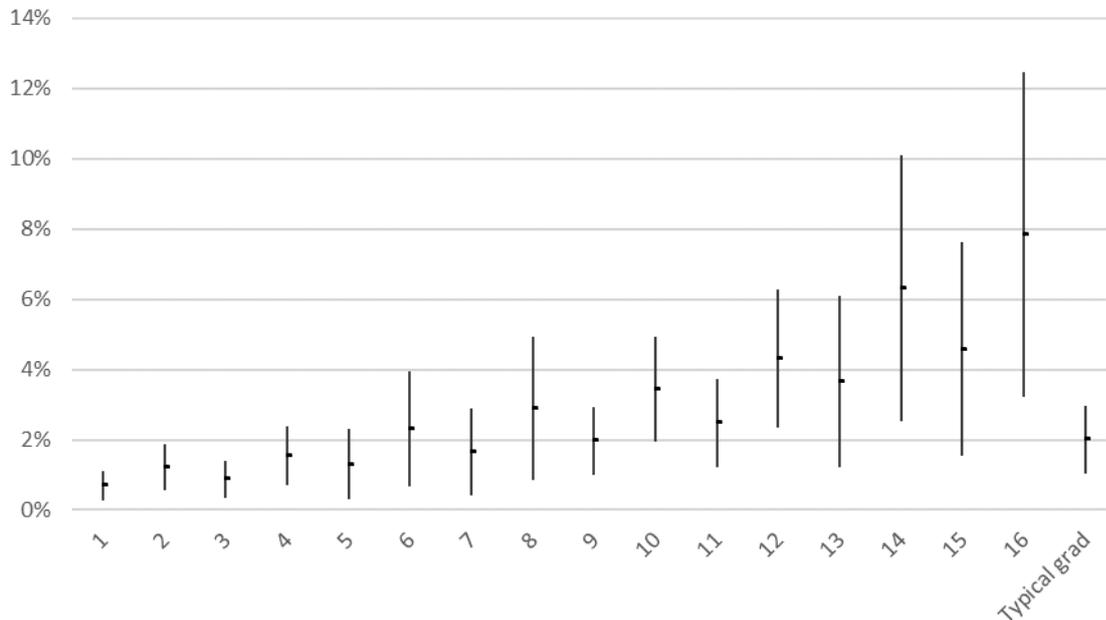
Notes: Figure 8 shows the predicted probability of entry to the teaching profession for sixteen different types of individuals, calculated using column 6 from Table 5. The vertical bars show the 90% confidence intervals.

Table 6: Groups in Figure 8

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Characteristic
								✓	✓	✓	✓	✓	✓	✓	✓	Female
			✓	✓	✓	✓						✓	✓	✓	✓	Parent was teacher
		✓	✓		✓	✓			✓	✓				✓	✓	Low conscientiousness
✓		✓		✓		✓		✓		✓		✓			✓	High openness

Notes: "Low" ("High") is operationalised as one standard deviation above (below) the mean value.

2.7 Conclusion

Despite the importance of teachers for pupils' academic attainment, and persistent shortages of teachers in many developed nations, we still know relatively little about the reasons that people choose to enter the teaching profession. This is in part because of the limited set of variables included in the graduate cohort datasets which have so far been used to track progression into the profession. This research has addressed this gap in the literature by using rich data from a household panel survey to shed new light on entry to the teaching profession.

The proportion of ethnic minority teachers in each new birth cohort has increased steadily over time. However, the odds of those from an ethnic minority going into teaching have fallen with each new

birth cohort, suggesting that the unconditional increase over time is due to the overall increase in the proportion of ethnic minority people in the UK. Another notable change is that each new birth cohort of teachers has seen an increase in the proportion of teachers whose parents also taught. Finally, the long-established trend towards an equalisation of the proportion of male and female teachers seems to have flattened off among the most recent birth cohort of teachers. Data on the current stock of teachers shows that they have relatively high job- and life-satisfaction relative to comparable occupational groups and graduates in general, but are also relatively dissatisfied with their leisure time compared to other graduates.

This research has also provided new evidence on the determinants of entry to teaching. In particular, the importance of personality type emerges clearly from this research. In particular, openness to new experience shows robust associations with current teaching status and entry to teaching. Agreeableness is also robustly related to current teaching status, though not entry to teaching, suggesting perhaps that it helps *keep* teachers in the profession. Conscientiousness is also negatively associated with entry to teaching. Indeed, groups with particular sets of characteristics are up to four times more likely to enter the teaching profession than the average graduate, suggesting that these relationships are of substantive importance for policymakers looking to tackle teacher shortages.

While these findings do not provide causal evidence on why people enter teaching, the richness of the data used to model entry to the profession means that they are likely to have predictive value. For example, the results from Section 5 and 6 could potentially be used by policymakers to target recruitment efforts. This is of course dependent on being able to observe the characteristics I find to be associated with entry to teaching in the population as a whole. The obvious place to get a larger sample of graduates is in graduate surveys. In the UK a new graduate survey, the New Destination of Leavers of Higher Education (NDLHE), is currently being developed and is set to be implemented in December 2018. This survey aims to collect information from at least 70% of all university-leavers, fifteen months after graduation. While the fifteen month delay is not ideal, since many graduates will already have made important career choices at this stage, the NDLHE survey does represent a good opportunity to gather data on, e.g. personality type that could help identify those with a high propensity to enter teaching. This data source also includes detailed information on the degree subject of graduates, which would allow targeted approaches to be made to those with qualifications in subjects in which there is currently a shortage, such as maths, physics and modern foreign languages (Chapter 1). Once those with a high-propensity for teaching had been identified, it would then be possible to target them with information about entering the teaching profession. This would also provide an opportunity to conduct an independent test of the predictions made in this paper about the propensity of certain groups to enter teaching.

Chapter 3: How are Working Conditions associated with Teacher Job Satisfaction and Desire to Move School?

3.1 Introduction

High teacher turnover in a school is associated with reduced pupil attainment, both because of the disruption caused by having to hire replacements and because the replacement teachers tend to be initially less effective (Atteberry et al., 2016; Gibbons et al., 2018; Ronfeldt et al., 2013). High levels of turnover in a school can also create acute staffing shortages. School leaders tend to respond to such shortages by making increased use of temporary teachers, lowering recruitment standards, or increasing class sizes (Smithers & Robinson, 2000), all of which have been linked with reduced pupil attainment (Mocetti, 2012; Fredriksson et al., 2012; Schanzenbach, 2006).

Where turnover reflects teachers leaving the profession altogether, there are also high opportunity costs incurred in training replacements. In England, approximately 33,000 people begin teacher training each year at an annual cost of around £700m to the public purse (NAO, 2016). However, 40% of those who enter teacher training are no longer teaching in public schools after five years (Allen et al., 2016). With a sufficient supply of teachers, such turnover has ambiguous effects on the quality of education, depending on the quality of the departing teacher. However, when there is an overall shortage of teachers, such wastage is unambiguously costly. Understanding the determinants of teacher job satisfaction and turnover is therefore important.

Early research using administrative data found that teachers are much more likely to leave their job if they work in schools with high levels of deprived, minority or low attaining children and those that stayed in the profession tend to move to schools with more affluent intakes (Boyd et al., 2005; Hanushek et al., 2004; Scafidi et al., 2007; Allen et al., 2018). This research prompted many to conclude that teachers were fleeing "difficult to teach" pupils and that there was little policy could do to address this. However, a recent reinterpretation of the literature on teacher turnover has begun to change this view.

Simon and Johnson (2015, p10) point out that in six studies in which efforts were made to control for the quality of teachers' working environment "all or most" of the relationship between student characteristics and teacher turnover disappeared. The inter-personal aspects of working conditions, such as leadership and teacher collaboration, appear to have the strongest relationship with retention (Simon & Johnson, 2015). As well as being associated with teacher retention, working conditions also predict job satisfaction (Collie et al., 2012; Malinen & Savolainen, 2016; Skaalvik & Skaalvik, 2009; Reeves et al., 2017). Indeed, some research suggests that job satisfaction mediates the relationship between conditions and turnover (Skaalvik & Skaalvik, 2011). This has important implications for

policy since it creates the potential for improving teacher retention by improving working conditions in schools.

Despite these recent advances, there is currently very little research using data from outside the USA, which raises questions about the applicability of these findings in other school systems. A number of papers have analysed bivariate relationships using the TALIS 2013 data (Micklewright et al., 2014; Sellen, 2016) and one paper has modelled job satisfaction using the 2008 data (Duyar et al., 2013; Fackler & Malmberg, 2016). This paper goes beyond the existing literature in three main ways. First, it uses exploratory factor analysis (EFA) to develop a new way of categorising and quantifying the working conditions measures in the TALIS survey. Second, this paper is the first to model the determinants of teachers' desire to leave their school using the TALIS data. Third, in later stages of the analysis, I link the TALIS data to department-level administrative data on teachers in England, including objective measures of pupil deprivation, to provide a richer set of control variables.

I find that the nature of leadership and management has a particularly strong association with both satisfaction and desire to leave, highlighting the central importance of school leaders in improving retention. This is consistent with findings from research using other datasets (Boyd et al., 2011; Ingersoll, 2001; Ladd, 2011). As well as replicating these findings, I identify a number of new correlates of improved teacher job satisfaction including feedback, student discipline and scope for career progression. Teachers being assigned to teach classes and subjects for which they feel well prepared is also strongly associated with reduced desire to move school.

These findings come from cross-sectional survey data, which makes it hard to rule out reverse causation. My interpretation of the associations in this paper therefore relies on two additional sources. First, theoretical accounts which provide descriptions of the psychological mechanisms by which poor working conditions increase the demands on teachers and/or reduce the resources available to them, leading to teacher burnout and then turnover (Bakker & Demerouti, 2007; Crawford et al., 2010; Fernet et al., 2013). Second, previous research using longitudinal data which has shown that changes in working conditions predict changes in teacher retention outcomes at later times points (Boyd et al., 2011; Donaldson and Johnson, 2010) and not the other way round (Kraft et al., 2016). Both sources of evidence help reduce concerns about reverse causality with respect to the associations identified in this paper.

The rest of the paper is organised as follows. Section 2 describes the data. Section 3 presents the results using the international data. Section 4 describes the factor analytic methods used to derive the latent working conditions variables used in the England-only data. Section 5 provides the results from models using the England-only data. Finally, section 6 concludes by discussing implications for research and policy.

3.2 Data

TALIS is an international teacher survey that collects information on teachers' beliefs, practices and working conditions. TALIS was first conducted in 2008 and then again in 2013; the next survey will be in 2018. The teacher questionnaire for the 2013 survey contains 50 question groups covering teacher and school characteristics, professional development, feedback, pedagogy, attitudes to teaching, school climate and job satisfaction. The survey employs a complex design in which schools are sampled with probability proportional to size (with some countries also employing stratification at this stage) and then teachers are randomly sampled from within sampled schools (Dumais & LaRoche, 2014). The target sample in each country is 20 teachers in 200 schools, or 4,000 respondents. The final teacher dataset for 2013 contains just under 120,000 respondents, though responses vary in their level of completeness and some items have high non-response rates. The TALIS data is supplied with weights that account for the complex survey design and non-response. These weights are applied to all estimates in this research using the `REPEST` command in the STATA software, which uses balanced repeated replication (Kish & Frankel, 1970) to account for all clustering in the complex survey design when estimating standard errors (Avvisati & Keslair, 2016). The weighting method used here means that larger countries such as Japan contribute proportionally more information to the results when pooling data across countries.

The TALIS data covers 35 different countries.[6] International surveys such as TALIS are often subject to concerns that responses cannot be compared across respondents who read and respond to the survey in different languages and may interpret the questions through different cultural lenses (Rutkowski & Svetina, 2014). Ideally, international surveys will display measurement invariance, which suggests that the questions have been interpreted in a similar way across countries. The TALIS 2013 data has been subjected to extensive empirical testing for invariance (Desa, 2016; Desa et al., 2014) through multi-group confirmatory factor analysis. These tests have shown that the data display weak or metric invariance, which allows the sort of cross-country regression analysis employed in this paper. However, the same tests generally show that the data do not show strong or scalar invariance, meaning average scores cannot be compared across different linguistic and cultural groups.

There are two outcome variables for this study: teacher job satisfaction and teacher desire to move school. The job satisfaction variable in the TALIS dataset, TJOBSATS, is a composite measure made up of eight different items from question 46, all relating to teachers' evaluations of their job (Desa et al., 2014), for example "The advantages of being a teacher outweigh the disadvantages" and "I would recommend my school as a good place to work." Responses to each item are measured on a four-point scale from Strongly Disagree to Strongly Agree. Negative statements are reverse coded so that a

---

[6] Technically, it includes 36. However, the USA is generally excluded from such analyses due to low response rates (Becker et al., 2014, p8).

higher score on TJOBSATS indicates higher job satisfaction. Table 22 in Appendix C shows the means (2.9-3.3) and standard deviations (0.6-0.8) of the responses to each question in the international sample. It is slightly negatively skewed, with a median value of 11.7 and a mean of 11.9. The distribution is bimodal, with peaks around 12 and 16, which likely correspond with respondents scoring 3 out of 4 and 4 out of 4 on all the component questions. In the rest of the analysis, this variable is standardised to give it a mean of zero and a standard deviation of one.

Figure 9: Histogram of the teacher job satisfaction variable
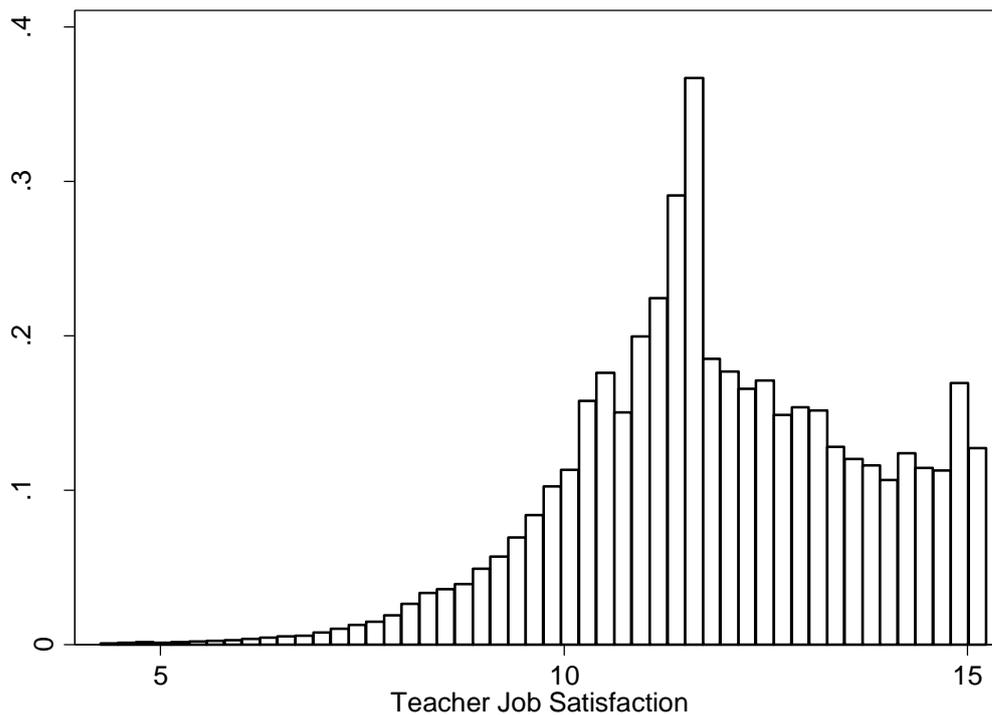
Notes: N = 112,120. The Job Satisfaction variable is calculated as an average of the satisfaction with working environment and satisfaction with profession complex scale scores. Each of these complex scale scores is created using confirmatory factor analysis and rescaled to have a mean 10 and standard deviation of 2 across the international sample. For more detail see Desa et al. (2014).

The second outcome measure for this study is desire to move to another school. This is measured using a single item from question 46, which asks teachers whether they agree with the statement "I would like to change to another school if that were possible", measured on the same four-point scale. It is worth noting that this item also makes up part of the job satisfaction score, but it is used separately here because it is of specific interest. As can be seen from Table 22 in Appendix C, this question has a mean (reversed) score of three (equivalent to responding Disagree) and standard deviation of 0.8.

3.3 Modelling Using International Data

The TALIS data contain many variables that are relevant to analysing teacher job satisfaction and retention (Borman & Dowling. 2008; Simon & Johnson, 2015). In order to enhance interpretability of my results, I make use of several of the latent variables that come in the data: professional development (the TEFFPROS variable in the TALIS data), teacher cooperation (TCCOOPS) and stakeholder engagement (TSCSTAKES). All three of these latent variables have been shown to display metric invariance and can therefore be used to model outcomes across the TALIS countries (Desa, 2016; Desa et al., 2014). I standardise all three variables to have mean zero and a standard deviation of one. Discipline is another aspect of working conditions that is relevant to teacher retention (Kraft et al., 2016; Simon & Johnson, 2015) and there is a composite score for this in the TALIS data (TCDISCS). However, this only relates to a single "target class" rather than the school in general. I therefore do not include it in my models in this section. In section 4 and 5 however, I develop my own measure of school discipline that helps avoids these pitfalls. Table 24 shows the pairwise correlations between the three latent variables, all of which have an absolute value lower than 0.3. This suggests that they are all measuring distinct latent variables.

The professional development latent variable captures how often the professional development at a school includes a range of features suggested by research to be associated with more effective professional development. It is measured by four manifest variables, each of which are measured on a four-point scale from "Not in any activities" to "Yes, in all activities." It is therefore intended as a measure of the *quality* of professional development. The four features of professional development it captures are: whether it involves learning with colleagues; whether it involves active learning methods; whether it involves collaborative learning activities; whether it takes place over an extended period of time, rather than being one off. Professional development has been shown to be related to teacher retention and job satisfaction in existing research (Kraft et al., 2016; Skaalvik & Skaalvik, 2011; Simon & Johnson, 2015).

The teacher cooperation latent variable captures how often teachers participate in a range of collaborative activities relating to their teaching. This is measured by eight manifest variables, which each refer to specific collaborative practices, such as "Exchange teaching materials with colleagues" or "Teach jointly as a team in the same class." Responses are measured on a six point scale from "Never" to "Once a week or more". Teacher cooperation has been shown to be related to teacher retention in existing research (Kraft et al., 2016; Simon & Johnson, 2015).

The stakeholder engagement latent variable captures teachers' responses to five statements relating to the engagement of important groups in the life of the school, e.g. "This school provides parents or guardians with opportunities to participate actively in school decisions". This is measured on a four-

point Likert type scale from "Strongly Disagree" to "Strongly Agree". Again, this has been shown to be relevant in the existing literature (Skaalvik & Skaalvik, 2011).

Besides these working conditions variables, the model includes: age (measured in years); a dummy indicating female gender; experience in teaching (measured in years); a dummy indicating whether a teacher has a permanent contract; a dummy indicting whether they have a Science, Technology, Engineering or Maths (STEM) degree, which has been shown to be relevant to retention (Allen & Sims, 2017); and a country indicator. Unfortunately, the international TALIS sample does not include information on the deprivation of pupils in the school, other than a categorical teacher self-report that is related to a single "target class". I therefore postpone use of pupil deprivation measures until Section 4, when I use linked survey and administrative data for England.

Table 7 shows the results from regressing the various working conditions measures, demographic variables and country indicators on the two outcome measures. Column 1 is an ordinary least squares (OLS) regression on the standardised teacher job satisfaction score. The coefficients show the standard deviation change in job satisfaction for a one unit change in each variable, holding the other covariates constant. Column 2 is an ordinal logistic regression on the four-point (Strongly Disagree to Strongly Agree) Likert response to whether a teacher would like to move schools. The coefficients are odds ratios and show the change in odds of moving one category towards the highest category (Strongly Agree) associated with a one unit change in each variable, holding the other covariates constant. A coefficient of less than one indicates a reduction in the odds. Both models include country fixed effects.

The models are based on around 85,000 observations with the missing observations largely due to the effective professional development variable, which is missing for 23,980 (20.6%) of respondents. The missing proportion is high because if any of the four underlying manifest variables used to create the scaled score were missing for an individual, the scaled score was not calculated for that individual. The analysis in Table 7 is therefore based on listwise deletion to of the cases with missing data. This is clearly undesirable in the sense that the data can no longer be claimed to be representative of the target population. The obvious alternative would be to impute the data using multiple imputation, which would maintain unbiased coefficient and standard error estimates (Cheema, 2014). However, guidelines on the imputation of missing data in education recommend that any variable with more than 15% of values missing should not be imputed because the necessary assumption of missing at random is untenable in such cases (Cheema, 2014). I therefore refrain from imputing the CPD complex scale score variable in this part of the analysis. In Section 3.5 I address this shortcoming by using multiple imputation on the underlying manifest variables (all of which have less than 15% of cases missing) and then derive my own complex scale scores based on these imputed manifest variables. The results with full imputation (Appendix C, Table 32) are very similar to the results

without imputation (Table 9, Table 10) which provides some reassurance that missing data is not likely to be biasing the results to a great extent in Table 7.

Table 7: Modelling job satisfaction and desire to move school across TALIS countries

	(1) Job Satisfaction (*Z scored*)	(2) Want to Move School (*Likert*)
Prof Development (*Z scored*)	0.10**	0.90**
	(0.01)	(0.01)
Teacher Cooperation (*Z scored*)	0.19**	0.81**
	(0.01)	(0.01)
Stakeholder Engagement (*Z scored*)	-0.01	1.00
	(0.01)	(0.01)
STEM Degree (*Dummy*)	0.01	0.93*
	(0.01)	(0.02)
Age (*Years*)	0.01**	0.99
	(0.001)	(0.01)
Female (*Dummy*)	0.02	0.96
	(0.02)	(0.03)
Experience (*Years*)	0.01	0.99**
	(0.01)	(0.001)
Permanent Contract (*Dummy*)	0.08**	1.00
	(0.02)	(0.04)
R Squared / Pseudo R Squared	0.13	0.10
N	85,116	84,559

Notes: Column 1 is an OLS regression and coefficients are marginal effects. Column 2 is a logistic regression and coefficients are odds ratios. Numbers in parentheses are standard errors. Standard errors are adjusted for clustering through the application of final and replicate weights. ** = $p < 0.01$, * = $p < 0.05$. Pseudo R Squared is McFadden's Pseudo R Squared. Both models include country fixed effects.

Looking first at column 1, effective professional development and teacher cooperation both have a positive relationship with job satisfaction, statistically significant at the 1% level. A one standard deviation (SD) increase in professional development is associated with a 0.1 SD increase in job satisfaction. I also find that a one SD increase in teacher cooperation is associated with a 0.2 SD increase in job satisfaction. This is somewhat smaller than the finding from Johnson et al. (2012) that a one SD increase in collegial support is associated with a 0.52 SD increase in job satisfaction. The third working condition, stakeholder engagement, has no statistically significant relationship with job satisfaction. This contrasts with the positive, statistically significant association with "community support" identified by Johnson et al. (2012). Age and having a permanent contract are both weakly positively correlated with job satisfaction. Column 2 shows the results of modelling teachers' desire to move school. Teacher cooperation is associated with reduced desire to move school, statistically significant at the 1% level. The coefficient indicates that a one SD increase in teacher cooperation is associated with around a 19% reduction in the odds that a teacher move one category towards strongly agreeing that they want to move school. Again, this is smaller than the finding from Johnson et al. (2012) that a one SD increase in collegial support is associated with a 57% reduction in odds of

wanting to move schools. A one SD increase in professional development is associated with a 10% reduction in the odds of moving one category up the response scale towards strongly agreeing that they want to move school, statistically significant at the 1% level. This finding contrasts with the finding from Ladd (2011) that there is no relationship between profession development and intention to move school. The final finding from Table 7 is that having a STEM degree and being more experienced are both negatively associated with desire to move school.

3.4 Measuring Working Conditions in English Schools

In the remainder of the paper, the sample is restricted to TALIS teachers in England. This has three advantages. First, it removes doubts about measurement invariance based on whether responses to the survey questions are comparable across different linguistic and cultural groups. This allows a much wider range of working conditions variables to be used in the analysis. Second, it allows the use of a set of variables derived from questions that were asked exclusively in the England version of the TALIS questionnaire (see Table 25 and Table 26 in Appendix C for more details). Third, it allows me to link in extra (school level) data on TALIS respondents in England from the School Workforce Census (SWC) dataset, including objective measures of pupil disadvantage, the proportion of pupils from ethnic minority backgrounds and the region and rural-urban status of the school. This allows a wider range of variables to be held constant in the analysis, while looking at the relationship between working conditions and the two outcome measures. The disadvantage of using only the English data is that the sample size falls to 2,060 linked teachers.

This dataset includes more than forty different variables measuring working conditions (Borman & Dowling, 2008; Simon & Johnson, 2015). Table 23 in Appendix C shows the level of missing values for each of these variables, ranging between 0.8 and 14.5%. Guidance in the literature states that variables with up to 15% of missing values should be dealt with through multiple imputation (Cheema, 2014). Multiple imputation has the benefit of ensuring the sample remains representative of the target population and allowing the estimation of unbiased standard errors. The corresponding downside is that it relies on the (strong) missing at random assumption. In the rest of this chapter I therefore present my main results using both the complete case (no imputation) sample and multiple imputation approaches. The results are qualitatively very similar.

Including all of these variables in the regression and interpreting the coefficients on each of them separately would be difficult. Moreover, some of the variables are likely to be measuring very similar concepts. Indeed, other research measuring working conditions in schools has identified groups of working conditions measures underlying the questionnaire responses (Corbell et al., 2008; Kraft et al., 2016; Ladd, 2011; Weiss, 1999). In the presence of a clear theoretical framework to guide aggregation, confirmatory factor analysis could be used to create composite scores. However, there are a number of competing theoretical frameworks (Deci & Ryan, 2008; Demerouti et al., 2001;

Eisenberger, 1986) and existing research has not settled on a dominant conceptualisation of working conditions. I therefore use exploratory factor analysis (EFA) because it is capable of both identifying latent variables underlying the measured working conditions variables and reducing the number of dimensions in the data in the process. This approach is common in related literature (Kraft et al., 2016; Ladd, 2011; Weiss, 1999).

Exploratory factor analysis aims to parcel common variance between groups of measured variables into distinct latent variables. A necessary condition for being able to identify latent variables is that groups of the measured variables share some common variance, otherwise the data is not factorable. The Kaiser-Meyer-Olkin (KMO) test quantifies this by comparing the pairwise association between measured variables (conditional on the other variables) with the sum of correlations. The rationale behind the KMO is that, if there are indeed latent factors to be identified, then pairwise conditional correlations should be relatively low, because conditioning on the other variables would remove the variation associated with the shared latent variable. A value close to zero means that partial correlations are relatively large compared to the sum of correlations and thus there appears to be little variation shared between groups of variables. A value close to one means that partial correlations are small relative to the sum of correlations and thus a high proportion of the variation appears to be shared between groups of measured variables. In this case, the KMO statistic has a value of 0.83, which is generally considered to merit factor analysis (Kaiser & Rice, 1974).

A second precondition for conducting an informative factor analysis is that the data comes from a large heterogeneous sample (Yong & Pearce, 2013). To see why, consider that if all observations varied in the same way along the same dimensions then there may be other latent variables present in the data, but not identifiable from observed variation (Gaskin et al., 2017). For the purpose of modelling, failing to identify a latent variable could create omitted variable bias even while the variable is measureable in the dataset. I argue that the TALIS data does indeed constitute a large heterogeneous sample. My dataset includes 2,060 teachers across 154 schools and Table 26 (Appendix C) shows substantial variation across the variables used in the analysis. The data is therefore appropriate for conducting an EFA.

The first step in conducting an EFA is to calculate the pairwise correlation between all the variables included in the analysis. Because many of the working conditions variables are ordinal I construct the correlation matrix using polychoric correlations, which assumes that the ordinal variables are imperfect measures of underlying, theorised-continuous variables. Simulations suggest this produces more accurate identification of latent variables than simply treating the variables as if they were continuous (Holgado-Tello et al., 2010). I construct the correlation matrix using the Polychoric command in the STATA software (Kolenikov & Angeles, 2004).

The process of extracting latent variables from the data begins by identifying the group of measured variables that are most strongly correlated with each other. Following this, the EFA process proceeds to identify the second group of measured variables which explain the most variation in each other, and so on. In order to ensure that this second set is distinct from the first, EFA involves rotating the factor axis in between extracting each factor. The two main types of factor rotation available are orthogonal, which rotates the factor axis by ninety degrees between extracting each factor, and oblique, which may rotate the axis by less than ninety degrees. Almost without exception, the other articles in this literature (Kraft et al., 2016; Ladd, 2011; Weiss, 1999) employ orthogonal (Varimax) rotation. However, the use of orthogonal rotation methods has been strongly criticised on the grounds that it assumes the latent variables are uncorrelated and this is unlikely to be true in many social scientific settings (Preacher & MacCallum, 2003). For example, it is plausible that schools with good leadership also offer good professional development. In any case, as Preacher and MacCallum (2003) point out, if the data is indeed suited to an orthogonal factor structure, then an oblique rotation will resemble that achieved by an orthogonal rotation. I therefore employ an oblique (Promax) factor rotation. Table 28 in Appendix C shows the factor loadings after rotation, which can be thought of as the weight with which each variable contributes to each latent variable. The uniqueness column shows the proportion of variance of each variable that is not shared with other variables. The rotated factor structure is very simple, with no variables loading on more than one factor.

An important question when conducting EFA is: how many factors should be retained? One commonly used method for deciding this is to retain all factors with an eigenvalue larger than one. An eigenvalue measures how much of the variance in the underlying variables the marginal factor can account for. An eigenvalue of one means the factor explains as much as one of the original variables. Intuitively then, an eigenvalue of less than one for the marginal factor suggests that the gain in terms of capturing more variation from the variables no longer outweigh the costs in terms of reduced parsimony. As can be seen from Figure 10 below, eight of the factors from the EFA have an eigenvalue of more than one. However, this method of choosing how many factors to retain relies on strong assumptions which cannot be empirically verified (Preacher & MacCallum, 2003). An alternative criterion is to retain as many factors as come before the last large drop, or "elbow point", on a plot of the marginal eigenvalues (Figure 10). The rationale for this is that the gradient of the plot captures the trade-off between parsimony (which is declining in the number of factors retained) and variance captured (which is declining in the eigenvalues). A discontinuous change in the gradient at an elbow point therefore reflects a sharp decline in the desirability of this trade-off and therefore a sensible place to optimise. In this case, both methods of determining how many factors to retain give the same answer: eight. I therefore retain the first eight factors.

Figure 10: Eigenvalues for each marginal factor in the EFA
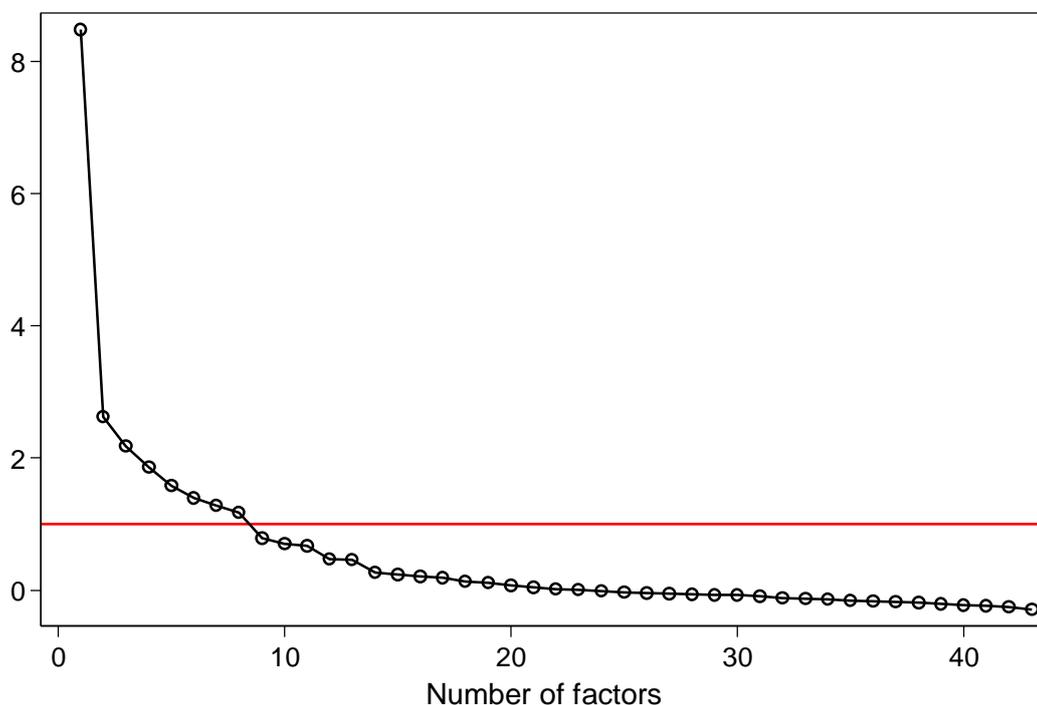
Notes: The red line indicates an eigenvalue of one.

I give each of the eight factors a name based on my own interpretation of what unifies the variables which make them up. They are: Leadership/Management; Teacher Cooperation (which is very similar to the variable of the same name used in Section 3); Feedback; Scope for Progression; Effective Professional Development (which is very similar to the variable of the same name used in Section 3); Preparation for Teaching Assignments; Discipline; Workload. Table 8 lists the specific questions associated with each factor and readers may wish to review these to satisfy themselves that the factors are coherent and meaningful. Between them, these eight factors are able to explain 94.5% of all the variation captured in the TALIS working conditions variables. Table 24 (Appendix C) shows the pairwise correlations for the eight factors, six of which have an absolute value of 0.3 or higher. The correlations between the different working conditions measures justify the use of oblique rotation in this case. The factors that emerge from the EFA are broadly consistent with what the existing literature has found to be important for determining job satisfaction and retention, e.g. school leadership (Boyd et al., 2011; Ingersoll, 2001; Ladd, 2011; Skaalvik & Skaalvik, 2011); teacher cooperation (Borman & Dowling, 2008; Boyd et al., 2011; Kraft et al., 2016; Reeves et al., 2017); discipline (Ingersoll, 2001); preparation for the job (Bogler & Nir, 2015; Donaldson & Johnson,

2010); and workload (Betoret, 2006; Collie et al., 2012; Ladd, 2011; Marinell et al., 2013; Smith & Ingersoll, 2004).

I use the information from the EFA to calculate predicted factor scores based on teachers' responses to the 42 manifest working conditions variables. Measuring each factor using a single number like this requires that the factor can indeed be measured on a unidimensional scale. Internal consistency (interrelatedness) of items is a necessary condition for such unidimensionality (Green et al., 1977). Cronbach's alpha (Chronbach, 1951) is a commonly used measure of inter-relatedness which calculates the mean of all split half reliabilities for items in the scale. The convention in the literature is to require a value of alpha between 0.7 and 0.9 on the grounds that anything lower suggests multidimensionality and anything higher suggests redundancy of items (Tavakol & Dennick, 2011). Table 29 shows the standardised alphas for the items that load on each factor. There is a debate about whether Alpha is an appropriate measure of reliability for two-item scales, so I also report the Spearman-Brown statistic for the discipline factor, as recommend by Eisinga et al. (2013). Workload has a very low alpha of 0.5. One plausible reason for the low reliability is that the sample contains both full-time (87%) and part-time (13%) teachers. However, recalculating the alpha only among full-time teachers gives the exact same value. In order to avoid using a scale with low reliability, and simplify making workload comparisons across full-time and part-time teachers, the models below use a single variable measuring whether teachers feel their workload is unmanageable on a four-point Likert scale from Strongly Disagree to Strongly Agree. Aside from Workload, the groups of manifest variables that measure each latent variable all have alphas above or very close to 0.7 (Feedback is 0.69) and are therefore retained.

Table 8: Wording of questions/variables that load on each factor

Factor	Wording of question which makes up that factor	TALIS Variable Code
Leadership/ Management	School provides staff with opportunities to participate in school decisions	44A
	I do not have the autonomy I need to do a good job as a teacher	47E
	Feedback provided based on a thorough assessment of teaching	31E
	The school has an effective school management team	47C
	The school management team give clear vision and direction	47D
	There is a collaborative school culture characterized by mutual support	44E
	There is a lack of employer support (for professional development)	27C
	The school has a culture of shared responsibility for school issues	44D
Teacher Cooperation	(How often) Teach jointly as a team in same class	33A
	(How often) Observe other teachers' classes and provide feedback	33B
	(How often) Engage in joint activities across different classes and age groups	33C
	(How often) Exchange teaching materials with colleagues	33D
	(How often) Engage in discussion about the learning of specific students	33E
	(How often) Work with other teachers to ensure common evaluations	33F
	(How often) Attend team conferences	33G
	(How often) Take part in collaborative professional learning	33H
Feedback	(Do you get) Feedback from student surveys about your teaching	28B
	(Do you get) Feedback following an assessment of your content knowledge	28C
	(Do you get) Feedback following a review of your students' tests scores	28D
	(Do you get) Feedback following self-assessment of your work	28E
	(Do you get) Feedback following surveys or discussions with parents	28F
Scope for Progression	I have scope to progress to a higher pay level	47L
	I have scope to progress into a leadership team role	47M
	I have scope to progress as a classroom teacher	47N
Effective Professional Development	(To what extent does PD) include a group of colleagues from my school/subject	25A
	(To what extent does PD) include opportunities to use active learning methods	25B
	(To what extent does PD) include collaborative learning with other teachers	25C
	(To what extent does PD) occur over several occasions spread out over weeks	25D
Preparation	Were contents of subjects you teach included in your formal education/training	12A
	Was pedagogy of subjects you teach included in your formal education/training	12B
	(Do you feel prepared for) Content of the subjects you teach	13A
	(Do you feel prepared for) Pedagogy of the subjects you teach	13B
Discipline	(Can you) Control disruptive behaviour in the classroom	34D
	(Can you) Get students to follow classroom rules	34H
Workload	(How many hours did you spend on) Your job last week	16
	(How many hours did you spend on) Planning and lesson preparation last week	18A
	(How many hours did you spend on) Marking/correcting students work last week	18C

Notes: For space reasons, not all questions are reproduced in full. Only questions with loadings >|0.3| are shown.

3.5 Modelling Using English Data

Table 9 shows the results from modelling teachers' desire to move school based on their personal characteristics, the characteristics of their school and the working conditions in their school. As in section 3, desire to move school is measured on a four-point scale from Strongly Disagree to Strongly Agree. All four models are therefore ordered logistic regressions and the coefficients show the percentage change in odds of being one category higher (closer to Strongly Agree), associated with a one unit change in the independent variable, holding the other variables constant. The demographic variables included in the model but not shown in the results are listed in the notes to Table 9.

Column 1 shows the relationship between the proportion of pupils who qualify for free school meals (FSM) and teachers' desire to move school, conditional on teacher demographic characteristics and school characteristics. It reveals the expected positive association, statistically significant at the 1% level. However in column 2, which also includes the seven working conditions, this association drops substantially and is no longer statistically significant. This is consistent with the findings from Simon and Johnson (2015). Four of the coefficients on the working condition variables reach statistical significance at conventional levels. The strongest relationship is with leadership. A one standard deviation in my measure of leadership reported by a teacher is associated with a 61% reduction in the odds that a teacher is one category higher (closer to Strongly Agree). This is very similar to the equivalent finding from Johnson et al. (2012) and Ladd (2011) who find a one SD increase in leadership is associated with a 60% and 52% reduction in the odds that a teacher intends to switch schools, respectively. I also find that a one SD increase in scope for progression is associated with a 28% reduction in the odds of being one category closer to Strongly Agree, and a one SD increase in Preparation is associated with a 13% reduction in the odds of being one category closer to Strongly Agree. Finally, moving one category closer to Strongly Agree that workload is unmanageable is associated with a 20% increase in desire to move school. This is comparable with the finding from Johnson et al. (2012) that a one SD improvement in workload is associated with a 37% reduction in the odds of intended turnover. The other four working conditions measures – collaboration, feedback, professional development and discipline – all show very little association with desire to move school.

Table 31 in Appendix C shows analogous results when the working condition variables are entered into the model one by one. The associations are generally slightly stronger than when added simultaneously, but remain in the same direction. Interestingly, the coefficients on leadership (preparation) are similar to those in column 2 when leadership (preparation) is entered into the model without the other working conditions or FSM. This suggests that the other working conditions measures and pupil deprivation have a separate relationship with job satisfaction, rather than being mediators or mechanisms between, e.g. leadership and satisfaction. In Table 32 (Appendix C) I run the same model using multiple imputation by chained equations to account for missing data on the covariates. The results are qualitatively very similar.

A serious concern when using survey data in this way is that the results will be affected by common source bias (Meier & O'Toole, 2010; Podsakoff & Organ, 1986). This occurs when both the dependent and independent variables contain measurement error due to being from a common source, such as a single questionnaire. The shared measurement error emanating from the common source in effect acts as an omitted variable and can cause large bias in the regression coefficients (Podsakoff et al., 2012). Favero and Bullock (2015) distinguish individual common source bias and environmental common source bias. In this setting, individual common source bias could result from the affective state of the teachers responding to the survey. For example, if a teacher is having a particularly bad day when they respond to the TALIS questionnaire, this might influence them to report more negatively on the working conditions in the school *and* report more negatively on their job satisfaction or desire to leave. A teacher having a particularly good day may do the opposite. This would inflate the estimate of the relationship between working conditions and the two outcome variables. Environmental common source bias may also be a concern in this setting if, for example, the organisational culture in a particular school affects the way that teachers interpret and respond to the TALIS questionnaire. Teachers in schools with a deferential culture, for example, may feel less inclined to give an honest assessment of leadership in their school, which would give a downward bias to the coefficient on the leadership variable.

I run two additional specifications of my model in Table 9 to test for common source bias. Column 3 includes school fixed effects to try and capture unobserved school-level environmental influences on teachers' survey responses. It also includes a measure of happiness at work as an independent variable, in order to try and capture individual-level time-varying affective state at the point a teacher responded to the survey. The coefficient on leadership and preparation are fairly stable and both remain statistically significant at the 5% level or higher. However, the association with scope for progression and workload become weaker and are no longer statistically significant. Column 4 takes a different approach to testing for individual-level common source bias by measuring working conditions W for teacher i in department j, using the mean of working conditions reported by other teachers in their department $\overline{W}_{i \neq i,j}$. The coefficient on leadership falls noticeably compared to column 2, but remains statistically significant at the 5% level. The coefficient on scope for progression switches sign and is no longer statistically significant at conventional levels. The coefficient on preparation remains stable and statistically significant at the 5% level. While neither of these additional specifications can eliminate the problem entirely (Favero & Bullock, 2015), they do provide some reassurance that the relationship between leadership, preparation and desire to move schools are robust to checks for common source bias. By contrast, the results from column 3 and 4 mean common source bias cannot be ruled out as the reason for the observed relationship between scope for progression and workload.

Table 9: Modelling desire to move school (ordered logit regression)

	(1)	(2)	(3)	(4)
FSM (%)	1.091**	1.008	0.918	1.002
	(0.007)	(0.007)	(0.050)	(0.007)
Leadership (Z score)		0.393**	0.565**	0.698*
		(0.031)	(0.052)	(0.102)
Collaboration (Z score)		0.946	0.902	0.995
		(0.064)	(0.095)	(0.111)
Scope for Progression (Z score)		0.723**	0.88	1.212
		(0.076)	(0.097)	0.186
Feedback (Z score)		1.072	1.122	0.966
		(0.083)	(0.098)	(0.116)
Prof Development (Z score)		1.016	1.005	1.052
		(0.085)	(0.099)	(0.14)
Preparation (Z score)		0.87*	0.817*	0.755*
		(0.047)	(0.049)	(0.085)
Discipline (Z score)		1.030	1.137	0.857
		(0.071)	(0.09)	(0.114)
Workload is Unmanageable (Likert†)		1.20*	1.01	0.99
		(0.1)	(0.099)	(0.058)
Enjoy Working at This School (Likert†)			0.146**	
			(0.025)	
Pseudo R Squared	0.04	0.14	0.89	0.04
N	1,933	1,508	1,505	785
Demographics	Y	Y	Y	Y
School Fixed Effects			Y	
Colleague Report				Y

Notes: Coefficients are odds ratios. Numbers in parentheses are standard errors. Standard errors are adjusted for clustering through the application of final and replicate weights. Demographic controls included in the model: female (dummy), age (years), full time (dummy), experience in teaching (years), pupils female (%), pupils ethnic minority (%), academy school (dummy). FSM = Free School Meals. Prof Development = effective professional development. Each column is a separate ordered logistic regression. Colleague report involves measuring working conditions W for teacher i in department j, using the mean of working conditions reported by other teachers in their department $\overline{W}_{i \neq i,j}$. Numbers in parentheses are standard errors. ** = p< 0.01, * = p< 0.05. † = measured on a four-point scale from Strongly Disagree (1) to Strongly Agree (4). Happiness with Job is derived from Question 46E "I enjoy working at this school". Workload is Unmanageable is derived from Question 47H "My workload is unmanageable". Pseudo R Squared is McFadden's Pseudo R Squared.

In order to give a sense of the material significance of these associations, Figure 11 shows predictive margins for the relationship between working conditions and whether a teacher wants to leave their school. In order to maximise interpretive clarity, the graphs below are based on a model identical to that in column 2 of Table 9, but with the outcome measured as a dummy indicating whether teachers agree (32.8% of respondents in England) or disagree (67.2% of respondents in England) that they would like to move schools. The model is therefore estimated using binary logistic regression. The graphs show the predicted probability of a teacher reporting that they either Disagree or Strongly Disagree that they would like to move to another school, for varying values of leadership, scope for

progression and preparation, when the other variables are set to their average values. The x axes are again measured as z scores and the y axes as probabilities between zero and one. Leadership shows by far the strongest correlation with desire to stay: a move from one SD below the mean to one SD above the mean is associated with around a 30 percentage point increase in the probability that a teacher desires to stay. The same change for scope for progression and preparation are associated with an increase of less than 0.1 in the probability that a teacher wishes to remain in their current school.

Figure 11: Predicted margins that a teacher wants to remain at their current school
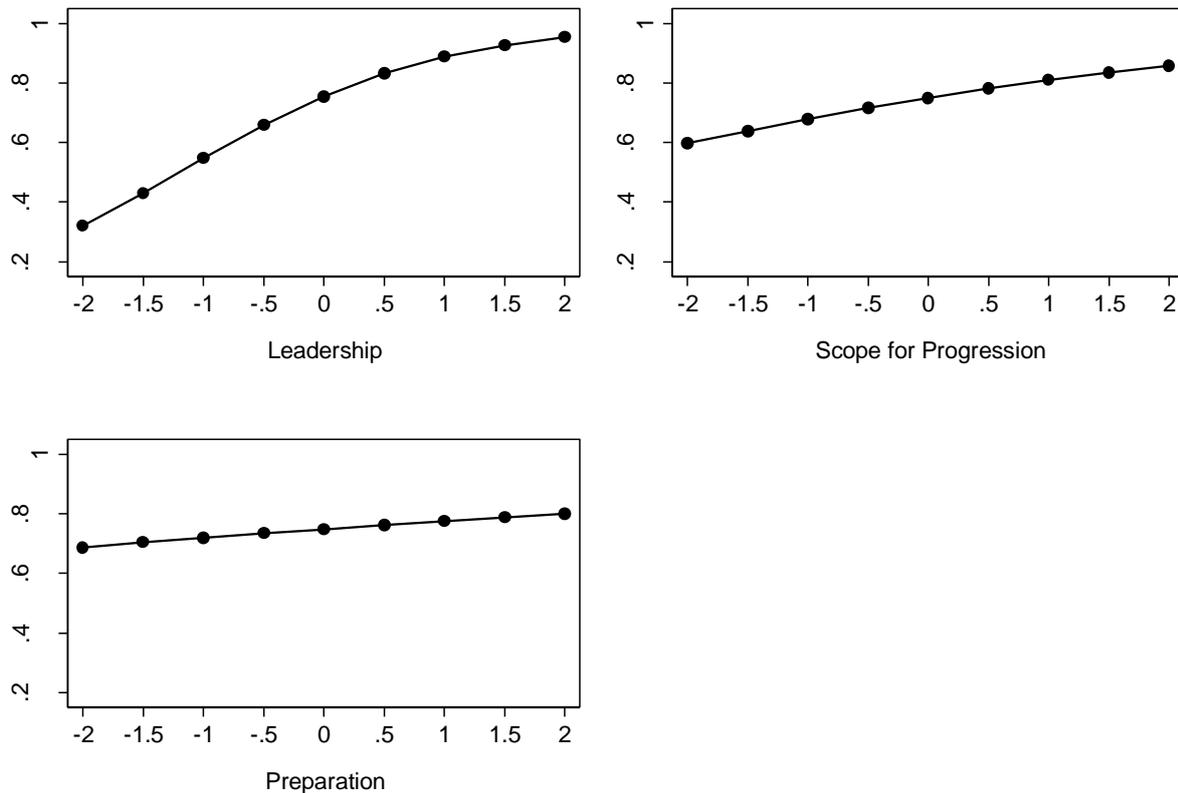
Notes: Shows the predicted margins from a binary logistic regression of whether or not a teacher wants to stay in their school (dummy) and the seven working conditions variables, plus the demographic and school variables included in column 2 of Table 9. All other variables in the model are evaluated at the mean (continuous variables) and mode (dummy variables). The x axes show z scores and the y axes show probabilities. 67.2% of respondents (after weighting) want to remain at their current school.

Table 10 repeats the analysis in Table 9 but using job satisfaction (z scored) as the outcome variable. Ordinary least squares is used to estimate the regression and the coefficients therefore show the standard deviation increase in job satisfaction associated with a one unit increase in each independent variable, conditional on the other independent variables. Unlike in Table 9, column 5 in Table 10 does not show any association between deprivation of the pupil intake and the outcome. Column 6 also shows no association, conditional on teacher demographics and school characteristics. In column 6,

Leadership again has the strongest association with the outcome, with a one SD increase in the former associated with a 0.38 SD increase in job satisfaction. This is comparable to the finding from Johnson et al. (2012) that a one SD increase in leadership is associated with a 0.51 SD increase in job satisfaction. This association is robust to the addition of school fixed effects in column 7 and using colleague scores for the independent variables in column 8. In column 6 a one SD increase in scope for progression and discipline is associated with a 0.07 SD and 0.18 SD increase in job satisfaction, respectively. The workload finding is comparable to Johnson et al. (2012) who find a one SD improvement in workload is associated with a 0.26 SD improvement in job satisfaction. Both coefficients are similar in column 7 and remain statistically significant at the 5% level. However in column 8 they change sign and are not statistically significant, which is consistent with the presence of individual-level common source bias. The coefficient on workload follows a similar pattern, remaining stable across column 6 and 7, but changing in magnitude and no longer being statistically significant at conventional levels in column 8. The coefficient on preparation is positive and fairly stable (between 0.03 and 0.1) across all models but is not statistically significant at conventional levels in column 6 or 7. Table 30 in Appendix C shows the results when the working conditions latent variables are added one-by-one. The associations are generally slightly stronger than when the variables are added simultaneously, but do not change sign. Table 32 in Appendix C shows the results from column 6 after imputation has been used to account for missing data on the covariates and the results are again qualitatively very similar to the complete case version.

Table 10: Modelling job satisfaction (OLS regression)

	(5)	(6)	(7)	(8)
FSM (%)	-0.003	-0.004	-0.005	-0.002
	(0.003)	(0.002)	(0.015)	(0.004)
Leadership (Z score)		0.379**	0.377**	0.159*
		(0.034)	(0.037)	(0.059)
Collaboration (Z score)		0.026	0.055	0.013
		(0.031)	(0.034)	(0.06)
Scope for Progression (Z score)		0.183**	0.180**	-0.078
		(0.029)	(0.034)	(0.074)
Feedback (Z score)		-0.045*	-0.055*	0.047
		(0.025)	(0.026)	(0.057)
Prof Development (Z score)		-0.022	0.020	-0.078
		(0.025)	(0.028)	(0.059)
Preparation (Z score)		0.041	0.026	0.111*
		(0.025)	(0.027)	(0.049)
Discipline (Z score)		0.073*	0.073**	-0.036
		(0.024)	(0.024)	(0.051)
Workload is Unmanageable (Likert†)		-0.305**	-0.287**	-0.03
		(0.033)	(0.037)	(0.029)
R Squared	0.015	0.41	0.50	0.04
N	1,945	1,510	1,510	758
Demographics	Y	Y	Y	Y
School Fixed Effects			Y	
Colleague Report				Y

Notes: Demographic controls include in the model: female (dummy), age (years), full time (dummy), experience in teaching (years), pupils female (%), pupils ethnic minority (%), academy school (dummy). FSM = Free School Meals. Prof Development = effective professional development. Each column is a separate ordinary least squares regression. Colleague report involves measuring working conditions W for teacher i in department j, using the mean of working conditions reported by other teachers in their department $\overline{W}_{i \neq i,j}$. Numbers in parentheses are standard errors. ** = p< 0.01, * = p< 0.05. † = measured on a Likert scale from Strongly Disagree (1) to Strongly Agree (4). Happiness with Job is derived from Question 46C "I enjoy working at this school". Workload is Unmanageable is derived from Question 47H "My workload is unmanageable".

3.6 Conclusion

Consistent with research from the USA (Simon & Johnson, 2015), the results show that the disadvantage of a school's intake does not have an association with either reduced teacher job satisfaction or increased turnover intentions, once working conditions have been controlled for. This has important implications for policy since it suggests that there is potential for improving teacher retention by improving working conditions in schools.

This research also highlights which aspects of working conditions are most important. The nature of leadership and management in a school consistently emerges as having the strongest association with both satisfaction and desire to move school. This association is robust to a number of checks for common source bias and is consistent with other research using different datasets (Boyd et al., 2011; Ingersoll, 2001; Ladd, 2011). The manifest variables which load most heavily on the leadership latent

variable (see Table 8 and also Table 28 in Appendix C) emphasise leadership that sets a clear direction and vision for the school, whilst also allowing opportunities for teachers to participate in decision making and support each other. While this research is not able to identify which aspects or types of leadership are most important, this combination of direction setting and enabling leadership behaviours is consistent with other research on the types of leadership that are the best predictors of teacher job satisfaction and commitment to their schools (Nguni et al., 2006; Bogler, 2001). Interestingly, four of the six pairwise correlations among the working conditions latent variables include leadership, which suggests that the other working conditions may be mediators between leadership and the outcomes. Further research should test the strengths of both the direct and mediated effects of leadership on job satisfaction and retention.

The relationship between preparation and desire to move school is also robust across my model specifications. As can be seen from Table 8, the manifest variables that make up the preparation latent variable focus on whether teachers have had formal or informal training in the subjects they teach. This is consistent with findings from Donaldson & Johnson (2010) who use longitudinal survey data to show that early-career teachers who are asked to teach outside of the subject in which they have been trained are more likely to leave both their school and the profession. This suggests that school leaders should try to give teachers, particularly those with less experience, assignments in the subjects in which they have a degree. Where this is not possible due to staffing shortages, additional subject training should be provided.

Several other variables were predictive of teacher job satisfaction, such as scope for progression (whether teachers feel they can gain new skills and promotions), discipline and workload. However, while all three of these variables had a statistically significant relationship with job satisfaction conditional on school fixed effects, they did not show an association when the working conditions variables were measured using colleague reports. This is suggestive of individual-level common source bias. However it may also be due to intra-departmental variations in working conditions. Scope for progression, for example, may be very different for teachers in the same department. Other studies have found that discipline (Ingersoll, 2001); and workload (Betoret, 2006; Collie et al., 2012; Ladd, 2011; Marinell et al., 2013; Smith & Ingersoll, 2004) are related to retention. Nevertheless, future research should therefore identify alternative data sources for measuring the outcome variables that do not share measurement error with the independent variables, such as administrative data on observed turnover and retention.

Despite these limitations, these results should be interpreted with respect to the limitations of this data. In particular, the cross-sectional nature of the data means that I cannot rule out changes in working conditions preceding changes in retention or job satisfaction, e.g. head teachers assigning teachers they expect to leave to classes for which the school has no appropriately trained teachers.

Interpretation of the results in this paper therefore relies on two other sources of evidence. First, psychologists have developed theoretical accounts (Bakker & Demerouti, 2007) which specify the mechanisms by which poor working conditions increase the demands on teachers (e.g. having to teach courses for which they are not prepared) and/or reduce the resources (e.g. help from colleagues) available to them. This imbalance between resources and demands leads to a depletion of energy levels and eventually results in teacher burnout and turnover (Crawford et al., 2010; Fernet et al., 2013). Second, other research using longitudinal data has shown that changes in working conditions precede changes in teacher retention outcomes. For example, the quality of school leadership (Boyd et al., 2011) and teachers assigned to classes for which they have not had appropriate training (Donaldson & Johnson, 2010) is predictive of turnover measured in subsequent years. Similarly, Kraft et al. (2016) find that school leadership, teacher collaboration and orderliness predict turnover at later time points, not the other way round. Theory and empirical results therefore help reduce concerns about reverse causality explaining the correlations in my cross-sectional data.

As well as potential reverse causality, there are also questions about mediation in the models. Table 27 in Appendix C shows that leadership has non-trivial correlations with four of the other working conditions latent variables. If the other working conditions are mediators or mechanisms through which leadership affects job satisfaction and desire to leave, rather than just covariates, then the coefficients in my regression can no longer be interpreted as ceteris paribus associations. In order to address both of these limitations, future research should utilise panel data on working conditions both to control for unobserved potential confounders and to try and rule out this sort of reverse causality.

This research can be of direct use to school leaders looking to improve the morale and retention of their teachers. Ultimately however, school leaders can only act on the findings of this research if they have the skills and resources to go about improving working conditions. It may therefore be necessary to find ways of sharing good practice between schools with variation in the quality of their working conditions. The provisions of interventions like those studied in Fryer (2017) and Jacob et al. (2015), would also help leaders to improve teachers' working conditions. Such training interventions also provide an opportunity to go beyond the conditional associations documented in this paper to identify causal effects using experimental research designs.

Chapter 4: Does Subject-Specific Professional Development Improve the Retention of Science Teachers?

4.1 Introduction

Teacher shortages are a persistent and widespread problem in public school systems and are particularly severe among science teachers (Dolton, 2006). While other subjects generally see shortages eliminated during economic downturns, shortages of science teachers tend to persist between economic cycles (Goldhaber et al., 2014b; Smithers & Robinson, 2008). Where no appropriately qualified teachers are available, research shows that school leaders tend to lower recruitment standards, make increased use of temporary teachers or increase class sizes (Moor, 2006; Smithers & Robinson, 2000), all of which have been linked with reduced pupil attainment (Fredriksson et al., 2012; Mocetti, 2012; Schanzenbach, 2006). Finding ways to reduce the shortage of appropriately qualified science teachers is therefore important.

The shortage of these teachers is due in part to the higher rates at which science teachers leave the profession (Kelly, 2004; Worth & De Lazzari, 2017). This in turn reflects the fact that teachers with STEM degrees face a higher outside pay ratio than other teachers (MAC, 2016), providing them with monetary incentives to move to jobs in other sectors. Several evaluations have shown that increasing science teacher pay towards what science teachers could earn outside of the profession has a positive effect on retention, at least in the short term (Bueno & Sass, 2016; Clotfelter et al., 2008; Feng & Sass, 2016).

Science teaching is also in some ways more demanding than teaching in other subjects, in that science teachers generally have to teach one or two subjects in which they do not have a degree. For example, a chemistry graduate will generally be expected to teach biology and physics, as well as chemistry. The high numbers of physicists choosing to teach mathematics instead of mixed science (Smithers & Robinson, 2008), suggests perhaps that this is seen as undesirable by teachers. Early-career teachers who are given multiple subjects to teach are also more likely to leave the profession (Donaldson & Johnson, 2010). The demands of teaching science therefore also help to explain the higher levels of wastage (leaving the profession) among science teachers.

Concerns about teacher retention have led to calls for improved continuing professional development (CPD) as a way of improving retention (Borman & Dowling, 2008). Detailed qualitative research has been used to suggest that professional development can improve retention through improving teachers' knowledge, self-efficacy and motivation (Coldwell, 2017; Taylor et al., 2011; Wolstenholme et al., 2012). However, the limited quantitative research which has been conducted in this area has so far yielded mixed findings. Kraft et al. (2016) use linked survey and administrative data and find that professional development is predictive of teacher turnover. An evaluation by Knibbs et al. (2017) also finds that a two-year intensive professional development programme improved retention. However, a randomised controlled trial of an intensive mentoring and professional development programme for beginning teachers in the US found no effects on retention (Glazerman et al., 2010). In Chapter 3 of this thesis, I also did not find a robust relationship between CPD and teachers' job satisfaction or

desire to move school. Further research is therefore required to identify which types of CPD do and do not help with retention.

Reviews of research on CPD have recommended that programmes should focus on developing teachers' subject knowledge, as well as how to teach that knowledge (Blank & de las Alas, 2009), making this one promising avenue for exploration. Unfortunately, research on the links between subject-specific professional development and retention is even more limited. Subject-specific professional development has been shown to improve teachers' knowledge of their subject (Goldschmidt & Phelps, 2010; Kutaka et al., 2017; Polly et al., 2015; van Driel et al., 2012) and increase the self-efficacy of both science (Lakshmanan et al., 2011) and maths teachers (Ross & Bruce, 2007). However, very little research has been conducted on whether this ultimately feeds through into improved retention. Reviews of the literature on science CPD have therefore called for more research on the importance of subject knowledge and effects of subject-specific CPD on other outcomes, including retention (Luft et al., 2015; van Driel et al., 2012).

I address this gap in the literature by evaluating the effect of science teacher participation in National STEM Learning Network (NSLN) courses in England. NSLN was set up in 2005 to provide courses aimed at improving science teachers' subject and pedagogical knowledge, career progression and, ultimately, retention in the profession (Wolstenholme et al., 2012). Courses vary from one day (71% of all records) to residential courses of three or more days in length (11%). Many courses are free to participants and there is a generous system of bursaries, which is jointly funded by charitable and public sector bodies. Courses cover a range of subjects from classroom management through to astronomy, but most focus on science content that is directly relevant to the national science curriculum. NSLN is delivered through a mixture of regional networks and the National STEM Learning Centre in the North of England, giving it broad coverage: 25% of all science teachers in England attended at least one course between the 2010/11 and 2012/13 school years. Wolstenholme et al. (2012) used data from a survey of 519 participants and telephone interviews with a further 25 to develop a detailed programme theory for the NSLN intervention, describing the mechanism or channels through which it is hypothesised to affect teacher retention. Their data suggests that participation in NSLN would affect retention in the profession through two routes: increased motivation and job satisfaction, and increasing teacher knowledge and self-efficacy.

My first set of results come from teacher-level cross-section regression models, which show that participation in NSLN courses is associated with increases in the odds of both retention in the participant's original school and in the profession overall. However, these findings are open to the obvious objection that unobservable variables that affect both participation in the programme and retention in the profession are confounding these estimates. For example, more motivated teachers may be both more likely to take an NSLN course and more likely to remain in the profession. In order to address this limitation I create a balanced, stacked panel of science departments and estimate double-difference models (comparing changes in retention in participating science departments to changes in retention in similar non-participating science departments) which show that participation is

associated with a reduction in wastage of around two percentage points, two years after first-participation. Estimates from triple-difference models (which further net out changes in retention in English departments in schools where the science department participates) also show that participation is associated with a reduction in wastage of around three percentage points. This is equivalent to around a third of average science department wastage, which was 9% in 2011. A placebo test in the year prior to treatment provides no reason to doubt the common trends assumption. This paper is therefore one of the first to provide evidence that subject-specific CPD can improve science teacher retention.

The rest of the paper is structured as follows: section 2 describes the data, section 3 presents the results from teacher-level regression models, section 4 presents the results from the department-level double- and triple-difference models, and section 5 concludes by discussing the implication for policy and future research.

4.2 Data and Descriptive Statistics

The first source of data used for this research is a record-level dataset of every time a person participates in an NSLN course between the 2010/11 and 2012/13 academic years. This data was supplied by NSLN. It contains an identifier for each participant, the school they were working in at the time, and the length and subject-focus of the course they attended. Individuals are free to participate in multiple NSLN courses if desired. Most participants (61%) only attend one NSLN course but many do attend more than one, and 10% attend four or more. Figure 13 in Appendix D shows the full distribution of dosage across participants, across all years in the data, and in the first year that an individual participates. In this analysis, I look at the effect of first observed participation one and two years after the first year that a teacher or department first participates. This has the disadvantage of ignoring variation in dosage that occurs through repeated participation. However, it also has the important advantage of avoiding the endogeneity of repeat participation, in the sense that those participants who have a positive first experience of NSLN are more likely to participate in future, and vice versa. In order to investigate the role of dosage, I investigate the effect of different numbers of days attended within the first year of participation in some of my models, which should keep the endogeneity of repeat participation to a minimum, while providing some insight into dose-response patterns.

The second source of data is National Pupil Database (NPD) and School Workforce Census (SWC) administrative data. Information on pupils and schools comes from the NPD, which is an administrative dataset containing information on all pupils in England since before 2010. This includes information on school type (e.g. academy), school location, pupil demographics and prior attainment, and school Ofsted rating. Information on teachers is taken from the SWC, which contains information on teachers' demographic characteristics (e.g. age, ethnicity, gender), professional

characteristics (e.g. years of experience, pay, degree subject) and class assignments, including which subjects they teach. In total, 3,016 records of the 26,776 (11%) obtained from NSLN could not be matched to teachers in the SWC. A large part of this is likely because NSLN courses are open to (non-teaching) science technicians working in schools and other science educators not based in publicly-funded schools, neither of whom would show up in SWC data.

The SWC also provides the information from which I construct my primary outcome variable, wastage, and my secondary outcome variable, turnover. Wastage measures teachers leaving the profession. When I am analysing individual-level data, the wastage variable takes the value one in period *t* for any individual who is teaching in a state funded school in England in period *t-1* but is not in period *t*, and zero otherwise. When I am analysing data at the departmental level, the wastage variables for department d in period *t* is defined as the proportion of teachers who were teaching in department d in period *t-1* who are no longer teaching in department d in period *t*. Turnover measures teachers leaving their school. When I am analysing individual level data, the turnover variables takes the value one in period *t* for any teacher who is teaching in school s in period *t-1* but is no longer teaching in school s in period *t*, and zero otherwise. When I am analysing data at the departmental level, the turnover variable for department d in period *t* is defined as the proportion of teachers who were teaching in period *t-1* in department d who are no longer teaching in department d in period *t*. Figure 12 shows the retention of the 2010 cohort of newly qualified teachers (NQTs) in their original 2010 school and in the profession overall. It shows that science teachers leave their original school and the profession at a slightly faster rate than teachers of other subjects.

Table 11 compares the characteristics of science teachers in general, science teachers who participate in NSLN courses and science teachers who participate in more than two days of NSLN courses in total. Compared to science teachers in general, NLSN participants are more likely to be male, unqualified, or on a temporary contract and tend to be younger, less experienced, have been at their school for less time and are paid slightly less. They also tend to work in schools with more disadvantaged pupils and slightly lower attainment. Heavy NSLN participants (those that take part in two days or more) are even younger and less experienced. This suggests that NSLN courses may be being used strategically to improve the knowledge of junior science teachers. These observable characteristics are controlled for in all models in sections 4.3 and 4.4. Information on the distributions of these variables for participants can be found in Table 33 in Appendix D.

Figure 12: Science and non-science teacher retention for the 2010 NQT cohort
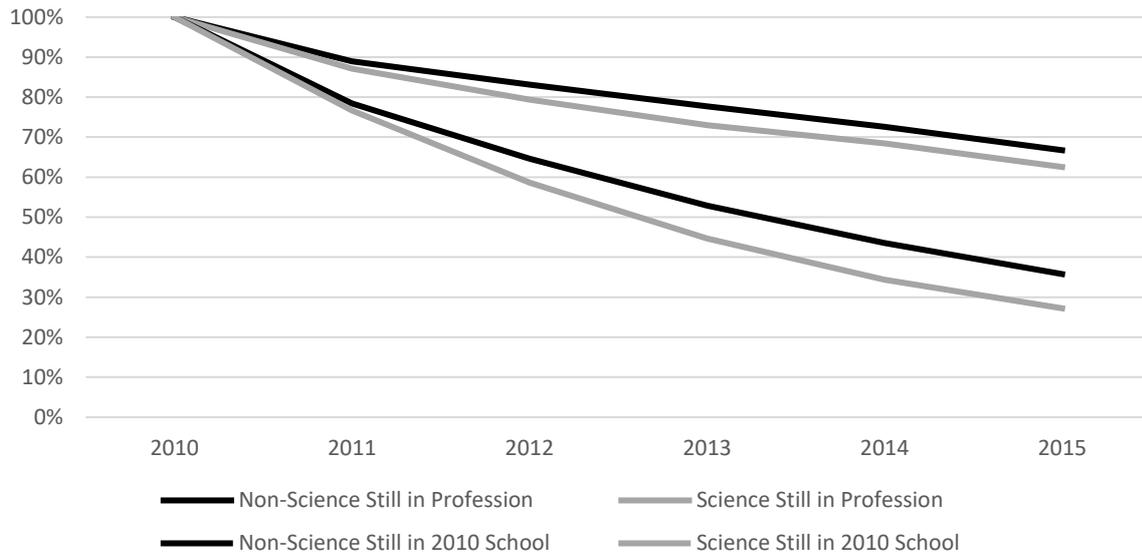
Notes: N = 16,048. The 2010 NQT cohort are the teachers who received newly qualified teacher status in the 2010/11 academic year. Teachers still in 2010 school are those who remain in the school in which they worked in the 2010/11 academic year.

Table 11: Comparing science teachers by participation levels

	Non-NSLN Science Teachers	NSLN Science Teachers	Heavy NSLN Science Teachers
NSLN Days Attended 2010-13 (mean)	0	1.2	6.9
Female (%)	63.4	17.7	8.3
Age (mean)	39.4	37.3	36.5
Qualified Status (%)	63.4	17.7	8.3
Permanent Contract (%)	63.4	17.8	8.4
Years Since Qualified (mean)	12.6	10.3	9.3
Years at Current School (mean)	6.9	5.9	5.4
Annual Pay £1,000 (mean)	38	35.4	34.7
Free School Meals (%)	14	16	16
Best 8 GCSE Point Score (mean)	353	349	347
N	22,672	6,342	2,980

Notes: Table includes only science teachers who were in service in 2012. Science teachers are those who teach science for most of their timetabled hours. Heavy participants are those that take part in more than two days of NLSN courses between 2010 and 2013.

4.3 Individual-Level Cross-Section Models

I begin by using logistic regression to model the log odds that teacher i exits either their school or the profession based on participation in an NSLN course (P_i), characteristics of the teacher (T_i), their school (S_i) and the calendar-year in which they first participate (Z_i), which captures year-specific unmeasured influences on retention:

1) $\log\left(\frac{p(exit_i)}{1-p(exit_i)}\right) \beta_0 + \beta_1 P_i + \beta_2 T_i + \beta_3 S_i + \beta_4 Z_i + \varepsilon_i$

Table 12 shows the results from several variants of this model. While I am analysing individual level data, both the wastage and turnover variables are binary, so all models use logistic regression and all coefficients are odds ratios. Standard errors are clustered at the school level. Column 1 reports the results of a regression of participation in NSLN courses on retention in the profession after one year. Columns 2 and 3 both report the results of regressing participation on retention in the profession after two years, but column 3 uses a high-dosage (>2 days in the first year) indicator of participation. Columns 4 to 6 repeat this pattern but with retention in original school as the outcome, instead of retention in the profession.

Looking first at the covariates in the lower panel of Table 12, the results corroborate previous findings that science teachers have much lower retention than teachers of other subjects. Looking at columns 2 and 5, I find that the odds of a science teacher remaining in their school or in the profession overall are lower than all other subjects, including maths. More specifically, the odds of a science teacher remaining in the profession two years later are 21% lower than all other subjects (excluding maths) and the odds of a science teachers remaining in their original school are 32% lower than all other subjects (excluding maths). Both of these findings are statistically significant at the 1% level. As can be seen from Table 34 and Table 35 in Appendix D, the other variables all enter the model with the expected sign. Higher pay, higher school inspection (Ofsted) ratings, higher pupil attainment and more experience are all associated with higher retention; while a higher proportion of deprived pupils is associated with reduced retention.

Turning to the coefficients on the treatment variables, there is a clear pattern visible. The first row of the table shows that first observed participation is associated with an 87% increase in the odds that a teacher is still in the profession one year later and a 43% increase in the odds that a teacher is still in the profession two years later. Heavy participation is associated with a 65% increase in the odds of being in the profession two years later, which is consistent with a positive dose-response gradient. All three of these coefficients are statistically significant at the 5% level or higher. By contrast, there is no clear relationship between participation and retention of teachers in the school they were working in at the time of participation, even among those heavy participants taking two or more days of courses.

Table 12: Individual-level cross-section regression results

	Remain in Teaching			Remain in Original School		
	1 Year Later (1)	2 Years Later (2)	2 Years Later (3)	1 Year Later (4)	2 Years Later (5)	2 Years Later (6)
NSLN Participation	1.869**	1.426**		1.174	0.920	
	(0.216)	(0.118)		(0.091)	(0.053)	
Heavy NSLN Participation			1.645*			1.057
			(0.318)			(0.140)
Subject Taught (ref = other)						
Science	0.777**	0.793**	0.774**	0.720**	0.685**	0.677**
	(0.036)	(0.029)	(0.029)	(0.026)	(0.019)	(0.020)
Maths	1.040	1.079	1.077	0.821**	0.810**	0.810**
	(0.043)	(0.035)	(0.035)	(0.025)	(0.019)	(0.020)
Has Science Degree	0.961	0.980	0.986	1.053	1.109**	1.114**
	(0.035)	(0.028)	(0.028)	(0.029)	(0.024)	(0.025)
N	100,009	100,007	95,276	100,009	100,009	95,278
Pseudo R-Squared	0.052	0.054	0.054	0.049	0.039	0.039

Notes: Each column is a separate logistic regression on individual-level teacher data. Coefficients are odds ratios. All models control for the following teacher and school characteristics: teacher gender, teacher experience, whether the teacher has a permanent contract, their fulltime equivalent hours, pay, the proportion of pupils at the school on free school meals, the proportion of pupils who are from an ethnic minority, the school's Key Stage 4 attainment measured using Best 8, the school's prior (KS2) attainment, and the school's Ofsted grade. Also included in the regression are region dummies, urban/rural indicators and calendar year dummies. A full version of the regression output can be found in Table 34 and Table 35 in Appendix D. Covariates for non-participating (control) individuals all take 2010 values. Standard errors are shown in parentheses and are clustered at the school level. ** = $p < 0.01$, * = $p < 0.05$. Pseudo R Squared is McFadden's Pseudo R Squared.

The results from these models are informative, but also have important limitations. First, the models are likely to suffer from omitted variable bias, not least because they do not control for working conditions (see Chapter 3 of this thesis). My data does not contain measures of these variables, so they cannot be explicitly controlled for. Second, and relatedly, to identify the causal effect of the programme, this model relies on the strong assumption of selection-on-observables. In order to address these weaknesses, in the next section I exploit the longitudinal nature of my data and introduce a number of panel data techniques.

4.4 Department-Level Panel-Data Models

I begin by aggregating the data to department-level by assigning teachers to departments based on the subject in which they do most of their teaching. This has several advantages for the analysis. First, because departments do not leave the data when turnover occurs, it allows me to observe turnover and wastage across a stable set of units. Second, it helps minimise concerns around the endogeneity of which individuals departments choose to send on NSLN course. Third, it allows me to utilise

continuous variation in the outcome variables, which are now measured as departmental turnover and wastage rates.

I create a balanced, stacked, three-year panel of science departments. For treated departments, variables in the first year of the panel take the values from the calendar year in which they first participated in NSLN, variables in the second year of the panel take the values from the year after first-participation occurs and variables in the third year of the panel take the values of variables from the second year after participation occurs. Ideally, I would include more pre-intervention years in the panel, as this would allow me to visually assess common trends over multiple periods (see Wong et al., 2015) and potentially allow me to relax the common trends assumption by modelling heterogeneous linear trends (Angrist & Pischke, 2014). Unfortunately however, the SWC data only begins in 2010, meaning that turnover and wastage can first be calculated in 2011 (with 2010 acting as the *t-1* period). Including two pre-intervention years in my panel would therefore require dropping departments which first participated in 2011 from my panel. Given that I only have data on departments first participating up to 2012, this would have undesirable consequences for the precision of my estimates. I therefore stick to having just one pre-intervention period in my panel.

When constructing a stacked panel it is obvious that treated units should use data from both the pre-intervention and post-intervention period. However, an important question remains about which years of data should be used by untreated units in each of the three panel years. My approach is to use matching to construct my stacked panel. I first take all departments which first participated in 2011 and match them on 2011 characteristics to departments which I never observe participating using nearest neighbour matching and a caliper of 0.1. I then apply the propensity score weights which effectively drops all units which could not be matched. All untreated departments still in the data then use information from the variables in 2011, 2012 and 2013 in the first, second and third panel year respectively. I then repeat this process for departments which first participated in 2012, with department which I never observe participating using information from the variables in 2012, 2013 and 2014 in the first, second and third panel year respectively. Figure 14 and Figure 15 in Appendix D show the estimated propensity scores, provide information on how they were estimated, and illustrate the extent of common support. I do not present balance tests because I rely on regression analysis for covariate adjustment. The matching is only intended to 1) provide a principled way of constructing the balanced panel for never-treated departments and 2) enforce common support. This leaves me with a balanced, stacked, three-year, department-level panel.

I use a difference-in-difference strategy (Equation 2 below) to analyse this panel. The outcome (Y_{dt}) is now the proportion of teachers that leave department d in year t. The T, S and Z terms all now have d and t subscripts to indicate that they are measured at the department level (d), in a given school year (t). The calendar year term now allows me to capture differences in year-specific changes in turnover

and wastage across the years included in my stacked panel. Equation 2 also includes departmental fixed effects (D) which capture time-invariant, department-level characteristics. I argue that within the narrow (three year) window of observation in my panel this helps to capture important aspects of working conditions as well as, for example, the level of motivation of teachers working in that department. Equation 2 also includes a panel-year dummy (K) which takes the value of 0 in the pre-treatment period and 1 in the treatment period. The final change I make in Equation 2 is to add an interaction term between the participation dummy (P) and the panel year term (K) which takes the value of 1 for participating departments in the treatment period and zero otherwise. In practice, I have two K dummies and two interactions, one for the year after treatment and one for the second year after treatment. The coefficient β_3 on these interactions shows the association of retention with participation, conditional on covariates and changes in retention in other non-participating science departments. Under the assumption that retention in participating and non-participating departments conditional on covariates would have changed in the same way over this period in the absence of the treatment (common trends), β_3 is therefore the causal effect of first observed participation on NSLN courses. By using non-participating science departments as a control structure, this specification allows me to account for unmeasured, time-varying factors which are common between participating science departments and non-participating science departments, such as changes in national policy relating to science teaching.

$$2)\ Y_{dt} = \beta_0 + \beta_1 P_d + \beta_2 K_t + \beta_3 (P_d * K_t) + \beta_4 T_{dt} + \beta_5 S_{dt} + \beta_6 D_d + \beta_7 Z_{dt} + \varepsilon_{dt}$$

Utilising non-participating science departments as a control structure accounts for a number of important unmeasured variables which are common between schools. However, it does not account for unmeasured variables which differ between schools, such as the quality of senior leadership. Fortunately, because NSLN courses are only available to science educators, I can also exploit other untreated departments within the same school as the participating science department as an additional control structure. English and maths departments are both good candidates for this, since they share with science departments the characteristic of being compulsory subjects, meaning the pupils in those departments do not vary within a school due to the subject choices made by pupils. However, using maths departments as the additional control structure is not desirable, since many NSLN participants also teach in maths departments, meaning this could introduce treatment contamination. I therefore use English departments, adding an additional dummy (SE) to indicate whether a department is an English department. Adding two further two-way interactions and the three way interaction between P, K and SE gives the triple-difference specification in Equation 3. β_7 is now the coefficient of interests, since it shows the association between retention and being in a participating science department, conditional on both changes in retention in non-participating science departments in other

schools and English departments in participating schools. Under the assumption that all variation in the outcome measure not due to observables is either shared between the participating science department and the matched science department, or between the participating science department and the English department in the same school, β_7 shows the causal effect of participating in NSLN courses. By using English departments as an additional control structure, this specification also allows me to account for unmeasured, time-varying factors which are common between participating science departments and non-participating English departments in the same school. This specification is therefore able to account for, e.g. the changing quality of senior leadership in a school over time. Figure 16 in Appendix D provides a graphical illustration of the rationale for the double- and triple-difference approaches.

$$3) \; Y_{dt} = \beta_0 + \beta_1 P_d + \beta_2 K_t + \beta_3 SE_d + \beta_4 (P_d * K_t) + \beta_5 (P_t * SE_d) + \beta_6 (K_t * SE_d) + \beta_7 (K_t * SE_d * P_t) + \beta_8 T_{dt} + \beta_9 S_{dt} + \beta_{10} D_d + \beta_{11} Z_{dt} + \varepsilon_{dt}$$

Table 13 shows the results from these models. The outcome variables are now continuous and the regressions are therefore estimated using ordinary least squares. The coefficients show the percentage point change in turnover/wastage associated with a one unit increase in the relevant independent variable. Column 7 shows the results from the double-difference turnover model. The coefficients are very close to zero and not statistically significant at conventional levels. The results in the triple-difference turnover model (Column 8) are also not statistically significant at conventional levels. The results from the departmental turnover models are therefore consistent with those from the individual logistic regression models in showing no effect.

Column 9 shows the results of the double-difference wastage model. One year after treatment the coefficient is again close to zero and not statistically significant at conventional levels. Two years after treatment however, participation is associated with a 2 percentage point reduction in wastage, statistically significant at the 1% level. In the triple-difference wastage model in column 10 participation is associated with a 2 percentage point reduction in wastage one year after participation, though this is not statistically significant at conventional levels. The coefficient on the treatment effect two years after participation shows a 3 percentage point reduction in wastage, which is again statistically significant at the 1% level. In summary, the double- and triple-difference models suggest that participation in subject-specific CPD reduces departmental wastage by between 2 and 3 percentage points after two years. This is equivalent to around a third of the average wastage rate for science departments, which was 9% in 2011. While participation is associated with improved

retention in the profession, it is not associated with increased retention in a participant's original school. The panel data wastage models also show a similar pattern of results to the individual level wastage models. Full output from the models in Table 13 can be found in Table 36 in Appendix D.

Table 13: Percentage point change in turnover/wastage associated with NSLN participation

	Turnover		Wastage	
	Double-Diff	Triple-Diff	Double-Diff	Triple-Diff
	(7)	(8)	(9)	(10)
One Year After Participation	-0.004	0.0004	-0.008	-0.018
	(0.010)	(0.012)	(0.006)	(0.008)
Two Years After Participation	-0.006	0.027	-0.021**	-0.029**
	(0.010)	(0.012)	(0.006)	(0.008)
N	4,579	4,579	4,579	4,579
R-Squared	0.69	0.74	0.59	0.64
Placebo (Year of Participation)	0.006	0.006	-0.005	-0.014
	(0.015)	(0.019)	(0.010)	(0.013)

Notes: Each column is a separate OLS regression. Also included in all regressions are region dummies, urban/rural indicators and calendar year dummies, as well as all of the time-varying covariates from the models in Table 11. Standard errors are shown in parentheses and are clustered at the school level. ** = p< 0.01, * = p< 0.05. N = number of groups. N for the placebo test is 1,595. Those who first participate in 2011 cannot be used in the placebo test as there is no turnover and wastage data for 2010 due to it being the first year of the data.

In order to identify the causal impact of the programme on retention, double-difference and triple-difference models rely on the assumption that the outcomes for participating departments and control departments move along common trends during the treatment period, conditional on observables. This assumption is necessary because these models cannot control for time-varying factors which are not common between treatment and control observations. A placebo test aims to indirectly falsify the common trends assumption by looking at whether there are differences in retention prior to the treatment occurring, conditional on the other covariates. If differences can be detected in the year prior to treatment occurring then it is implausible that the common trends assumption holds during the treatment period. If no difference can be detected in the year prior to treatment then this does not provide evidence against the common trends assumption. In order to test this I create a new stacked panel in which the first year of the panel contains data on departments in the year prior to first-participation (rather than the year of first-participation) and the second year of the panel contains data on departments in the year that first-participation occurs (rather than the year after). I then run the same models and look for any differences in the trajectory of turnover and wastage in this period. The bottom panel of

Table 13 reports the results of the placebo test, with coefficients very close to zero and not statistically significant at conventional levels. The placebo test therefore does not provide any reason to doubt the

common trends assumption. Figure 17 in Appendix D plots the coefficient from the triple-difference turnover and wastage models to allow for visual inspection of how the treatment effects emerge over time.

4.5 Conclusion

Advanced economies regularly suffer from shortages of qualified science teachers. This is in part because scientists leave the teaching profession at a faster rate than their peers. Finding ways to improve the retention of these teachers is therefore important. However, there has been very little research evaluating specific interventions for achieving this. This chapter is among the first to investigate the link between subject-specific CPD, which combines training on subject knowledge with training on pedagogical techniques, and the retention of science teachers.

I replicate existing findings (Kelly, 2004; Worth & De Lazzari, 2017) that science teachers are both more likely to leave their current school and to leave the teaching profession as a whole than non-science teachers. Fortunately, this evaluation of the NSLN suggests that subject-specific CPD can help address this issue. My individual-level cross sectional models show that participation in NSLN programmes is associated with a 43% increase in the odds that a teacher remains at their school two years after first observed participation, rising to 65% among those who participate in more than two days of courses in the first year of participation. However, participation is not associated with increased retention in a teacher's original school. My department-level panel data models, which account for a range of unobserved variables, come to qualitatively similar conclusions: first observed participation in NSLN courses is associated with a drop in wastage of between two to three percentage points after two years. This result is materially significant: average science department wastage was 9% in 2011. The findings support the use of such courses in order to retain science teachers in the profession.

The findings of this research should of course be interpreted in light of its limitations. The placebo test in the year prior to first observed participation does not show any statistically significant 'treatment' effect, suggesting treated and untreated departments do not differ systematically in terms of changes in turnover or wastage immediately prior to treatment. This is reassuring. However, it would be preferable to be able to investigate the trends in outcomes over additional periods in the pre-intervention period. Unfortunately, the panel data used to measure the outcome variables begins only shortly before the treated schools used in this evaluation begin to participate. A second important limitation that should be borne in mind while interpreting the findings from this evaluation is that any treatment effects identified here are, strictly speaking, only relevant to those who have participated in the programme in the period for which I have data. There is no guarantee that the same effects would

result should the programme to be expanded beyond the relatively young and inexperienced group of teachers who have tended to participate up until now.

A third important limitation relates to the definition of my treatment variable. The NSLN data only covers 2010 onwards meaning that I am forced to define treatment as first *observed* participation. It is therefore possible that some of the schools which I use as "never treated" control schools have been treated at some point between 2005 (when the programme began) and 2009 (which is the last year for which I do not have data). How might this affect my results? The double and triple difference models employed in this chapter control for prior levels of turnover and wastage in both treatment and control schools. To the extent that control schools treated between 2005 and 2009 have already benefited from the programme by 2011 (which is the first year of treatment I evaluate) my models will therefor net out the effects of treatment prior to 2010. Nevertheless, I cannot rule out that some control schools are still benefiting from receiving the treatment three of more years after participation. If so, this will downward bias my estimates of the programme impact, making my estimates of the treatment effect conservative. While these limitations are important, the downward bias they would introduce are unlikely to change the overall conclusion of this paper that participation in NSLN courses reduced wastage from science departments.

The finding that participation increases retention in the profession but does not increase retention in participating schools is interesting. Wastage is a component of turnover since anyone who leaves the profession also leaves their school (the average turnover rate among science departments in 2011 was 14%). This implies that the reduced number of teachers leaving the profession from a participating department is being offset by an increased number of teachers leaving their department to move to another school. Further research is needed to understand why participants move schools more often than non-participants. In the meantime however, this finding has important implications for the way in which NSLN courses are funded, because individual schools are unlikely to capture the benefits of investing in their staff through sending them on the programme. Government and the charitable sector should therefore continue to fund the programme on the grounds that it benefits the school system overall, through reducing teacher shortages.

Future research should also investigate the mechanism through which NSLN participation affects retention by collecting longitudinal survey data on the mechanisms outlined in Wolstenholme et al. (2012) such as self-efficacy and career progression, neither of which are observable in the school workforce census.

Chapter 5: Conclusion

Teachers have a powerful influence on their pupils (Hanushek, 2011). Despite this, public school systems suffer from recurring shortages of qualified, experienced staff (Dolton, 2006). England is currently in the midst of one such shortage caused, in part, by declining early-career retention. This thesis has investigated why people enter and exit the teaching profession, adding to existing knowledge by using better data, studying the subject in new contexts, or answering questions which have not previously been addressed by academic research. The advent of very large survey datasets has enabled me to model the processes of entry and exit in greater detail than previous research, going beyond demographic characteristics to investigate the importance of personality in influencing entry to the profession. Similarly, I have been able to quantify the nature and importance of teachers' working environments for their job satisfaction and intention to leave. The rich descriptive analysis of survey datasets in Chapter 2 and 3 is complemented by the causal analysis in Chapter 4, which exploits panel data on teachers in England. This conclusion summarises the findings and contribution of each of the three papers, and considers future directions for research on the recruitment and retention of teachers.

Chapter 2 reports on the first analysis of entry to the teaching profession to use household panel survey data. This data allows me to describe the changing characteristics of those entering the teaching profession from full time education, stretching back 75 years. The analysis reveals that the well-established trend toward a more male workforce has halted among the most recent cohort of teachers. The data also puts teachers' attitudes to their jobs in comparative perspective, showing that they have higher job satisfaction, but lower satisfaction with their leisure time than graduates in general. Teachers are also much more likely to have personalities characterised by high levels of openness to new experience than graduates in general. When modelling entry to the profession, I also find strong, robust associations between openness and choosing to become a teacher. I show that just four characteristics of teachers can be used to identify groups with a predicted probability of entry up to four times higher than the average graduate, information that could be used to improve the targeting of efforts to recruit teachers.

Further research could empirically test the predictions made by these models. For example, as the Millennium Cohort Study (MCS) participants move into the Labour market over the next five years, the information contained within that dataset could be used to assess the validity of the predictions made by the models in this paper. Alternatively, if it were possible to collect information on the four variables referred to above in the New Destination of Leavers of Higher Education (NDLHE), then it would be possible to test whether groups with high predicted propensities to enter teaching are more likely to respond to postal information about a career in teaching. Survey experiments (Mutz, 2011) could also be used to test whether those with a high propensity to enter teaching were more likely to

subsequently begin teacher training, following the receipt of marketing materials. In any case, future research on entry to the teaching profession should include personality measures, particularly openness to new experience and neuroticism.

Chapter 3 reports on one of the first analysis of the relationship between working conditions and retention outside of America. Indeed, the international survey data used allow me to model this relationship across 35 different countries. The TALIS data provide unusually rich and comprehensive measures of teacher working conditions. My analysis therefore includes a number of working conditions such as scope for progression, feedback and preparation, which have not previously been used to model teacher job satisfaction and turnover intentions. By using oblique rotation methods in my factors analysis, which is unusual in this literature but methodologically more appropriate, I am also able to document how working conditions are correlated within and across schools. I replicate the finding from the literature on US schools that school leadership has the strongest relationship with retention. I also extend existing knowledge by showing that preparation and scope for progression are both related to teachers' desire to move school.

An important limitation of this chapter is the lack of data on observed turnover behaviour. Fortunately the next round of TALIS, due to be collected in 2018, will be linked to the School Workforce Census in England, which will allow actual turnover and attrition to be modelled using the same working conditions measures. Currently Kraft et al. (2016) is the only paper to utilise a school-level panel of working conditions, which enables them to use fixed effects and a number of other methods able to get closer to a causal analysis. The arrival of the TALIS 2018 data will also allow comparisons to be made within countries across time. A final direction for research in this area is to integrate the working conditions literature, which focuses on aspects of teachers working lives, and has largely been developed by economists, with the school climate literature (Thapa et al., 2013), which focuses on a broader set of measures related to school life such as norms and values, and has largely been developed by psychologists.

The research reported in Chapter 4 is some of the first to evaluate the effects of subject-specific professional development on teacher retention. My double- and triple-difference models, combined with panel data on all teachers in England, provide evidence on the causal impact of participation under the standard common-trends assumption. Using both science departments in non-participating schools and non-science departments in participating schools simultaneously as control structures enables me to rule out a number of the most plausible threats to the internal validity of the evaluation. The results suggest that participation in the NSLN programmes does indeed improve retention of teachers in the profession, although not necessarily in their original school. The research therefore suggests that the NSLN is an effective way of reducing the current shortage of science teachers. The

findings suggest that the benefits of increased retention are not captured by participants' original schools, which suggests an ongoing role for government in funding this training.

This evaluation does have a number of important limitations. Foremost among these is that it relies on the common trends assumption. Given that quasi-experimental evidence suggests that it helps retain teachers, it may be worth conducting a randomised controlled trial of the NSLN intervention to see if the findings can be reproduced experimentally. An RCT might also allow individual-level analysis, which would provide more precise information about dose-response relationships. A related direction for future research would be to conduct a cost-benefit analysis in order to determine whether the enhanced retention from the programme represents good value for money. This would require estimating the cost per participant, per day and the benefits, evaluated in terms of the reduction in training and recruitment costs achieved through enhanced retention.

Thinking more generally about research on the supply of teachers, there is a need for more research on the effectiveness of incentives for both recruitment and retention. In the US, a number of recent studies have found that bonus payments can be effective ways of increasing the supply of teachers, at least in the short run (Bueno & Sass, 2016; Clotfelter et al., 2008; Feng & Sass, 2016). In particular, there is potential to evaluate the effects of the bursary and scholarship training incentives currently available for shortage subject teachers. The subject-eligibility and value of these payments have varied across years, which provides useful, plausibly-exogenous variation to identify both the impact of the incentives on recruitment and the dose-response relationship. In addition, the recent announcement that student loan forgiveness will be offered to MFL and science teachers and the early-career bonus payments to maths teachers provide an opportunity to investigate the effects of incentives on retention.

Appendix A: Supplementary results from Chapter 1

Table 14: Net retention by ITT cohort with wastage rates frozen at 2009 levels

Cohort	NQTS	Net retention by cohort, true historic rates of wastage (Years Since NQT)							Net retention by cohort, wastage frozen at 2009 levels (Years Since NQT)						
		+1	+2	+3	+4	+5	+6	+7	+1	+2	+3	+4	+5	+6	+7
2009	22,300	87%	83%	79%	76%	72%	68%	67%	87%	83%	79%	76%	72%	68%	67%
2010	24,100	87%	82%	77%	73%	70%	66%		87%	83%	79%	76%	72%	68%	
2011	20,600	88%	83%	77%	73%	69%			87%	83%	79%	76%	72%		
2012	23,000	88%	81%	75%	71%				87%	83%	79%	76%			
2013	23,600	87%	80%	74%					87%	83%	79%				
2014	24,200	87%	79%						87%	83%					
2015	25,500	87%							87%						
2009	22,300	19,401	18,509	17,617	16,948	16,056	15,164	14,941	19,401	18,509	17,617	16,948	16,056	15,164	14,941
2010	24,100	20,967	19,762	18,557	17,593	16,870	15,906		20,967	20,003	19,039	18,316	17,352	16,388	
2011	20,600	18,128	17,098	15,862	15,038	14,214			17,922	17,098	16,274	15,656	14,832		
2012	23,000	20,240	18,630	17,250	16,330				20,010	19,090	18,170	17,480			
2013	23,600	20,532	18,880	17,464					20,532	19,588	18,644				
2014	24,200	21,054	19,118						21,054	20,086					
2015	25,500	22,185							22,185						

Notes:

The top two panels show the net retention rates for each cohort of NQTs in each year. The top left panel shows the historically accurate retention rates, as observed in each cohort/year. The top right panel shows the historically accurate retention rates for the 2009 cohort and the same 2009 retention rates applied to each subsequent cohort, as if retention rates had been frozen at 2009 level.

The bottom two panels show the number of NQTs still working in state funded schools in England for each cohort in each year. The numbers in the bottom left panel are the historically accurate retention figures implied by the wastage rates in the top left panel. The numbers in the bottom right panel are fictional retention figures calculated using the fictional (frozen at 2009) retention rates shown in the top right panel.

The total retained NQTs since 2009 can be found by summing the diagonals (highlighted in grey) in the two lower panels. The total retained NQTs in the bottom left panel (using the historically accurate retention rates) is 105,217. The total retained NQTs in bottom right panel (using the fictional 2009 cohort retention rates) is 109,615. The additional net wastage that has occurred since 2009 due to increasing early-career wastage is the difference between these two figures: 4,398. To put this in context, the total shortfall of EBACC teachers in 2017/18 is 2,080. Declining early-career retention is therefore an important reason for the current teacher shortage.

Source: ITT Number Census Main Tables 2017, SFR 68.

Appendix B: Supplementary results from Chapter 2

Table 15: Counts of teachers in the Understanding Society data

	In the Teaching Profession				Entering the Teaching Profession			
	Ever	1st Job	W1-W6	1st Job & W6	1935-54	1955-74	1975-94	95-2016
Any Type of Teacher	2,485	1,123	1,761	234	56	439	358	263

Notes: Table 15 shows unweighted frequencies of teachers who can be identified as either being in or entering the profession, conditional on interval censoring of individual employment histories. W1 is Wave 1 of Understanding Society data collection (2009-10), W6 is Wave 6 (2014-15). The left panel and right panel do not match up because of censoring.

Table 16: Dichotomised measures of job/income/leisure-time/life satisfaction by professional group

Satisfaction with…	Teachers	Nurses	Govt/ Admins	Protective Officers	All Other Graduates
Job (Dummy)	71.5%	63.9%**	56.2%**	69.4%	73.3%
Income (Dummy)	59.1%	52.8%*	48.7%**	56.4%	57.0%
Leisure Time (Dummy)	43.6%	45.4%	43.3%	46.6%	51.0%**
Overall (Dummy)	73.1%	71.9%	62.4%**	71.5%	70.2%

Notes: All satisfaction variables have been dichotomised, with all responses at or above "Mostly or completely satisfied given a value of 1 and all others given a value of 0. ** = $p < 0.01$, * = $p < 0.05$.

Table 16B: Pairwise correlations between big five personality variables

	F1	F2	F3	F4	F5
F1: Agreeableness	1.00				
F2: Conscientiousness	**0.33**	1.00			
F3: Extraversion	0.16	0.20	1.00		
F4: Neuroticism	-0.06	-0.18	-0.18	1.00	
F5: Openness	0.19	0.21	0.25	-0.10	1.00

Notes: The table shows the pairwise correlations between the five dimensions of personality. Only one of the ten correlations (highlighted in bold) has an absolute value greater than 0.3. These low correlations suggest that the five dimensions capture distinct underlying latent variables, which is necessary for a ceteris paribus interpretation of the coefficient in my models.

Table 17: Population-level conditional associations with teaching, by birth cohort

	(1)	(2)	(3)	(4)
	Pre 1945	Baby Boomers	Gen X	Millennials
Ethnic Minority	0.836	0.730	0.609*	0.496**
	(0.190)	(0.0904)	(0.107)	(0.0746)
Female	5.865**	2.637**	2.006**	2.792**
	(0.979)	(0.226)	(0.261)	(0.362)
Graduate	13.79**	11.19**	17.38**	10.58**
	(2.225)	(1.077)	(3.803)	(1.527)
Mother/Father Teach	2.562	1.810**	1.340	1.341
	(1.044)	(0.305)	(0.253)	(0.238)
Father Graduate	1.641*	1.075	1.238	1.141
	(0.314)	(0.115)	(0.190)	(0.176)
N	9,613	15,337	11,698	14,211

Notes: Table 17 probes the bivariate associations in Figure 6 by looking at associations conditional on other characteristics, across each birth-cohort. Each column is a separate logistic regression on a separate group of teachers. The outcome being modelled is whether somebody can be identified as having taught at some point in their life. The coefficients are odds ratios. The conditional odds of a female from the Pre 1945 birth cohort are 487% higher than males from the same generation, but this reduces in the next two birth-cohorts, before increasing slightly among millennials. Figure 6 showed a steady increase in the proportion of teachers who have a mother or father who taught. By contrast, Table 17 shows that those with a parent teacher have higher odds of becoming teachers in the Pre 1945 and Baby Boomer generations, but the association becomes smaller and no longer statistically significant at conventional levels among Gen Y and Millennials. Similarly, the conditional association between having a graduate father and choosing teaching that is visible among the Pre 1945 generation is no longer visible in later birth-cohorts. ** = p< 0.01, * = p< 0.05. Numbers in parentheses are standard errors. The N in the final row is the number of individuals in the data from each birth cohort.

Table 18: Descriptive statistics for the independent variables in Chapter 2

	Min	Max	Mean	SD	Missing (%)
Female	0	1	0.51		0
Ethnic minority	0	1	0.20		0
Graduate	0	1	0.21		0
First gen immigrant	0	1	0.12		0
Second plus gen immigrant	0	1	0.77		0
Father is a graduate	0	1	0.16		0
Mother/father taught when 14	0	1	0.04		0
Personality:					
Agreeableness	1	7	5.63	0.01	64.8
Conscientiousness	1	7	5.50	0.01	64.8
Extraversion	1	7	4.59	0.01	64.8
Neuroticism	1	7	3.54	0.01	64.8
Openness	1	7	4.55	0.01	64.8
No. of close friends	0	100	5.25	0.04	60.6
Self-efficacy	1	4	3.12	0.01	68

Notes: Weights applied for the wave in which the data was collected. Missing values are due overwhelmingly to non-administration rather than non-response (see Section 2.5 for a discussion).

Table 19: Logistic regression models of current teaching status with imputed missing variables and dummies indicating missingness

	(1)	(2)	(3)	(4)
Female	2.251**	2.210**	2.234**	2.267**
	(0.189)	(0.195)	(0.198)	(0.200)
Ethnic minority	0.890	0.873	0.853	0.854
	(0.138)	(0.137)	(0.135)	(0.137)
Graduate	10.89**	10.47**	10.43**	10.13**
	(1.114)	(1.119)	(1.120)	(1.092)
First gen immigrant	0.399**	0.388**	0.403**	0.395**
	(0.0981)	(0.0970)	(0.101)	(0.0989)
Second plus gen immigrant	0.918	0.890	0.905	0.888
	(0.161)	(0.159)	(0.163)	(0.160)
Father is a graduate	0.985	0.968	0.967	0.964
	(0.0941)	(0.0931)	(0.0930)	(0.0928)
Mother/father taught when 14	1.219	1.182	1.166	1.164
	(0.171)	(0.166)	(0.165)	(0.164)
Personality (z scored):				
Agreeableness		1.174*	1.172*	1.174*
		(0.0587)	(0.0587)	(0.0587)
Conscientiousness		0.958	0.956	0.934
		(0.0482)	(0.0482)	(0.0475)
Extraversion		1.019	1.011	0.999
		(0.0476)	(0.0479)	(0.0476)
Neuroticism		1.064	1.065	1.116*
		(0.0521)	(0.0520)	(0.0569)
Openness		1.276**	1.272**	1.247**
		(0.0694)	(0.0695)	(0.0690)
Missing dummy for Personality		1.047	0.736	0.748
		(0.141)	(0.143)	(0.147)
No. of close friends (z scored)			1.059	1.057
			(0.0398)	(0.0400)
Missing dummy for friend			1.735	1.738
			(0.396)	(0.400)
Self-efficacy (z scored)				1.169*
				(0.0583)
Missing dummy for self-efficacy				0.934
				(0.136)
Number of Observations	18,022	18,022	18,022	18,022

Notes: Each column is a separate logistic regression on a separate group of teachers. Coefficients are odd ratios. ** = p< 0.01, * = p< 0.05. Numbers in parentheses are standard errors. Models also contains country dummies for the four nations of the UK.

Table 20: Table 4 but with the sample held constant across columns

	(1) Demographics	(2) Demographics & Personality	(3) Demographics & Personality & Networks	(4) Demographics & Personality & Networks & Efficacy
Female	2.211**	2.179**	2.180**	2.208**
	(0.206)	(0.214)	(0.214)	(0.216)
Ethnic minority	0.859	0.843	0.846	0.843
	(0.160)	(0.157)	(0.157)	(0.158)
Graduate	10.40**	9.904**	9.817**	9.606**
	(1.196)	(1.185)	(1.179)	(1.159)
First gen immigrant	0.605	0.584	0.585	0.580
	(0.178)	(0.175)	(0.175)	(0.173)
Second plus gen immigrant	1.135	1.105	1.103	1.089
	(0.250)	(0.246)	(0.246)	(0.243)
Father is a graduate	1.001	0.985	0.983	0.980
	(0.103)	(0.102)	(0.102)	(0.102)
Mother/father taught when 14	1.293	1.259	1.246	1.243
	(0.193)	(0.187)	(0.186)	(0.185)
Personality (z scored):				
Agreeableness		1.157*	1.156*	1.158*
		(0.0598)	(0.0596)	(0.0596)
Conscientiousness		1.000	0.999	0.980
		(0.0525)	(0.0524)	(0.0523)
Extraversion		0.995	0.987	0.977
		(0.0476)	(0.0479)	(0.0474)
Neuroticism		1.047	1.049	1.091
		(0.0530)	(0.0530)	(0.0588)
Openness		1.285**	1.280**	1.260**
		(0.0714)	(0.0714)	(0.0714)
No. of close friends (z scored)			1.066	1.064
			(0.0413)	(0.0414)
Self-efficacy (z scored)				1.126
				(0.0608)
Number of Observations	13,649	13,649	13,649	13,649

Notes: Each column is a separate logistic regression on a separate group of teachers. Coefficients are odd ratios. ** = $p < 0.01$, * = $p < 0.05$. Numbers in parentheses are standard errors. Models also contains country dummies for the four nations of the UK. Number of observation drops due to personality, social network and self-efficacy questions only be asked in certain waves. Pseudo R Squared is McFadden's Pseudo R Squared.

Table 21: Associations between teaching and different sets of variables

	(1)	(2)	(3)
Personality (z scored):			
Agreeableness	1.171*		
	(0.0588)		
Conscientiousness	0.958		
	(0.0481)		
Extraversion	1.018		
	(0.0477)		
Neuroticism	1.063		
	(0.0521)		
Openness	1.277**		
	(0.0699)		
No. of close friends (z scored)		1.086	
		(0.0387)	
Self-efficacy (z scored)			1.174**
			(0.0538)
Number of Observations	13,714	14,834	13,900

Notes: Each column is a separate logistic regression. ** = $p< 0.01$, * = $p< 0.05$. Numbers in parentheses are standard errors. Models also contain demographic and family background variables, as well as country dummies for the four nations of the UK. Pseudo R Squared is McFadden's Pseudo R Squared.

Appendix C: Supplementary results from Chapter 3

Table 22: Components of the teacher job satisfaction score

TALIS Variable	Question Wording	Mean Score	Standard Deviation
TT2G46A	The advantages of being a teacher outweigh the disadvantages	2.9	0.8
TT2G46B	If I could decide again, I would still choose to work as a teacher	3.0	0.8
TT2G46C	I would like to change to another school if that were possible	3.0	0.8
TT2G46D	I regret that I decided to become a teacher	3.3	0.7
TT2G46E	I enjoy working at this school	3.2	0.7
TT2G46F	I wonder whether it would have been better to choose another profession	3.0	0.8
TT2G46G	I would recommend my school as a good place to work	3.1	0.7
TT2G46J	All in all I am satisfied with my job	3.2	0.6

Notes: The table lists the mean and standard deviation (across the TALIS sample) of the variables which make up the teacher job satisfaction score. C, D and F are reverse scored, making a higher score "better" in all cases.

Table 23: Missing Data in the England Sample

Variable	Percent Missing
TT2G12A	0.72
TT2G12B	0.72
TT2G13A	0.8
TT2G13B	1
TT2G16	1.96
TT2G18A	2.96
TT2G18C	3.45
TT2G20A	2.56
TT2G20B	3.29
TT2G25A	13.5
TT2G25B	13.94
TT2G25C	13.86
TT2G25D	14.5
TT2G28B6	7.41
TT2G28C6	7.29
TT2G28D6	6.29
TT2G28E6	7.49
TT2G28F6	6.97
TT2G31A	7.05
TT2G31D	7.01
TT2G31E	6.89
TT2G31H	7.93
TT2G33A	5.93
TT2G33B	6.09
TT2G33C	6.05
TT2G33D	6.09
TT2G33E	6.17
TT2G33F	6.29
TT2G33G	6.25
TT2G33H	6.45
TT2G34D	6.17
TT2G34H	6.17
TT2G44A	6.57
TT2G44D	6.65
TT2G44E	6.89
TT2G27C	4.89
TT2G47C_EN~2	7.01
TT2G47D_EN~2	7.17
TT2G47E_EN~2	7.25
TT2G47L_EN~2	7.01
TT2G47M_EN~2	7.33
TT2G47N_EN~2	7.17

Notes: Levels of missing data on the working conditions variables in the England sample. All variables are below 15% which makes multiple imputation appropriate (Cheema, 2014). Results from the analysis with multiple imputation can be found in Table 32.

Table 24: Pairwise correlations between the complex scale scores used in the international modelling

	STEFFPROS	STCOOPS	TCSTAKES
Profession Development (STEFFPROS)	1.00		
Teacher Cooperation (STCOOPS)	0.27	1.00	
Stakeholder Engagement (TCSTAKES)	0.002	-0.009	1.00

Notes: The table shows the pairwise correlations between the three complex scale scores measuring working conditions in the international analysis. Zero of the three correlations have an absolute value greater than 0.3. This suggests that the three complex scale scores are measuring distinct latent variables, which is necessary for a ceteris paribus interpretation of the coefficients in Table 7.

Table 25: England-Specific variables included in the england-only analysis

TALIS Variable Code	TALIS Item Wording
TT2G47C_ENGX	This school has an effective school management team.
TT2G47D_ENGX	The school management team give clear vision and direction.
TT2G47E_ENGX	I do not have the autonomy I need to do a good job as a teacher.
TT2G47J_ENGX	Parents are supportive of my role as their children's teacher.
TT2G47L_ENGX	I have scope to progress as a classroom teacher.
TT2G47M_ENGX	I have scope to progress into a leadership team role.
TT2G47N_ENGX	I have scope to progress to a higher pay level.

Notes: Table 25 lists the eight working conditions variables that were asked exclusively to the England samples in TALIS 2013.

Table 26: Variation in the variables used in the exploratory factor analysis

Variable	Mean	S.D.	Min	25th Percentile	Median	75th Percentile	Max
TT2G44A	2.53	0.79	1	2	3	3	4
TT2G34D	3.42	0.69	1	3	4	4	4
TT2G34H	3.51	0.62	1	3	4	4	4
TT2G20A	0.8	0.4	0	1	1	1	1
TT2G20B	1.67	0.47	1	1	2	2	2
TT2G31H	2.84	0.75	1	2	3	3	4
TT2G12A	1.77	0.42	1	2	2	2	2
TT2G12B	1.74	0.52	0	2	2	2	2
TT2G28B6	0.87	0.99	0	0	0	2	2
TT2G28C6	0.79	0.98	0	0	0	2	2
TT2G28D6	1.47	0.89	0	0	2	2	2
TT2G28E6	0.97	1	0	0	0	2	2
TT2G28F6	0.81	0.98	0	0	0	2	2
TT2G31A	2.31	0.85	1	2	2	3	4
TT2G31D	2.7	0.73	1	2	3	3	4
TT2G31E	2.6	0.74	1	2	3	3	4
TT2G16	51.11	11.01	30	44	50	60	91
TT2G18A	8.46	5.44	0	5	7	10	50
TT2G18C	6.67	5.13	0	3	5	10	90
TT2G42F	3.06	0.71	1	3	3	3	4
TT2G43D	2.95	0.68	1	3	3	3	4
TT2G27C	2.07	0.86	1	1	2	3	4
TT2G44D	2.66	0.74	1	2	3	3	4
TT2G33A	2.53	1.7	1	1	2	3	6
TT2G33B	2.9	1.34	1	2	3	4	6
TT2G33C	2.32	1.31	1	1	2	3	6
TT2G33D	4.98	1.22	1	4	5	6	6
TT2G33E	5.2	1.16	1	5	6	6	6
TT2G33F	4.2	1.43	1	3	4	5	6
TT2G33G	3.65	1.84	1	2	4	5	6
TT2G33H	3.76	1.41	1	3	4	5	6
TT2G25A	2.52	0.9	1	2	2	3	4
TT2G25B	2.33	0.79	1	2	2	3	4
TT2G25C	2.16	0.83	1	2	2	3	4
TT2G25D	1.87	0.87	1	1	2	2	4
TT2G44E	2.68	0.78	1	2	3	3	4
TT2G47N_ENGX2	2.67	0.78	1	2	3	3	4
TT2G47M_ENGX2	2.66	0.77	1	2	3	3	4
TT2G47L_ENGX2	2.85	0.67	1	3	3	3	4
TT2G47C_ENGX2	2.76	0.78	1	2	3	3	4
TT2G47D_ENGX2	2.73	0.81	1	2	3	3	4
TT2G47E_ENGX2	2.78	0.76	1	2	3	3	4

Notes: The table shows the variation in the set of variables used in the exploratory factor analysis.

Table 27: Pairwise correlations between the eight working conditions latent variables

	F1	F2	F3	F4	F5	F6	F7	F8
F1: Leadership	1.00							
F2: Collaboration	**0.34**	1.00						
F3: Feedback	**0.54**	**0.42**	1.00					
F4: Scope for Progression	**0.41**	0.44	0.28	1.00				
F5: Prof Development	**0.40**	0.25	0.29	0.19	1.00			
F6: Preparation	0.19	0.24	0.25	0.16	0.16	1.00		
F7: Discipline	0.00	0.05	0.01	0.14	-0.20	0.13	1.00	
F8: Workload	-0.22	**0.35**	0.06	0.01	-0.12	0.12	-0.11	1.00

Notes: The table shows the pairwise correlations between the eight factors initially extracted from the exploratory factor analysis. Six of the pairwise correlations (highlighted in bold) have an absolute value greater than 0.3. These correlations suggest that an orthogonal factor rotation would have been inappropriate. This provides justification for my decision to use an oblique (Promax) factor rotation.

Table 28: Factors loadings

Variable	F1	F2	F3	F4	F5	F6	F7	F8	Uniqueness
12A						0.77			0.43
12B						0.78			0.37
13A						0.70			0.40
13B						0.63			0.44
16								0.60	0.63
18A								0.49	0.73
18C								0.55	0.72
20A									0.85
20B									0.79
25A					0.61				0.66
25B					0.76				0.41
25C					0.73				0.43
25D					0.63				0.60
28B6				0.71					0.56
28C6				0.72					0.53
28D6				0.67					0.56
28E6				0.68					0.52
28F6				0.85					0.37
31A									0.83
31D									0.62
31E	0.33								0.58
31H									0.70
33A		0.39							0.82
33B		0.45							0.64
33C		0.42							0.73
33D		0.66							0.63
33E		0.76							0.48
33F		0.80							0.46
33G		0.61							0.66
33H		0.67							0.48
34D							0.82		0.34
34H							0.80		0.34
44A	0.79								0.35
44D	0.89								0.27
44E	0.91								0.25
27C	0.40								0.64
47C	0.98								0.18
47D	0.97								0.19
47E	0.33								0.79
47L			0.75						0.38
47M			0.91						0.19
47N			0.90						0.28

Notes: The table shows the factor loadings following Promax rotation. Loadings less than 0.32 are not included in the table. All variables load only on a single factor. F1 = Leadership, F2 = Collaboration, F3 = Feedback, F4 = Scope for Progression, F5 = Prof Development, F6 = Preparation, F7 = Discipline, F8 = Workload. See Table 8 for question wording for each variable (not shown here for space reasons).

Table 29: Testing the interrelatedness of items

	Factor	No. of Items	Alpha	Spearman-Brown
1	Leadership	8	0.86	
2	Teacher Cooperation	8	0.76	
3	Scope for Progression	3	0.70	
4	Feedback	5	0.84	
5	Effective PD	4	0.74	
6	Preparation	4	0.69	
7	Discipline	2		0.78
8	Workload	3	0.5	

Notes: The table shows Cronbach's Alpha for the eight working conditions latent variables and the Spearman-Brown for the two working conditions variables composed of only two manifest variables. Values above 0.7 are generally considered acceptable. The Cronbach's Alpha for Workload is clearly below 0.7 and this factor is therefore not used in the analysis.

Table 30: Regressing working conditions on job satisfaction one-by-one

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Leadership (Z Score)	0.548 ** (0.028)							
Collaboration (Z Score)		0.237 ** (0.026)						
Scope Progression (Z Score)			0.464 ** (0.024)					
Feedback (Z Score)				0.21 ** (0.025)				
Prof Development (Z Score)					0.207 ** (0.026)			
Preparation (Z Score)						0.178 ** (0.028)		
Discipline (Z Score)							0.086 * (0.026)	
Workload Unmanageable (Likert)								-0.509 ** (0.024)

Notes: Demographic controls included in the model: female (dummy), age (years), full time (dummy), experience in teaching (years), pupils female (%), pupils ethnic minority (%), academy school (dummy). Prof Development = effective professional development. Column 1 is an ordinary least squares regression, column 2 is an ordered logistic regression. Numbers in parentheses are standard errors. ** = $p < 0.01$, * = $p < 0.05$. Workload is unmanageable is measured on a Likert scale from Strongly Disagree (1) to Strongly Agree (4).

Table 31: Regressing working conditions on desire to move school one-by-one

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Leadership (Z Score)	0.317 ** (0.021)							
Collaboration (Z Score)		0.647 ** (0.032)						
Scope Progression (Z Score)			0.415 ** (0.034)					
Feedback (Z Score)				0.684 ** (0.04)				
Prof Development (Z Score)					0.664 ** (0.047)			
Preparation (Z Score)						0.731 ** (0.037)		
Discipline (Z Score)							1.02 n.s. (0.082)	
Workload is Unmanageable (Likert)								1.87 ** (0.127)

Notes: Demographic controls included in the model: female (dummy), age (years), full time (dummy), experience in teaching (years), pupils female (%), pupils ethnic minority (%), academy school (dummy). Prof Development = effective professional development. Column 1 is an ordinary least squares regression, column 2 is an ordered logistic regression. Numbers in parentheses are standard errors. ** = $p < 0.01$, * = $p < 0.05$.

Table 32: Modelling job satisfaction and desire to move schools with multiple imputation

	(1) Job Satisfaction	(2) Desire to Move School
FSM (%)	-0.007	-0.017
	(0.004)	(0.025)
Leadership (Z score)	0.463**	0.512**
	(0.026)	(0.029)
Collaboration (Z score)	0.029	1.004
	(0.031)	(0.031)
Scope for Progression (Z score)	0.212**	0.879*
	(0.026)	(0.029)
Feedback (Z score)	-0.018	0.994
	(0.031)	(0.039)
Prof Development (Z score)	-0.047	1.018
	(0.026)	(0.04)
Preparation (Z score)	0.055*	0.936*
	(0.022)	(0.027)
Discipline (Z score)	0.113**	1.004
	(0.023)	(0.027)
Workload Unmanageable	-0.141**	1.005
(Likert)	(0.019)	(0.023)
(Pseudo) R Squared	0.40	0.14
N	2,028	2,028

Notes: Multiple imputation by chained equations was used to account for missing covariate data by creating five imputed datasets. Missing covariates were imputed using all other covariates included in the model. Prior to imputation, all variables had less than 5% of observations missing, except 25A-D which has between 10 and 11% of observations missing. After imputation all covariates had no missing data. Demographic controls included in the model: female (dummy), age (years), full time (dummy), experience in teaching (years), pupils female (%), pupils ethnic minority (%), academy school (dummy). FSM = Free School Meals. Prof Development = effective professional development. Column 1 is an ordinary least squares regression. The coefficients show the standard deviation change in job satisfaction associated with a one unit increase in the relevant independent variable, conditional on the other variables in the model. Column 2 is an ordered logistic regression. The coefficients show the odds ratio of being one category higher (stronger desire of wanting to move school) associated with a one unit increase in the relevant independent variable, conditional on the other variables in the model. Numbers in parentheses are standard errors. ** = $p < 0.01$, * = $p < 0.05$. Pseudo R Squared is McFadden's Pseudo R Squared.

Appendix D: Supplementary results from Chapter 4

Figure 13: Dosage histograms for NSLN participants
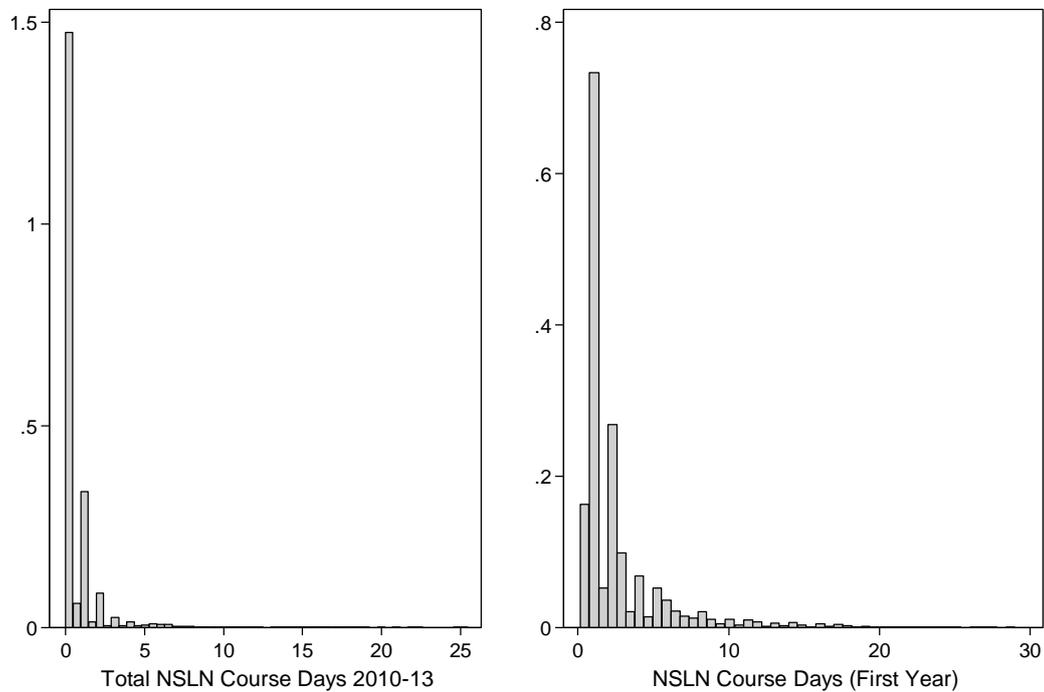
Notes: The figure shows the distribution of the number of days of training that all NSLN participants experienced. The left hand panel shows this for all years in the data. The right hand panel shows this for the first year in which a participant attended a course. Non-participants are not included. N=12,521 teachers. Teachers with more than 30 days of participation are also not included in the graph, in order to increase visual clarity. This amounts to 65 participants in the left hand panel and 20 participants in the right hand panel.

Figure 14: Common support graph for the 2011 first-participant matching
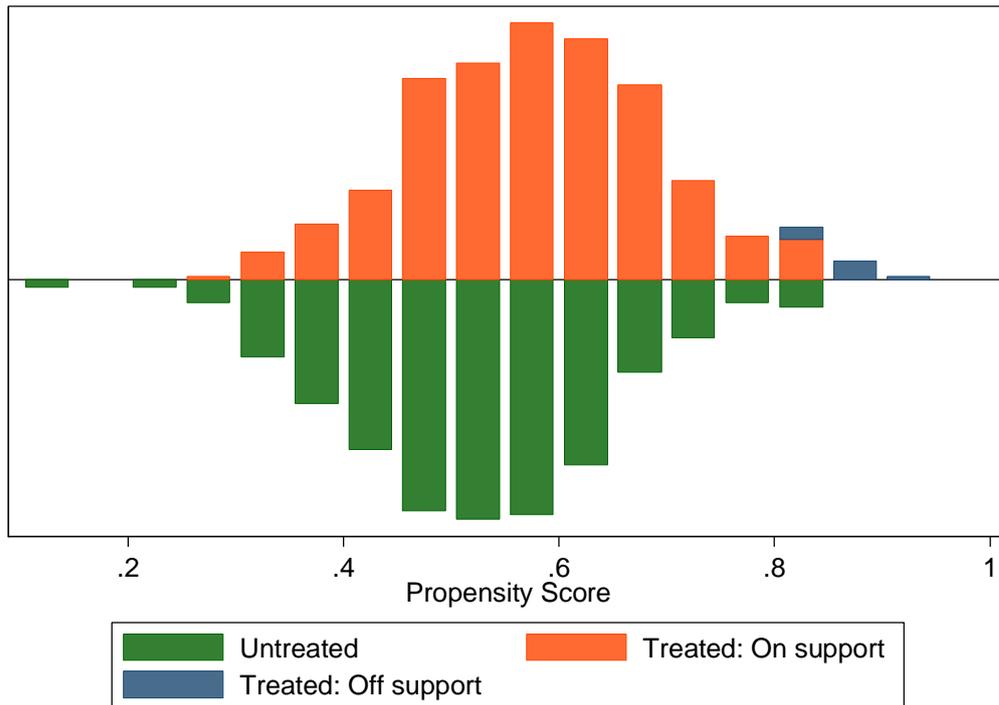
Notes: Shows the propensity scores for those departments that first participated in 2011 (orange & blue) and never-treated departments (green). 371 observations could not be matched within the caliper and were therefore dropped from the analysis. Propensity scores are estimated using the following variables: % teachers ethnic minority, % teachers female, mean age teacher, mean experience teachers, % teachers permanent contract, mean teacher pay, mean full-time equivalent, % female pupils in school, % pupils free school meals in school, average pupil grade on best 8 GCSE qualifications, mean KS2 prior attainment, school Ofsted grade, school sixth form indicator, school type indicator, school region indicator, urban/rural indicator.

Figure 15: Common support graph for the 2012 first-participant matching
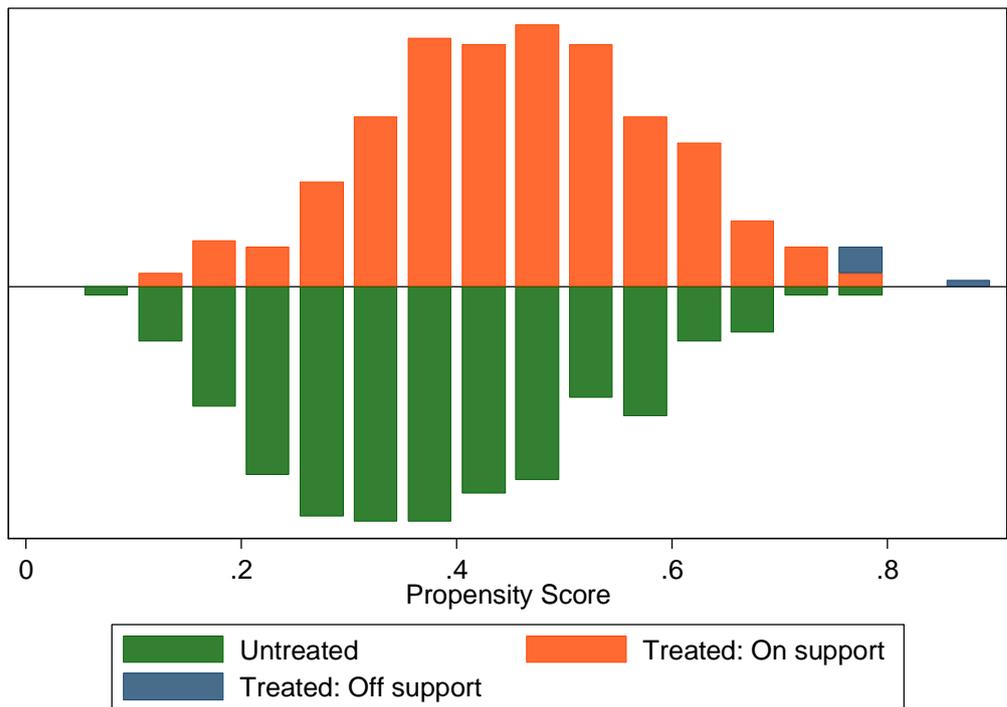
**Notes:** Shows the propensity scores for those departments that first participated in 2012 (orange & blue) and never-treated departments (green). 406 observations could not be matched within the caliper and were therefore dropped from the analysis. Propensity scores are estimated using the following variables: % teachers ethnic minority, % teachers female, mean age teacher, mean experience teachers, % teachers permanent contract, mean teacher pay, mean full-time equivalent, % female pupils in school, % pupils free school meals in school, average pupil grade on best 8 GCSE qualifications, mean KS2 prior attainment, school Ofsted grade, school sixth form indicator, school type indicator, school region indicator, urban/rural indicator.

Figure 16: Illustrating the rationale for a triple-difference specification
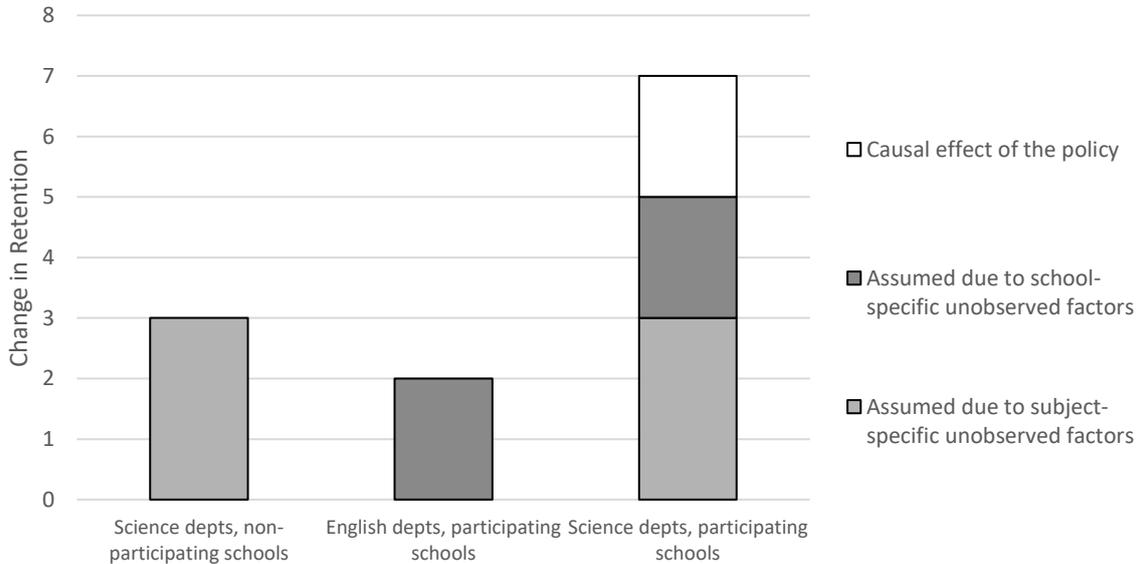
Notes: Figure 16 shows the change in retention in three school departments. The first bar shows that there has been a three percentage point increase in retention in sciences departments in schools that do not participate in NSLN courses. The second bar shows that there has been a 2 percentage point increase in retention in English departments in schools that do participate in NSLN courses. The third bar shows that there has been a 7 percentage point increase in retention in science departments in schools that do participate in NSLN courses.

A naïve estimate of the impact of NSLN attendance might conclude that it improves retention by seven percentage points. This would be inaccurate in cases where unobservable time-varying factors are also influencing retention, even when conditioning on observed variables.

However, if we assume that all time-varying factors are common between participating and non-participating departments, then we can net out the effect of these time-varying common factors by conditioning on the change in retention in other departments.

The double-difference specification nets out the change in observed retention in science departments in non-participating schools by conditioning on the change in retention in non-participating science departments $P_d = 0$ (the light grey area in the first and third bars of Figure 16.)

The triple-difference strategy goes one step further by also netting out the change in observed retention in English departments in participating schools. It does this by conditioning on <u>both</u> the change in retention in non-participating science departments (the light grey area in the first and third bar) and the change in retention in English departments in participating schools (the dark grey area in the second and third bar in Figure 16). These two types of non-participating departments therefore act as "control structures", which capture the effect of two different types of unobserved time-varying factors.

Under the common trends assumptions that all time varying unobserved factors have a common influence on treated and untreated departments, any remaining variation therefore represents the effect of NSLN participation. This is captured by the white area in the third bar of Figure 16.

Figure 17: Coefficient plots for turnover and wastage from the triple-difference models
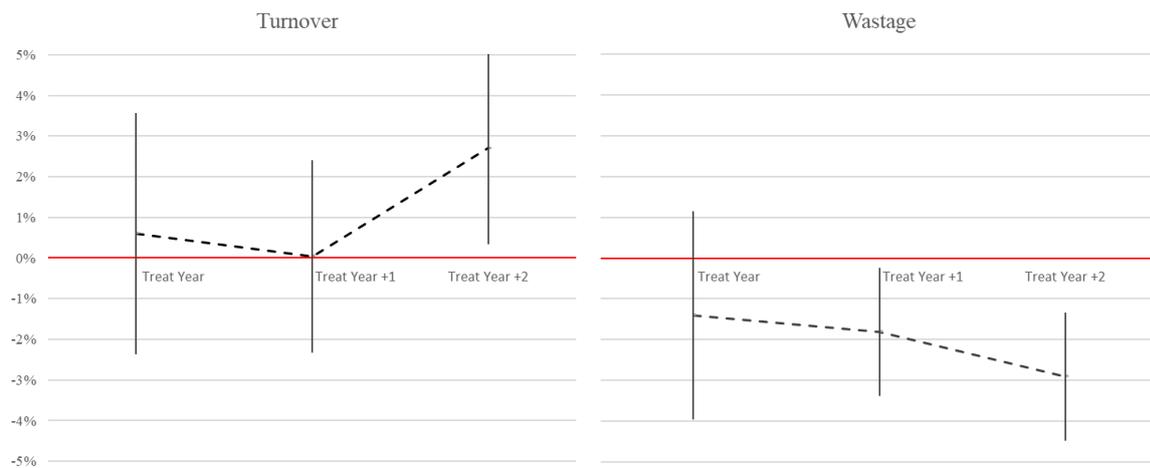
Notes: Figure 17 plots the point estimates and 95% confidence interval estimates (error bars) for the treatment effects in the year of first-participation (placebo), the year after first observed participation and the second year after participation. The left-hand side graph shows the triple-difference estimates for the turnover dependent variable taken from column 3 in Table 12. The right-hand side shows the triple-difference estimates for wastage dependent variable taken from column 6 in Table 12.

Table 33: Participant characteristics

	Proportion (%)	25th Percentile	50th Percentile	75th Percentile
NLSN Attendance (days)		1	1	3
Age (years)		28	35	44
Experience (years)		2.3	7.2	14.2
Tenure (years)		1.2	3.2	7.3
Annual Pay (£, 1000)		27	35.2	41.3
Free School Meals (%)		6.3	11.4	22.4
Best 8 GCSE Point Score		323	341	359
Female	57.7			
Qualified Status	96.4			
Permanent Contract	91.4			
Ofsted Outstanding	23.6			
Good	42.9			
Requires Improvement	30.2			
Inadequate	3.3			

Notes: N = 11,642. Data are for all secondary participants, with values taken from the year that a teacher first participates. Ofsted is the National Schools Inspectorate for England.

Table 34: Logistic regression of NSLN participation on retention in the teaching profession

	One Year Later	Two Years Later	Two Years Later
NSLN Participant (0/1)	1.869**	1.426**	
	(0.216)	(0.118)	
Heavy (>2 days) NSLN participant (0/1)			1.645*
			(0.318)
Subject Taught (ref = other subjects)			
Majority Science Teacher	0.777**	0.793**	0.774**
	(0.0362)	(0.0290)	(0.0285)
Majority Maths Teacher	1.040	1.079	1.077
	(0.0429)	(0.0345)	(0.0346)
Has Science Degree (0/1)	0.961	0.980	0.986
	(0.0351)	(0.0276)	(0.0282)
Female (0/1)	1.047	1.034	1.032
	(0.0280)	(0.0214)	(0.0218)
Ethnic Minority (0/1)	0.777**	0.769**	0.764**
	(0.0281)	(0.0217)	(0.0220)
Age (years)	0.970**	0.965**	0.965**
	(0.00192)	(0.00150)	(0.00153)
Experience (years)	0.977**	0.971**	0.971**
	(0.00222)	(0.00172)	(0.00174)
Permanent Contract (0/1)	3.122**	2.179**	2.210**
	(0.127)	(0.0780)	(0.0806)
Full Time Equivalent (%)	1.252	1.186	1.175
	(0.259)	(0.201)	(0.201)
Pay (£1000)	1.004**	1.004**	1.004**
	(0.000192)	(0.000150)	(0.000151)
Pupils Receiving Free School Meals (%)	0.480**	0.331**	0.350**
	(0.0827)	(0.0429)	(0.0463)
Pupils ethnic minority (%)	1.131	1.182**	1.160*
	(0.0906)	(0.0719)	(0.0717)
Average best 8 grades at Key Stage 4	1.002	1.001	1.001
	(0.000802)	(0.000614)	(0.000628)
Prior (Key Stage 2) attainment	0.984	0.961	0.959
	(0.0621)	(0.0467)	(0.0475)
Ofsted Rating (ref = 'Outstanding')			
Ofsted 'Good'	1.036	0.996	1.000
	(0.0381)	(0.0282)	(0.0289)
Ofsted 'Requires Improvement'	0.877*	0.844**	0.847**
	(0.0361)	(0.0266)	(0.0273)
Ofsted 'Inadequate'	0.700**	0.651**	0.647**
	(0.0547)	(0.0398)	(0.0403)
N	100,009	100,007	95,276
Pseudo R-Squared	0.052	0.054	0.054

Notes: Each column is a separate logistic regression. Coefficients are odds ratios. Also included in the regression are region dummies, urban/rural indicators and calendar year dummies. Covariates for non-participating (control) individuals all take 2010 values. Standard errors are in parentheses. ** = p< 0.01, * = p< 0.05. Ref = reference category for am ordinal variable. (0/1) = dummy variable. Pseudo R Squared is McFadden's Pseudo R Squared.

Table 35: Logistic regression of NSLN participation on retention in original school

	One Year Later	Two Years Later	Two Years Later
NSLN Participant (0/1)	1.174*	0.920	
	(0.0909)	(0.0529)	
Heavy (>2 days) NSLN participant (0/1)			1.057
			(0.140)
Subject Taught (ref = all other subjects)			
Majority Science Teacher	0.720**	0.685**	0.677**
	(0.0259)	(0.0193)	(0.0195)
Majority Maths Teacher	0.821**	0.810**	0.810**
	(0.0253)	(0.0193)	(0.0195)
Has Science Degree (0/1)	1.053	1.109**	1.114**
	(0.0294)	(0.0240)	(0.0246)
Female (0/1)	1.082**	1.099**	1.098**
	(0.0225)	(0.0176)	(0.0181)
Ethnic Minority (0/1)	0.970	0.975	0.972
	(0.0281)	(0.0218)	(0.0223)
Age (years)	0.995*	0.997	0.997
	(0.00161)	(0.00128)	(0.00132)
Experience (years)	0.989**	0.987**	0.986**
	(0.00202)	(0.00157)	(0.00160)
Permanent Contract (0/1)	4.078**	2.980**	3.039**
	(0.125)	(0.0841)	(0.0879)
Full Time Equivalent (%)	0.585	0.650*	0.644*
	(0.128)	(0.106)	(0.106)
Pay (£1000)	1.002**	1.002**	1.002**
	(0.000145)	(0.000111)	(0.000112)
Pupils Receiving Free School Meals (%)	0.685*	0.460**	0.477**
	(0.0937)	(0.0479)	(0.0508)
Pupils ethnic minority (%)	1.079	1.113*	1.080
	(0.0667)	(0.0522)	(0.0519)
Average best 8 grades at Key Stage 4	1.003**	1.002**	1.002**
	(0.000620)	(0.000471)	(0.000485)
Prior (Key Stage 2) attainment	1.044	1.066	1.058
	(0.0513)	(0.0401)	(0.0407)
Ofsted Rating (ref = 'Outstanding')			
Ofsted 'Good'	1.021	0.990	0.992
	(0.0291)	(0.0216)	(0.0222)
Ofsted 'Requires Improvement'	0.874**	0.832**	0.828**
	(0.0277)	(0.0202)	(0.0207)
Ofsted 'inadequate'	0.661**	0.620**	0.613**
	(0.0402)	(0.0299)	(0.0303)
N	100,009	100,009	95,278
Pseudo R-Squared	0.049	0.039	0.039

Notes: Each column is a separate logistic regression. Coefficients are odds ratios. Also included in the regression are region dummies, urban/rural indicators and calendar year dummies. Covariates for non-participating (control) individuals all take 2010 values. Standard errors are in parentheses. ** = p< 0.01, * = p< 0.05. Reference category for Ofsted rating is Outstanding. Pseudo R Squared is McFadden's Pseudo R Squared.

Table 36: Full regression output from the double- and triple-difference models

	Turnover		Wastage	
	Double-Diff	Triple-Diff	Double-Diff	Triple-Diff
	(7)	(8)	(9)	(10)
Effect Year 2	-0.00412	0.000433	-0.00808	-0.0182
	(0.00959)	(0.0116)	(0.00628)	(0.00813)
Effect Year 3	-0.00593	0.0265*	-0.0211**	-0.0285**
	(0.00960)	(0.0116)	(0.00629)	(0.00814)
Science Degree	-0.0137	0.0301	-0.0423	0.00706
	(0.0308)	(0.0373)	(0.0194)	(0.0268)
Female	-0.0203	-0.123**	-0.0586*	-0.120**
	(0.0287)	(0.0346)	(0.0188)	(0.0243)
Ethnic Minority	0.0714	0.121*	0.00647	0.0322
	(0.0329)	(0.0397)	(0.0216)	(0.0279)
Age	-0.0171**	-0.00784*	-0.0106**	-0.00558*
	(0.00206)	(0.00248)	(0.00135)	(0.00175)
Experience	0.00572*	-0.00153	0.000968	-0.00312
	(0.00234)	(0.00282)	(0.00153)	(0.00198)
Tenure	-0.0130**	-0.0104**	-0.00595**	-0.00450*
	(0.00167)	(0.00202)	(0.00109)	(0.00142)
Permanent Contract	0.0880*	0.103*	0.0623**	0.0393
	(0.0272)	(0.0328)	(0.0178)	(0.0231)
Full Time Equivalent	0.0350	0.0404	0.0544	0.100*
	(0.0440)	(0.0531)	(0.0288)	(0.0373)
Pay	0.00000207	0.00000161	0.00000113	0.00000252**
	(0.000000856)	(0.00000103)	(0.000000561)	(0.000000726)
% Pupils Female	0.229	-0.718**	0.0908	-0.249
	(0.172)	(0.208)	(0.108)	(0.149)
% Pupils Free School Meals	0.205	0.00881	0.255**	0.184
	(0.120)	(0.146)	(0.0758)	(0.105)
% Pupils Eth Min	0.345	0.212	0.255*	0.318*
	(0.141)	(0.172)	(0.0890)	(0.123)
Best 8 KS4	-0.000243	-0.000408	0.0000623	-0.0000151
	(0.000145)	(0.000176)	(0.0000913)	(0.000126)
Prior (KS2) Attainment	0.0314	0.0651	0.0153	0.0490
	(0.0278)	(0.0337)	(0.0175)	(0.0242)
Ofsted Good	-0.0148	0.0234	-0.00653	0.00462
	(0.0141)	(0.0171)	(0.00886)	(0.0122)
Ofsted Req. Imp.	-0.00527	0.0285	0.00610	0.00719
	(0.0162)	(0.0197)	(0.0102)	(0.0141)
Ofsted Inadequate	0.0188	-0.0258	0.0331	0.0115
	(0.0212)	(0.0258)	(0.0134)	(0.0185)
N	4,579	4,579	4,579	4,579
R-Squared	0.694	0.741	0.589	0.639

Notes: Each column is a separate OLS regression. Also included in all regressions are region dummies, urban/rural indicators and calendar year dummies, as well as all of the time-varying covariates from the models in Table 1. Standard errors are in parentheses. ** = $p < 0.01$, * = $p < 0.05$. N = number of groups. Calendar year dummies are also included in the model.

Bibliography

Allen, R., Belfield, C., Greaves, E., Sharp, C., & Walker, M. (2016). *The longer-term costs and benefits of different initial teacher training routes*. Institute for Fiscal Studies Report 118.

Allen, R., Burgess, S., & Mayo, J. (2018). The teacher labour market, teacher turnover and disadvantaged schools: new evidence for England. *Education Economics*, *26*(1), 4-23.

Allen, R. and Sims, S., (2017). *Improving science teacher retention: do National STEM Learning Network professional development courses keep science teachers in the classroom?* Wellcome Trust.

Allen, R. & Sims, S. (2018a). Do low-income pupils get low-quality teachers? Forthcoming in *Oxford Review of Education*.

Allen, R. & Sims, S. (2018b). Identifying schools with high use and high loss of Newly Qualified Teachers. National Institute Economic Review, *243*, 27-36.

Angrist, J. D., & Pischke, J. S. (2014). *Mastering 'metrics: The path from cause to effect*. Princeton University Press.

Atteberry, A., Loeb, S., & Wyckoff, J. (2017). Teacher churning: Reassignment rates and implications for student achievement. *Educational Evaluation and Policy Analysis*, *39*(1), 3-30.

Avvisati, F. & Keslair, F. (2014). *REPEST: Stata module to run estimations with weighted replicate samples and plausible values*. Statistical Software Components S457918, Boston College Department of Economics.

Bacolod, M. P. (2007a). Do alternative opportunities matter? The role of female labor markets in the decline of teacher quality. *The Review of Economics and Statistics*, 89(4), 737-751.

Bacolod, M. (2007b). Who teaches and where they choose to teach: College graduates of the 1990s. *Educational Evaluation and Policy Analysis*, 29(3), 155-168.

Bakker, A. B., & Demerouti, E. (2007). The job demands-resources model: State of the art. *Journal of managerial psychology*, *22*(3), 309-328.

Becker, A., Ed (2014). *TALIS User Guide for the International Database*. OECD.

Belanger, J., Normandeau, S. and Larrakoetxea, E. eds. (2014). *TALIS 2013 Technical Report*. Paris: OECD.

Bentolila, S., Michelacci, C., & Suarez, J. (2010). Social contacts and occupational choice. *Economica, 77*(305), 20-45.

Betoret, F. D. (2006). Stressors, Self-Efficacy, Coping resources, and burnout among secondary school teachers in Spain. *Educational Psychology*, *26*(4), 519–539.

Blank, R. K., & De Las Alas, N. (2009). *The effects of teacher professional development on gains in student achievement: How meta-analysis provides scientific evidence useful to education leaders*. Council of Chief State School Officers.

Bloom, N., Lemos, R., Sadun, R., & Van Reenen, J. (2015). Does management matter in schools?. *The Economic Journal*, 125(584), 647-674.

Bogler, R. (2001). The influence of leadership style on teacher job satisfaction. *Educational Administration Quarterly, 37*(5), 662-683.

Bogler, R., & Nir, A. E. (2015). The contribution of perceived fit between job demands and abilities to teachers' commitment and job satisfaction. *Educational Management Administration & Leadership*, *43*(4), 541-560.

Borghans, L., Weel, B. ter, & Weinberg, B. A. (2008). Interpersonal styles and labor market outcomes. *The Journal of Human Resources, 43*(4), 815–858.

Borman, G. D., & Dowling, N. M. (2008). Teacher attrition and retention: A meta-analytic and narrative review of the research. *Review of Educational Research*, *78*(3), 367-409.

Boyd, D., Grossman, P., Ing, M., Lankford, H., Loeb, S., & Wyckoff, J. (2011). The influence of school administrators on teacher retention decisions. *American Educational Research Journal*, *48*(2), 303-333.

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2005). Explaining the short careers of high-achieving teachers in schools with low-performing students. *The American Economic Review*, 95(2), 166-171.

Bueno, C. & Sass, T. (2016). The effects of differential pay on teacher recruitment, retention and quality. In 2016 Fall Conference: The Role of Research in Making Government More Effective. Appam.

Burgess, S. (2015). *Human capital and education: The state of the art in the economics of education*, COEURE: Cooperation on European Research in Economics.

Callanan, G. A., & Greenhaus, J. H. (2008). The baby boom generation and career management: A call to action. *Advances in Developing Human Resources, 10*(1), 70-85.

Cattell, R.B. and Vogelmann, S., (1977). A comprehensive trial of the scree and KG criteria for determining the number of factors. *Multivariate Behavioral Research, 12*(3), 289-325.

Cheema, J. R. (2014). A review of missing data handling methods in education research. *Review of Educational Research*, 84(4), 487-508.

Chetty, R. Friedman, J., Rockoff, J. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood, *American Economic Review, 104*(9), 2633-79.

Chevalier, A., & Dolton, P. (2004). Teacher shortage: Another impending crisis. *CentrePiece*, Winter, 15-21.

Chevalier, A., & Dolton, P. (2004). *The labour market for teachers*. University College Dublin, Centre for Economic Research Working Paper WP04/11.

Chevalier, A., Dolton, P., & McIntosh, S. (2007). Recruiting and retaining teachers in the UK: An analysis of graduate occupation choice from the 1960s to the 1990s. *Economica, 74*(293), 69-96.

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

Clotfelter, C., Ladd, H. F., Vigdor, J., & Wheeler, J. (2006). High-poverty schools and the distribution of teachers and principals. *North-Carolina Law Review, 85*, 1345-1379.

Cobb-Clark, D. A., & Tan, M. (2011). Noncognitive skills, occupational attainment, and relative wages. *Labour Economics, 18*(1), 1-13.

Coldwell, M. (2017). Exploring the influence of professional development on teacher careers: A path model approach. *Teaching and Teacher Education*, *61*, 189-198.

Collie, R. J., Shapka, J. D., & Perry, N. E. (2012). School climate and social–emotional learning: Predicting teacher stress, job satisfaction, and teaching efficacy. *Journal of Educational Psychology*, *104*(4), 1189–1204.

Corbell, K. A., Reiman, A. J., & Nietfeld, J.L. (2008). The perceptions of success inventory for beginning teachers: measuring its psychometric properties. *Teaching and Teacher Education, 24* (6), 1551–1563.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98-104.

Crawford, E. R., LePine, J. A., & Rich, B. L. (2010). Linking job demands and resources to employee engagement and burnout: a theoretical extension and meta-analytic test. *Journal of Applied Psychology*, 95(5), 834.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.

Deakin, G, James, N., Tickner, M. and Tidswell, J. (2010). *Teachers' workload diary survey 2010 (Research Report DfE-RR057)*. London: Department for Education.

Deci, E. L., & Ryan, R. M. (2000). The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227-268.

Deci, E. L. & Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian Psychology, 49*(3), 182-185.

Demerouti, E., Bakker, A. B., Nachreiner, F., & Schaufeli, W. B. (2001). The job demands-resources model of burnout. *Journal of Applied psychology*, *86*(3), 499.

Denton, F. T., Feaver, C. H., & Spencer, B. G. (1994). Teachers and the birth rate. *Journal of Population Economics*, *7*(3), 307-329.

Desa, D., 2016. Understanding non-linear modelling of measurement invariance in heterogeneous populations. *Advances in Data Analysis and Classification*, pp.1-25.

Desa, D., Gonzalez, E. and Mirazchiyski, P. (2014). Construction of scales and indices. In: Belanger, J., Normandeau, S. and Larrakoetxea, E. eds. *TALIS 2013 Technical Report*. OECD, pp.145-295.

Devos, C., Dupriez, V., & Paquay, L. (2012). Does the social working environment predict beginning teachers' self-efficacy and feelings of depression? *Teaching and Teacher Education*, 28(2), 206–217.

DfE [Department for Education] (2010). *A profile of teachers in England from the School Workforce Census 2010*. Department for Education Research Report DFE-RR151.

DfE (2013) Teacher Supply Model: a technical description. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/269894/Teacher_supply_model_-_a_technical_description.pdf

DfE (2016). *National pupil projections – future trends in pupil numbers: July 2016*. Department for Education. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/536880/SFR25Projns_2016_Text.pdf

DfE (2017a). *Graduate labour market statistics 2016*. Department for Education. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/610805/GLMS_2016_v2.pdf

DfE (2017b). *Mathematics early-career payments. Guidance for teachers and schools*. Department for Education. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/650868/Mathematics_early-career_payments-final.pdf

DfE (2017c). *Reducing teacher workload action plan: Update and next steps*. February 2017. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/594680/Teacher_Workload_Action_Plan.pdf

DfE (2017d). *Initial Teacher Training (ITT) Census for the academic year 2017 to 2018, England. SFR 68/2017.*

Dolton, P. J. (2006). Teacher supply. *Handbook of the Economics of Education*, *2*, 1079-1161.

Dolton, P., & Marcenaro-Gutierrez, O. (2011). *2013 Global Teacher Status Index*. Varkey Gems Foundation.

Donaldson, M. L., & Johnson, S. M. (2010). The price of misassignment: The role of teaching assignments in Teach for America teachers' exit from low-income schools and the teaching profession. *Educational Evaluation and Policy Analysis, 32*(2), 299-323.

Dumais, J. & LaRoche, S. (2014). Sampling Design. In: Belanger, J., Normandeau, S. and Larrakoetxea, E. eds. TALIS 2013 Technical Report. OECD, pp.73-84.

Duyar, I., Gumus, S. & Sukru Bellibas, M. (2013). Multilevel analysis of teacher work attitudes: The influence of principal leadership and teacher collaboration. *International Journal of Educational Management*, 27(7), pp.700-719.

Eisenberger, R., Huntington, R., Hutchison, S., & Sowa, D. (1986). Perceived organizational support. *Journal of Applied psychology*, *71*(3), 500.

Eisinga, R., Te Grotenhuis, M., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown?. *International Journal of Public Health*, 58(4), 637-642.

Fackler, S., & Malmberg, L., (2016). Teachers' self-efficacy in 14 OECD countries: Teacher, student group, school and leadership effects. *Teaching and Teacher Education*, 56, 185-195.

Falch, T., Johansen, K., & Strøm, B. (2009). Teacher shortages and the business cycle. *Labour Economics, 16*(6), 648-658.

Favero, N., & Bullock, J. B. (2015). How (not) to solve the problem: An evaluation of scholarly responses to common source bias. *Journal of Public Administration Research and Theory*, *25*(1), 285-308.

Feng, L., & Sass, T. R. (2016). The Impact of Incentives to Recruit and Retain Teachers in "Hard-to-Staff" Subjects. *Journal of Policy Analysis and Management, 37*(1), 112-135.

Fernet, C., Austin, S., Trépanier, S., Dussault, M., & Dussault, M. (2012). How do job characteristics contribute to burnout ? Exploring the distinct mediating roles of perceived autonomy, competence, and relatedness. *European Journal of Work and Organizational Psychology, 22*(2), 123-137.

Figazzolo, L. (2012). Terms and conditions of employment of teachers in relation to teacher shortages and Education for All. ILO/UNESCO.

Filer, R. K. (1986). The role of personality and tastes in determining occupational structure. *ILR Review*, 39(3), 412-424.

Fredriksson, P., Öckert, B., & Oosterbeek, H. (2012). Long-term effects of class size. *The Quarterly Journal of Economics*, 128(1), 249-285.

Fryer Jr, R. G. (2017). *Management and student achievement: Evidence from a randomized field experiment*. National Bureau of Economic Research Working Paper No. w23437.

Furlong, J., Barton, L., Miles, S., Whiting, C., & Whitty, G. (2000). *Teacher education in transition: Re-forming professionalism?* Buckingham: Open University Press

Gaskin, C. J., Lambert, S. D., Bowe, S. J., & Orellana, L. (2017). Why sample selection matters in exploratory factor analysis: implications for the 12-item World Health Organization Disability Assessment Schedule 2.0. *BMC Medical Research Methodology*, 17(1), 40.

Gibson, S., Oliver, L. and Dennison, M. (2015). *Workload challenge: Analysis of teacher consultation responses*. Department for Education Research Report DfE-RR445. London: Department for Education.

Glazerman, S., Isenberg, E., Dolfin, S., Bleeker, M., Johnson, A., Grider, M., & Jacobus, m. (2010). *Impacts of comprehensive teacher induction: Final results from a randomized controlled study (NCEE 2010-4027)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Gibbons, S., Scrutinio, V., & Telhaj, S. (2018). *Teacher turnover: Does it matter for pupil achievement?*. Centre for Economic Performance Discussion Paper No. dp1530.

Goldhaber, D., & Liu, A. (2003). *Occupational choices and the academic proficiency of the teacher workforce*. Washington, DC: Department of Education, National Center for Education Statistics.

Goldhaber, D., Krieg, J., & Theobald, R. (2014a). Knocking on the door to the teaching profession? Modelling the entry of prospective teachers into the workforce. *Economics of Education Review*, 43, 106-124.

Goldhaber, D., Krieg, J., Theobald, R., & Brown, N. (2014b). *The STEM and special education teacher pipelines: Why don't we see better alignment between supply and demand*? CEDR Working Paper 2014-3.

Goldschmidt, P., & Phelps, G. (2010). Does teacher professional development affect content and pedagogical knowledge: How much and for how long?. *Economics of Education Review*, 29(3), 432-439.

Graziano, W., & Tobin, R. (2017). Agreeableness and the Five Factor Model. In Thomas A. Widiger (Ed.), *The Oxford Handbook of the Five Factor Model*. New York: Oxford University Press.

Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement*, 37, 827-838.

Ham, R., Junankar, P. N., & Wells, R. (2009). *Occupational choice: Personality matters*. IZA Discussion Paper No. 4105.

Hampson, S. E., & Goldberg, L. R. (2006). A first large cohort study of personality trait stability over the 40 years between elementary school and midlife. *Journal of Personality and Social Psychology*, 91(4), 763.

Hamre, B., & Pianta, R. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure?. *Child Development*, 76(5), 949-967.

Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review*, 30(3), 466-479.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Why public schools lose teachers. *Journal of Human Resources*, 39(2), 326–354.

Higton, J., Leonardi, S., Richards, N., Choudhoury, A., Sofroniou, N., & Owen, D. (2017). *Teacher workload survey 2016. DfE research report rr633*. London: Department for Education.

Holgado–Tello, F. P., Chacón–Moscoso, S., Barbero–García, I., & Vila–Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity*, 44(1), 153-166.

Howe, N., & Strauss, W. (2009). *Millennials rising: The next great generation*. Vintage.

Howson, J., & McNamara, O. (2012). Teacher workforce planning: the interplay of market forces and government polices during a period of economic uncertainty. *Educational Research*, 54(2), 173-185.

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal*, 38(3), 499–534.

Ingersoll, B. R., & Kralik, J. M. (2004). *The impact of mentoring on teacher retention: What the research says*. Education Commission of the States Review.

Ingersoll, R. M., & Perda, D. (2010). Is the supply of mathematics and science teachers sufficient?. *American Educational Research Journal*, 47(3), 563-594.

Jackson, C. (2016). *What do test scores miss? The importance of teacher effects on non-test score outcomes*. National Bureau of Economic Research Working Paper No. w22226.

Jackson, J., & Roberts, B. (2017). Conscientiousness. In Thomas A. Widiger (Ed.), *The Oxford Handbook of the Five Factor Model*. New York: Oxford University Press.

Jacob, R., Goddard, R., Kim, M., Miller, R., & Goddard, Y. (2015). Exploring the causal impact of the McREL Balanced Leadership Program on leadership, principal efficacy, instructional climate, educator turnover, and student achievement. *Educational Evaluation and Policy Analysis*, 37(3), 314-332.

Jeong, U. Y. (2009). *Teacher policy in England: an historical study of responses to changing ideological and socio-economic contexts*. PhD Thesis Bath: University of Bath.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of Personality: Theory and research*, 2, 102-138.

Johnson, S. M., Kraft, M. A., & Papay, J. P. (2012). How context matters in high-need schools: The effects of teachers' working conditions on their professional satisfaction and their students' achievement. *Teachers College Record*, 114(10), 1-39.

Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark IV. *Educational and Psychological Measurement*, 34(1), 111-117.

Kelly, S. (2004). An event history analysis of teacher attrition: Salary, teacher tracking, and socially disadvantaged schools. *The Journal of Experimental Education*, *72*(3), 195-220.

Kish, L., & Frankel, M. R. (1970). Balanced repeated replications for standard errors. *Journal of the American Statistical Association*, 65(331), 1071-1094.

Knibbs, S., Mollidor, C., Lindley, L., Allen, R., & Sims, S. (2017) *Evaluation of the high-potential middle leaders programme: Secondary*. London: National College for Teaching and Leadership.

Knoll, S., & Koršnáková, P. (2014). In: Belanger, J., Normandeau, S. & Larrakoetxea, E. Eds. TALIS 2013 Technical Report. Paris: OECD.

Kolenikov, S. & Angeles, G. (2004). *The use of discrete data in PCA: theory, simulations, and applications to socioeconomic indices*. Chapel Hill: Carolina Population Center, University of North Carolina.

Kraft, M. (2018). Teacher effects on complex cognitive skills and social-emotional competencies. *Journal of Human Resources*, 0916-8265R3.

Kraft, M. A., Marinell, W. H., & Shen-Wei Yee, D. (2016). School organizational contexts, teacher turnover, and student achievement: Evidence from panel data. *American Educational Research Journal*, 53(5), 1411-1449.

Krueger, A. B., & Schkade, D. (2008). Sorting in the labor market: do gregarious workers flock to interactive jobs? *Journal of Human Resources*, 43(4), 859-883.

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity*, 47(4), 2025-2047.

Kukla-Acevedo, S. (2009). Leavers, movers, and stayers: The role of workplace conditions in teacher mobility decisions. *The Journal of Educational Research*, 102(6), pp. 443-452.

Kutaka, T. S., Smith, W. M., Albano, A. D., Edwards, C. P., Ren, L., Beattie, H. L., ... & Stroup, W. W. (2017). Connecting teacher professional development and student mathematics achievement: a 4-year study of an elementary mathematics specialist program. *Journal of Teacher Education*, 68(2), 140-154.

Labaree, D. F. (2000). On the nature of teaching and teacher education: Difficult practices that look easy. *Journal of Teacher Education*, *51*(3), 228-233.

Ladd, H. F. (2011). Teachers' perceptions of their working conditions: how predictive of planned and actual teacher movement? *Educational Evaluation and Policy Analysis*, 33(2), 235–261.

Lakshmanan, A., Heath, B. P., Perlmutter, A., & Elder, M. (2011). The impact of science content and professional learning communities on science teaching efficacy and standards-based instruction. *Journal of Research in Science Teaching*, 48(5), 534-551.

Loewenberg Ball, D., & Forzani, F. M. (2009). The work of teaching and the challenge for teacher education. *Journal of Teacher Education*, *60*(5), 497-511.

Luft, J. A., Dubois, S. L., Nixon, R. S., & Campbell, B. K. (2015). Supporting newly hired teachers of science: Attaining teacher professional standards. *Studies in Science Education*, 51(1), 1-48.

Lynch, S., Worth, J., Bamford, S., & Wespieser, K. (2016). *Engaging teachers: NFER* analysis of teacher retention. Slough: NFER.

Lynn, P., & Kaminska, O. (2010). *Weighting strategy for Understanding Society*. Understanding Society Working Paper Series, No. 2010-05. Institute for Social and Economic Research.

Lyons, S. T., Duxbury, L. E., & Higgins, C. A. (2006). A comparison of the values and commitment of private sector, public sector, and parapublic sector employees. *Public Administration Review*, 66(4), 605-618.

MAC [Migration Advisory Committee] (2016). *Partial review of the Shortage Occupation List: Review of Teachers*. London: Migration Advisory Committee.

Malinen, O., & Savolainen, H. (2016). The effect of perceived school climate and teacher efficacy in behavior management on job satisfaction and burnout: A longitudinal study. *Teaching and Teacher Education*, 60, 144–152.

Mannheim, K. (1970). The problem of generations. *Psychoanalytic Review*, 57(3), 276-322.

Marinell, W., & Coca, V. (2013). *Who stays and who leaves? Findings from a three-part study of teacher turnover in NYC middle schools*. New York: Research Alliance for New York City Schools.

Matthias, C. (2014). *Qualitative research with shortage subject teaching candidates: The journey to teacher training*. National College for Teaching & Leadership.

McAdams, D. P., & Pals, J. L. (2006). A new Big Five: Fundamental principles for an integrative science of personality. *American Psychologist*, 61(3), 204-217.

Meier, K. J., & O'Toole, L. J. (2010). *Organizational performance: Measurement theory and an application: Or, common source bias, the Achilles heel of public management research*. APSA 2010 Annual Meeting Paper.

Melero, E. (2010). Training and promotion: allocation of skills or incentives?. *Industrial Relations: A Journal of Economy and Society*, 49(4), 640-667.

Menon, M. E., & Saitis, C. (2006). Satisfaction of pre-service and in-service teachers with primary school organization: Evidence from Greece. *Educational Management Administration & Leadership*, *34*(3), 345-363.

Micklewright, J., Jerrim, J., Vignoles, A., Jenkins, A., Allen, R., Ilie, S., Bellabre, F., Barrera, F., & Hein, C. (2014). *Teachers in England's secondary schools: Evidence from TALIS 2013*. London: Department for Education.

Mocetti, S. (2012). Educational choices and the selection process: before and after compulsory schooling. *Education Economics*, 20(2), 189-209.

Moor, H., Jones, M., Johnson, F., Martin, K., Cowell, E., & Bojke Pharmerit, C. (2006). *Mathematics and science in secondary schools: the deployment of teachers and support staff to deliver the curriculum*. Department for Education and Skills Research Report RR708. London: Department for Education and Skills.

Müller, K., Alliata, R., & Benninghoff, F. (2009). Attracting and retaining teachers: A question of motivation. *Educational Management Administration & Leadership*, 37(5), 574-599.

Mutz, D. C. (2011). *Population-based survey experiments*. Princeton University Press.

NAO [National Audit Office] (2016). *Training new teachers*. London: National Audit Office.

Nguni, S., Sleegers, P., & Denessen, E. (2006). Transformational and transactional leadership effects on teachers' job satisfaction, organizational commitment, and organizational citizenship behavior in primary schools: The Tanzanian case. *School Effectiveness and School Improvement*, 17(2), 145-177.

Nieken, P., & Störmer, S. (2010). *Personality as predictor of occupational choice: Empirical evidence from Germany. Diskussionspapiere des Schwerpunktes Unternehmensführung am Fachbereich BWL 2(8)*.

OECD (2014). *TALIS 2013 results: an international perspective on teaching and learning*. Paris: OECD.

Ofsted (2015). *The annual report of Her Majesty's Chief Inspector of Education, Children's Services and Skills 2014/15*. London: Ofsted.

ONS [Office for National Statistics] (2015). *Trends in births and deaths over the last century.* Office for National Statistics. Accessed from: https://visual.ons.gov.uk/birthsanddeaths/

Ost, B., & Schiman, J. (2015). Grade-specific experience, grade reassignments, and teacher turnover. *Economics of Education Review,* (46), 112–126.

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

Parker, M. A., Ndoye, A., & Imig, S. R. (2009). Keeping our teachers! Investigating mentoring practices to support and retain novice educators. *Mentoring & Tutoring: Partnership in Learning*, 17(4), 329–341.

Parry, E., & Urwin, P. (2011). Generational differences in work values: a review of theory and evidence. *International Journal of Management Reviews*, 13(1), 79–96.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, 63, 539-569.

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management*, *12*(4), 531-544.

Polly, D., McGee, J., Wang, C., Martin, C., Lambert, R., & Pugalee, D. K. (2015). Linking professional development, teacher outcomes, and student achievement: The case of a learner-centered mathematics program for elementary school teachers. *International Journal of Educational Research*, 72, 26-37.

Preacher, K. J., & MacCallum, R.C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics: Statistical issues in psychology, education, and the social sciences*, 2(1), pp.13-43.

Quimby, J. L., & Santis, A. M. (2006). The influence of role models on women's career choices. *The Career Development Quarterly*, 54(4), 297-306.

Ranguelov, S., De Coster, I., Norani, S., & Paolini, G. (2012). *Key Data on Education in Europe 2012*. Education, Audiovisual and Culture Executive Agency, European Commission.

Reback, R. (2004). The impact of college course offerings on the supply of academically talented public school teachers. *Journal of Econometrics*, 121(1), 377-404.

Reeves, P. M., Pun, W. H., & Chung, K. S. (2017). Influence of teacher collaboration on job satisfaction and student achievement. *Teaching and Teacher Education*, 67, 227–236.

Richardson, P. W., & Watt, H. M. (2006). Who chooses teaching and why? Profiling characteristics and motivations across three Australian universities. *Asia-Pacific Journal of Teacher Education*, 34(1), 27-56.

Roness, D., & Smith, K. (2009). Postgraduate Certificate in Education (PGCE) and student motivation. *European Journal of Teacher Education*, 32(2), 111-134.

Roness, D., & Smith, K. (2010). Stability in motivation during teacher education. *Journal of Education for Teaching*, 36(2), 169-185.

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal*, 50(1), 4-36.

Ross, J., & Bruce, C. (2007). Professional development effects on teacher efficacy: Results of randomized field trial. *The Journal of Educational Research*, 101(1), 50-60.

Rothstein, J., & Rouse, C. E. (2011). Constrained after college: Student loans and early-career occupational choices. *Journal of Public Economics*, 95(1), 149-163.

Rutkowski, L. & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement*, 74(1), 31-57.

Scafidi, B., Sjoquist, D. L., & Stinebrickner, T. R. (2007). Race, poverty, and teacher mobility. *Economics of Education Review*, *26*(2), 145-159.

Schanzenbach, D. W. (2006). What have researchers learned from Project STAR?. *Brookings Papers on Education Policy*, (9), 205-228.

Schleicher, A. (2011). *Building a High-Quality Teaching Profession: Lessons from around the World*. Pairs: OECD Publishing.

Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy Scale. In J. Weinman, S. Wright, & M. Johnston Eds. *Measures in health psychology: A user's portfolio*. pp. 35-37. NFER-NELSON.

Sellen, P., (2016). *Teacher workload and professional development in England's secondary schools: Insights from TALIS.* London: Education Policy Institute.

Shugart, S. S., & Hounshell, P. B. (1995). Subject matter competence and the recruitment and retention of secondary science teachers. *Journal of Research in Science Teaching*, 32(1), 63-70.

Simon, N. S., & Johnson, S. M. (2015). Teacher turnover in high-poverty schools: What we know and can do. *Teachers College Record*, 117(3), 1-36.

Sims, S. (2016). *High-stakes accountability and teacher turnover: How do different school inspection judgements affect teacher's decisions to leave their school?* Department of Quantitative Social Science Working Paper No. 16-14, UCL, London.

Sims, S. (2017). *TALIS 2013: Working conditions, teacher job satisfaction and retention*. Department for Education Statistical Working Paper. London: Department for Education.

Sims, S. (2018a, 29th March). 'The DfE claims teacher retention is "broadly stable" – it's a whopping great fib'. tes.

Sims, S. (2018b). *What happens when you pay shortage-subject teachers more money? Simulating the effect of early-career salary supplements on teacher supply in England*. London: Gatsby.

Sinclair, C. (2008). Initial and changing student teacher motivation and commitment to teaching. *Asia-Pacific Journal of Teacher Education*, 36(2), 79-104.

Skaalvik, E. M., & Skaalvik, S. (2009). Does school context matter? Relations with teacher burnout and job satisfaction. *Teaching and Teacher Education*, 25(3), 518-524.

Skaalvik, E. M., & Skaalvik, S. (2011). Teacher job satisfaction and motivation to leave the teaching profession : Relations with school context, feeling of belonging, and emotional exhaustion. *Teaching and Teacher Education*, 27(6), 1029–1038.

Skaalvik, E. M., & Skaalvik, S. (2017). Motivated for teaching? Associations with school goal structure, teacher self-efficacy, job satisfaction and emotional exhaustion. *Teaching and Teacher Education*, 67, 152–160.

Slater, H., Davies, N. M., & Burgess, S. (2012). Do teachers matter? Measuring the variation in teacher effectiveness in England. *Oxford Bulletin of Economics and Statistics*, *74*(5), 629-645.

Smith, T. M., & Ingersoll, R. M. (2004). What are the effects of induction and mentoring on beginning teacher turnover? *American Educational Research Journal*, 41(3), 681–714.

Smithers, A. & Robinson, P. (2000). *Coping with teacher shortages*. Centre for Education and Employment Research, University of Buckingham.

Smithers, A., & Robinson, P. (2008). *Physics in schools IV: Supply and retention of teachers*. Centre for Education and Employment Research, University of Buckingham.

Strauss, W., & Howe, N. (1991). *Generations: The history of America's future, 1584 to 2069*. New York: William Morrow.

Sutin, A. (2017). Openness. In Thomas A. Widiger (Ed.), *The Oxford Handbook of the Five Factor Model*. New York: Oxford University Press.

STRB [School Teachers Review Body] (2017). *School teachers review body twenty-seventh report.* School Teachers Review Body.

Sutin, A. R., & Costa, P. T., Jr. (2010). Reciprocal influences of personality and job characteristics across middle adulthood. *Journal of Personality*, *78(1)*, 257–288.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53-55.

Taylor, M., Yates, A., Meyer, L. H., & Kinsella, P. (2011). Teacher professional leadership in support of teacher professional development. *Teaching and Teacher Education*, 27(1), 85-94.

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research*, *83*(3), 357-385.

Understanding Society (2015) *Long Term Content Plan*. University of Essex, Institute for Social and Economic Research.

University of Essex. Institute for Social and Economic Research, NatCen Social Research and Kantar Public, [producers]: Understanding Society: Waves 1-6, 2009-2015 [computer file]. 8th Edition. Colchester, Essex: UK Data Service [distributor], November 2016. SN: 6614

van Driel, J. H., Meirink, J. A., van Veen, K., & Zwart, R. C. (2012). Current trends and missing links in studies on teacher professional development in science education: A review of design features and quality of research. *Studies in Science Education*, *48*(2), 129-160.

Watt, H. M., & Richardson, P. W. (2007). Motivational factors influencing teaching as a career choice: Development and validation of the FIT-Choice scale. *The Journal of Experimental Education*, 75(3), 167-202.

Watt, H. M., Richardson, P. W., Klusmann, U., Kunter, M., Beyer, B., Trautwein, U., & Baumert, J. (2012). Motivations for choosing teaching as a career: An international comparison using the FIT-Choice scale. *Teaching and Teacher Education*, 28(6), 791-805.

Weiss, E. M. (1999). Perceived workplace conditions and first-year teachers' morale, career choice commitment, and planned retention: a secondary analysis. *Teaching and Teacher Education,* 15(8): 861–879.

White, P., & Smith, E. (2005). What can PISA tell us about teacher shortages?. *European Journal of Education*, 40(1), 93-112.

Wolstenholme, C., Coldwell, M. & Stevens, A. (2012). *The impact of Science Learning Centre continuing professional development on teachers' retention and careers: final report*. Sheffield: CEIR.

Wong, M., Cook, T. D., & Steiner, P. M. (2015). Adding design elements to improve time series designs: No Child Left Behind as an example of causal pattern-matching. *Journal of Research on Educational Effectiveness*, *8*(2), 245-279.

World Bank. (2013). *Achieving learning for all*. New York: World Bank.

Worth, J. & De Lazzari, G. (2017). *Teacher* Retention and Turnover Research Research. *Update 1: teacher retention by subject*. Slough: National Foundation for Educational Research.

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology*, *9*(2), 79-94.

You, S., & Conley, S. (2015). Workplace predictors of secondary school teachers' intention to leave: An exploration of career stages. *Educational Management Administration & Leadership*, *43*(4), 561-581.

Zieger, L., Sims, S., Jerrim, J. (2018) Comparing the quality of working conditions in schools across ten countries. Unpublished manuscript.

Zweimüller, J., & Winter-Ebmer, R. (2003). On-the-job-training, job search and job mobility. *Swiss Journal of Economics and Statistics*, 139(4), 563-576.