IDENTIFYING SCHOOLS WITH HIGH USAGE AND HIGH LOSS OF NEWLY QUALIFIED TEACHERS

Rebecca Allen* and Sam Sims**

In England, teacher shortages have worsened in recent years and one contributor is the declining rates of retention among newly qualified teachers (NQTs). We employ a method developed in the health-statistics literature to identify schools that both recruit an unusually high level of NQTs and lose an unusually high level of NQTs from the profession. We show that this small group of schools, which are likely characterised by poor working conditions, are responsible for a disproportionately large amount of attrition from the teaching profession. This has a material effect on overall teacher shortages and comes at a high cost to taxpayers. Policy solutions, including improving the flow of information to NQTs to help them avoid such schools, are discussed.

* UCL Institute of Education. Email: rebecca.allen@ucl.ac.uk

** UCL Institute of Education. Email: s.sims@ucl.ac.uk

1. Introduction

Research consistently shows that the quality of teaching is among the most important school-based determinants of pupil learning (Hanushek and Rivkin, 2012; Burgess, 2015). Moving a child from an average to a top teacher means they will learn in six months what would otherwise have taken twelve (Wiliam, 2016). Moreover, good teachers seem to have a disproportionately strong impact on pupils from disadvantaged homes, suggesting that they can help close socio-economic gaps in pupil attainment (Hamre and Pianta, 2005; Slater *et al.,* 2012). Teachers also have an important impact on pupils' non-cognitive abilities, or socio-emotional competencies (Jackson, 2016; Kraft, 2017). The beneficial effects of good teachers also extend beyond the school gates. Chetty *et al.,* (2011) estimate that moving a class of children from a bottom ventile to a median quality teacher is associated with increased lifetime earnings for that class of around a quarter of a million dollars. Ensuring a sufficient supply of high-quality teachers should therefore be a priority for education policymakers.

Despite this, shortages of teachers are a common problem in public school systems around the world (Dolton, 2006). Indeed, when the Teaching and Learning International Survey asked school leaders in 34 countries what they perceived to be the biggest constraint on improving the quality of teaching, the most frequent response was the lack of appropriately qualified staff (OECD, 2014). In recent years, demographic trends and the economic cycle have contributed to an increase in the severity of these shortages in many countries (Dolton*,* 2006; Dolton and van der Klaauw, 1999; Falch *et al*., 2009). In England, the context for this study, the government recently commissioned a detailed review of teacher supply from the Migration Advisory Committee, led by Professor Alan Manning. Their 162-page report concluded that maths, science (especially physics), computer science and modern foreign language teachers are all in short supply (MAC, 2016). Figures from the government's own Teacher Supply Model also indicate that there has been a shortage of maths teachers in every year since 2012 and an overall shortage of science teachers every year since 2013 (Sims, 2018).

These shortages are partly due to declining retention among early career teachers (Sims, 2018), which brings with it a distinctive set of problems. Research has repeatedly shown that teachers acquire skills quickly during the first three to five years of their career (Wiswall, 2013; Papay and Kraft, 2015; Henry *et al.,* 2012).[i] Early career attrition from the teaching profession therefore requires inexperienced teachers to be replaced with new inexperienced teachers, which holds down the average level of experience across the school system. Training the replacement teachers is also a costly activity for the taxpayer. The total cost of teacher training in 2013/14 was £700m (National Audit Office, 2016) and the average cost per teacher is £23,000 (Allen *et al*., 2016). High early-career teacher attrition therefore represents a poor return on this public investment in training new teachers.

The reasons for declining retention of early-career teachers are not entirely clear and are mirrored in other public sector professions such as nursing, where one-third of newly qualified nurses do not join the National Health Service in England (Oulton, 2006). Quality of working conditions, i.e. the degree of balance between support provided and demands made on employees, appears to be critical in both teaching and nursing. Empirical studies consistently find the most important determinant of teacher retention is the quality of working conditions in schools (Simon and Johnson, 2015; Kraft et al., 2016; Sims, 2017). The inter-personal characteristics of the working environment, such as the quality of school leadership and extent of teacher collaboration, appear to be particularly important (Simon and Johnson, 2015). Moreover, two studies which analysed experienced teachers and early-career teachers separately found that working conditions have a stronger influence on the retention of those that have recently qualified (Kukler-Acavedo, 2009; Boyd et al., 2010). Detailed qualitative research suggests that this is because, while experienced teachers already have the skills and confidence necessary to succeed even in spite of bad working conditions, early-career teachers are less likely to develop even basic skills and competencies while working in an unsupportive school (Hong, 2012; Johnson and Birkeland, 2003a).

This has led to concerns among researchers (Holme et al., 2017) that certain schools are experiencing a vicious cycle in which poor working conditions lead to poor retention, which undermines the support networks within schools, resulting in further turnover, and so on. For a case study of one such school, see Allen and Sims (forthcoming). Holme et al. (2017) identify 4.4% of Texas schools as displaying this sort of chronic instability while Allensworth et al. (2009) find that 14.8% of Chicago schools display such characteristics. High turnover of this sort is undesirable because it damages pupil attainment through the disruption it causes (Ronfeldt et al., 2012; Atteberry et al., 2016) and has a disproportionate effect on disadvantaged pupils (Allen et al, 2017).

In this paper, we adapt a simple technique from the medical statistics literature - funnel plots - in order to identify schools that both use and lose an unusually high number of newly qualified teachers (NQTs). Funnel plots allow us to quantify what counts as "unusually high" levels of use and loss of NQTs and avoid misclassifying small schools as displaying chronic instability due to the inherent volatility of recruitment and turnover in schools with lower headcounts. We identify 122 schools (0.5% of all schools in England) that consistently both use and lose an unusually high level of NQTs. The number of NQTs that left the profession from these schools over this period is 577, which reflects an NQT attrition rate three times higher than the national average. We show that if these schools had attrition rates equivalent to the average school then 376 additional teachers would have progressed beyond their NQT year. For context, this is equivalent to 22% of the nationwide shortfall

of teachers in 2015. Alternatively, the costs of training these NQTs who leave these schools amounts to £11m.

Our analysis suggests that a small number of schools are, wittingly or unwittingly, consistently running a 'recruit-burnout-replace' staffing model. Given that there is currently a shortage of teachers in England, and teachers generally do not have trouble finding work (Hobson *et al.,* 2009), this suggests that teachers are unable to reliably judge the quality of working conditions when selecting a school to work in. The article concludes with a discussion of policy options which could help increase the flow of information to teachers in order to better inform their choice of where to work, and in doing so make a recruit-burnout-replace approach to school staffing unsustainable for schools, thus forcing them to improve working conditions for teachers instead. We argue that publishing funnel plots for NQT wastage rates would be a simple, reliable and low cost way of achieving this goal.

2. Data

Our analysis uses the School Workforce Census, an annual census of all staff in state-funded schools in England, collected from November 2010 onwards. We use five sweeps of data from 2010 to 2014. These five sweeps have been "patched" using information from the Database of Teacher Records pension data, meaning that there is only a very small chance that missing data for a teacher in a particular year is misclassified as turnover. We do not use the latest sweep of the data because at the time of writing it had not yet been patched in this way. Teachers are consistently identified across all five years in the panel using their Teacher Reference Number (and supplemented by information from the Database of Teacher Records and/or fuzzy matching on names and birth dates where the Teacher Reference Number is missing). We restrict our analysis to those teachers in primary and secondary schools who work for at least 1 day a week in one or more of these years and to schools with at least seven teachers in each sweep of the School Workforce Census. This yields between 458,307 and 465,020 teachers in each of the five years, of whom around 5% each year are identified as NQTs. We identify 101,816 NQTs across the five years 2010-2014.

The School Workforce Census provides basic demographic information on each teacher, including their gender, age, ethnicity and qualifications. For this study we simply use an indicator of whether or not they have Qualified Teacher Status and the year in which they obtained Qualified Teacher Status to establish whether or not they are an NQT. We use the school identifier to show whether teachers are moving school at the end of a year, taking care to ensure that school re-organisations and mergers are not inadvertently identified as a job move by the teacher. We can also observe when a teacher

stops working at a state-funded school in England because they do not reappear in the dataset in the following year. We are interested in both whether the NQT leaves their first school at the end of their first year (turnover) and whether an NQT leaves the profession at the end of their first year (attrition).

Figure 1 shows the retention of each new cohort of NQTs, expressed as the proportion still working in state-funded schools in England. It shows that, prior to 2008, trainee cohorts displayed fairly stable levels of retention, following a broadly similar downward trajectory. From 2008 onwards however, the pattern changes, with each new cohort following a steeper downward trajectory. Retention of early-career teachers has therefore been declining for around nine years.

Figure 1: NQT Retention by Cohort
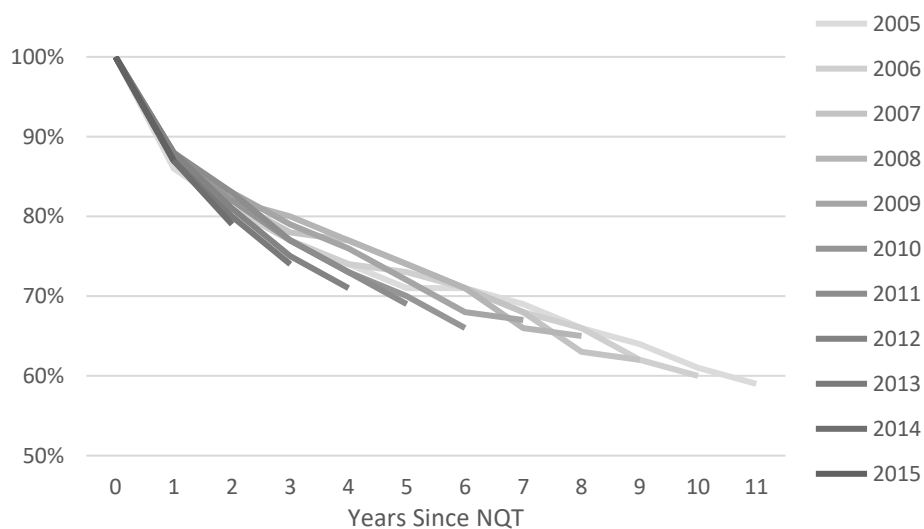
Note: We show every other cohort for clarity. Note that 2010 cohort
denotes the group of trainees who were in their NQT year in the 201/11 academic year.
Source: Initial Teacher Training Main Census.

Around 4% of teachers in any one year are NQTs. However these NQTs are distributed very unevenly across schools. Figure 2 shows the distribution of NQT use among primary and secondary schools, expressed as the proportion of all staff slots between 2010 and 2014. Assuming class sizes and teaching contact hours within schools are approximately equal, this can be thought of as the probability that a child within the school is taught by a newly qualified teacher.[1] On the far left of Figure 2 there are 1292 schools that did not employ a single NQT over the six-year period from 2010 to 2014. These schools are extreme examples of the wider group of 4311 schools who use disproportionately few (less than 4%) of NQTs. This group is however much smaller than the group of

---

[1] Strictly speaking, this would be an overestimate, since NQTs only teach 90% of the class hours of a fully qualified teacher.
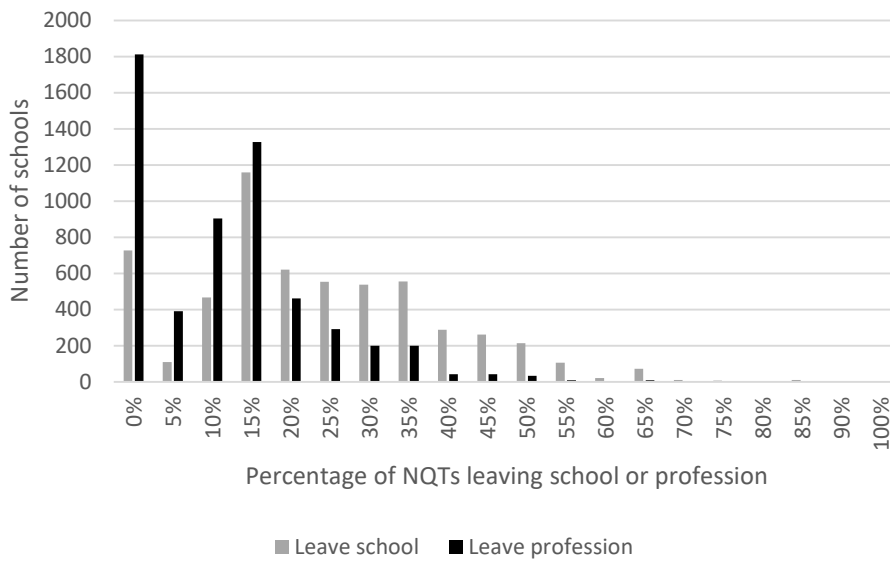
7724 schools that use disproportionately high (more than 4%) levels of NQTs. On the right hand side of the distribution there is a long tail of 1591 schools who have more than 10% of the workforce made up by NQTs across this period. There are just 38 schools where pupils are typically exposed to the classroom teaching of an NQT for over 1-in-5 classes.

Figure 2: Proportion of teachers in NQT year 2010-14 for all state-funded schools ~~[AND CHANGE AXIS ON FIGURE TO SAY 2014]~~
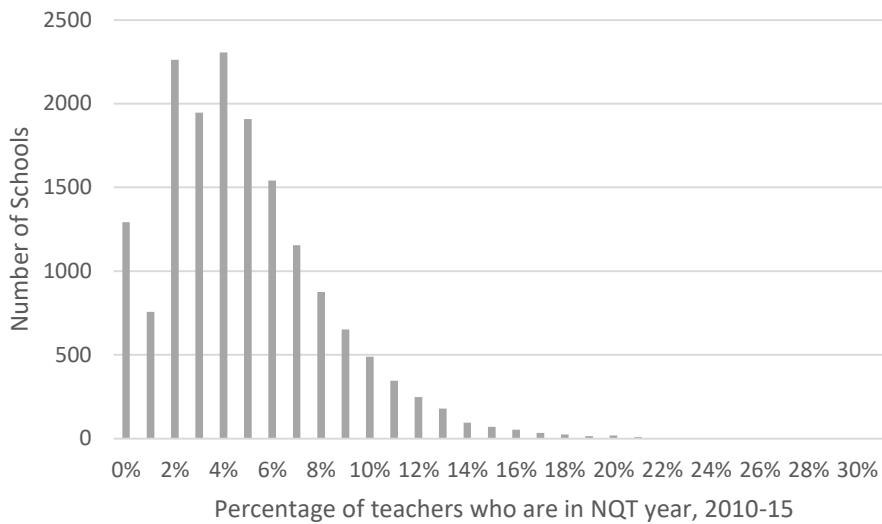
As well as looking at the use of NQTs by schools, we are interested in the rate at which NQTs leave schools. The proportion of NQTs that leave the profession has been between 12% and 13% per annum for each cohort since 2010 (Figure 1). **Error! Reference source not found.** shows the distribution of average 5-year NQT turnover and attrition rates across all schools that had six or more NQTs over this period. Again, the distribution is very uneven. There are 727 schools that have lost no NQTs over the five year period. Again, these are part of a wider group of 3106 schools who have lost disproportionately few NQTs (<12.5%). By comparison, there are 2615 schools who have experienced disproportionately high loss of NQTs. There are far fewer schools that experience zero loss of NQTs to other schools, relative to losing them from the profession. Moreover, there is a fatter tail of schools with high levels of loss to other schools, relative to the tail of schools with high levels of loss from the profession.

Figure 3: Proportion of teachers in NQT year that leave school or profession 2010-14 for all state-funded schools

Number of schools

2000
1800
1600
1400
1200
1000
800
600
400
200
0

0% 5% 10% 15% 20% 25% 30% 35% 40% 45% 50% 55% 60% 65% 70% 75% 80% 85% 90% 100%

Percentage of NQTs leaving school or profession

■ Leave school ■ Leave profession

3. Identifying high use and high loss of NQTs

Our challenge is to decide when the variation in NQT use/loss that we observe in Figure 2 and **Error! Reference source not found.** reflects more than just random variation. A simple approach to identifying schools that overuse NQTs would be to rank schools based on the proportion of NQTs that they use/lose and publish these as a league table. It is plausible that this would compel poorly ranked schools to improve their working conditions for NQTs either through a 'naming and shaming' dynamic (Bevan and Wilson 2013) or through a market mechanism in which poorly ranked schools are forced to improve in order to compete for teachers who can observe school's position in the ranking. However, this approach is problematic because the ranking is unlikely to reflect underlying differences in the working conditions in schools. To see why, consider that a proportion of the variation on which the ranking is based will reflect natural, random variation in NQT use/loss from year to year. Because primary schools often have low teacher headcounts (e.g. ten members of teaching staff), even a small amount of random (unrelated to working conditions) variation in the number of NQTs who leave a school (e.g. from 1 of 2 NQTs to 2 of 2 NQTs) will manifest in substantial changes in the percentage of NQT loss (e.g. from 50% to 100%). League tables would then be based on a spurious ranking largely reflecting random variation. Moreover, in regions of the rankings which are densely packed with schools, even small changes in the percentage of NQTs lost could translate into large changes in the rank order of schools. Using more years of data would reduce volatility at the cost of using information that may be less relevant, given that working conditions in schools can change quite quickly.

Similar analytical challenges are common in manufacturing, where engineers are concerned with creating a sufficiently uniform product, and in medicine, where those involved in clinical governance are concerned with eliminating poor quality medical care across practitioners or establishments. In the manufacturing context, Wiliam Shewhart (1925) developed a distinction between common-cause and special-cause variation which has proved useful for our purposes in this article. Common-cause variation is the natural variation exhibited by any *stable* process. For example a given production line will produce some variation in product, which can be thought of as a random draw from the *stable distribution* of the product that the production line creates. This variation can be affected only by changing the underlying process - altering or replacing the production line. However it cannot be eliminated since it represents the variation which is inherent to any given production line. Special cause variation, by contrast, is the result of factors extrinsic to the process, and can therefore only be affected by identification of the special cause. For example, a heat wave during summer might affect the output from a production line. The extreme heat is extrinsic to the manufacturing process and is not part of its stable operation. It can be eliminated through action on the special causes of the process - e.g. shading the windows of the plant – but not through action on the production line itself.

The concepts of common- and special-cause variation were later embodied in control charts. These charts plot the mean value of a process along with "control limits" around that mean - usually plus or minus three standard deviations (Shewhart, 1931). Because common-cause variation is assumed to result from random variation inherent to a stable process, and the probability of a random draw occurring a given a distance from the mean is known to occur with a certain probability, this allows the user to quantify the probability that a given observation has occurred as a result only of common-cause variation. Control charts were later adopted in medicine (see Mohammed *et al.,* 2001) and adapted into funnel plots (Spiegelhalter, 2005). The latter allow the control limits to vary with the size of an institution, in order to account for the precision of the estimate. For example, annual death rates after a certain procedure from a large hospital are a more precise estimate of the underlying quality of the procedure in that hospital, relative to annual death rates from a small hospital.

These ideas can also be applied to the use and loss of NQTs by schools. Common-cause variation in the use of NQTs would capture the variation in school staffing needs due to e.g. a group of teachers reaching retirement at the same time, as well as natural variation in the experience profile of the pool of applicants in the local labour market e.g. a high numbers of NQTs looking for a job in a particular year. Common-cause variation in loss of NQTs would capture the natural variation in trainees' capability and enjoyment of the job during their first year, for example. All of these

constitute common cause variation to the extent that the number of teachers retiring or NQTs looking for jobs are random variation in a stable underlying process. Special-cause variation, by contrast, would result from a non-stable process of school staffing need. We suggest that one important cause of such need is the vicious-cycle dynamic described in the introduction by which turnover damages the working conditions within a school, particularly the collegial support structures, which itself leads to further turnover, and so on. The variation observed here is not resulting from a random draw from a stable underlying distribution but is instead the result of a changing underlying distribution.

Figure 4 shows a funnel plot applied to the use of NQTs by schools. The plot was created using the `funnelcompar` command in Stata (Gini and Forni, 2009). It plots the proportion of teachers who are in their NQT year at a school, against a count of the number of teacher slots in that school, between 2010 and 2014, for each secondary school in a given area. Each year is counted separately, so if a school typically employs 100 teachers at one time then it has 500 teacher slots across five years. The dashed line shows the average proportion of teacher slots filled by NQTs across the time period. This functions as our estimate of the proportion of NQTs used by schools displaying only common cause variation. The solid lines show the control limits containing the region in which we would expect 95% of random draws from a binomial distribution centred on this mean to fall.[2]. We create one plot per region of England. Our plots therefore make contextualised comparisons of schools within a region and phase where local labour market conditions are similar.

We do not explicitly condition on or account for other observable characteristics of the school that might be associated with high NQT-use through e.g. regression. The types of pupil demographic information which we could use to contextualise our estimates in this way are indeed associated with e.g. high recruitment of NQTs (Allen & Sims, 2018) and low retention (Allen & Burgess 2017). Conditioning on these characteristics of schools would therefore change our conclusions. Crucially however, pupil demographics and other information which is available in administrative datasets has been shown to have either very little or zero predictive power for retention, *if this is estimated conditional on measures of the working environment within schools* (Simon and Johnson, 2015; Sims, 2017). Because the relationship between pupil demographics and retention is largely spurious, conditioning on pupil demographics would therefore strip out the effect of poor working conditions form our estimates, which would be undesirable for our purposes. Data on working conditions is not

---

[2] Note that any limits can be shown depending on the balance of false-negative and false-positive misclassification desired. Shewhart maintained that this should be calibrated empirically. For example, in the case of teachers, this could be done by monitoring the attrition rate of NQTs across the system.

available for all schools in England. We therefore believe that our use of funnel plots is the optimal method for our purposes.

Figure 4 is in fact the funnel plot for use of NQTs by schools in North-West England. The dashed line shows the mean level of NQT use across secondary schools in the North West. Most schools lie within the 95 percent limits, thus any variation between them is hard to distinguish from random variation. Rankings of schools within these limits would therefore be dubious. However, the chart also shows that 60 of 458 schools appear to have systematically high use of NQTs over five years. This is much higher than the 11 (2.5% of) schools we would expect to see above the control limit line if a purely random allocation of NQTs took place. Therefore, we conclude that some schools are systematically using unusually high numbers of NQTs over this period.

Figure 4: Proportion of teachers who are in NQT year at a school 2010-14 (secondary schools in the North West region)
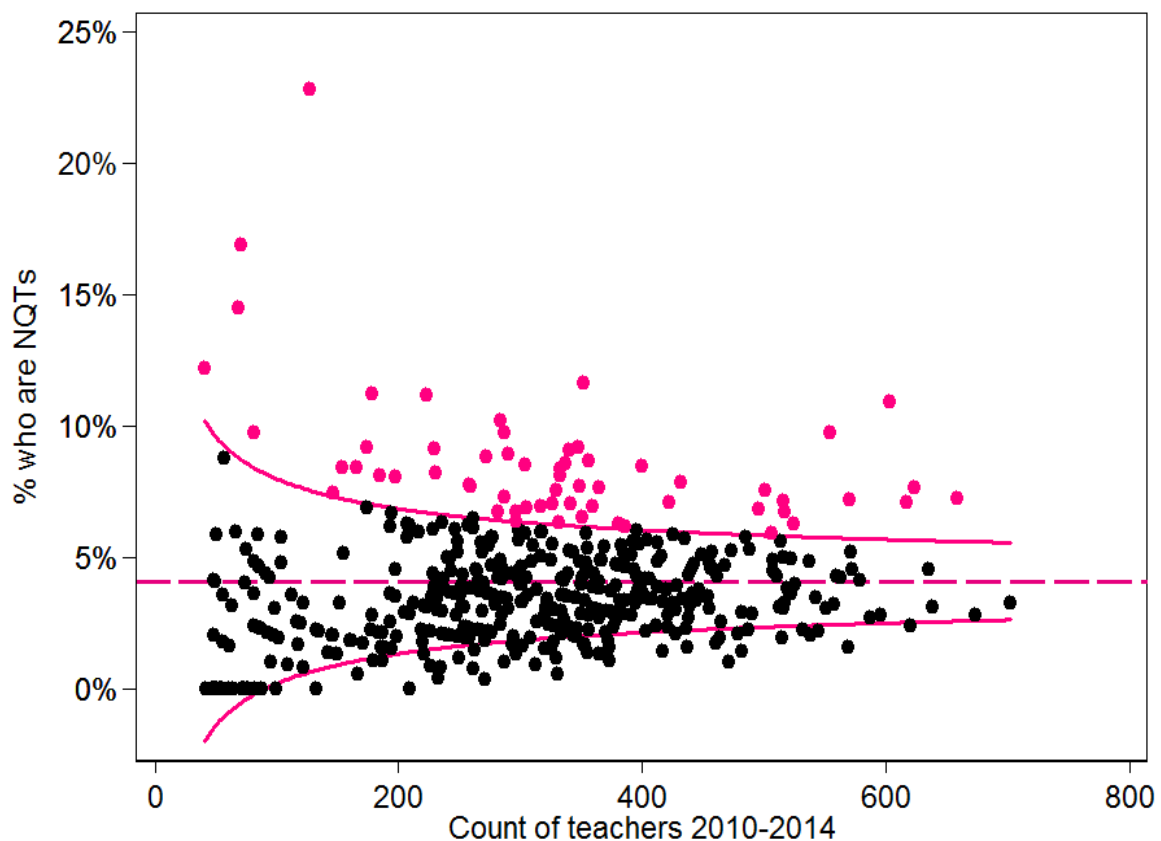
We can draw similar funnel plots to see whether secondary schools in the North West also have unusually high loss of their NQTs at the end of this first post-qualification teaching year. Figure 5 shows that just 22 of the 458 secondary schools in the North West have loss of NQTs that is more

than we would expect through chance alone, given a common cause process. Under a random process we would expect this figure to be 11 schools. We have been careful to check that this NQT attrition is not largely attributable to new teachers rotating jobs across a multi-academy trust (it is not) or resulting from the end of Teach First placements when attrition is very high (few of the schools in the table below participate in Teach First).

Figure 5: Proportion of teachers in NQT year who leave a school 2010-14 (secondary schools in the North West region)
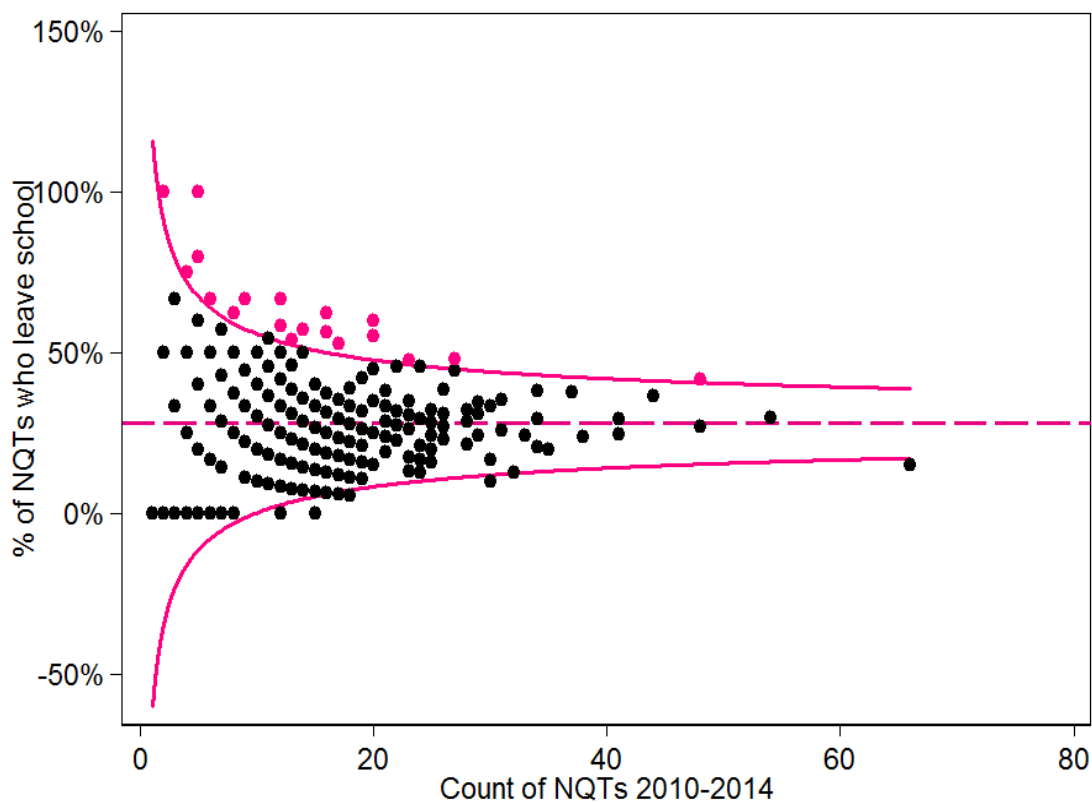
We repeated the analysis shown above for every region, separately for primary and secondary schools. By running these 18 separate analyses, we are implicitly assuming that all schools within the same phase and region are subject to similar common cause teacher labour market processes. Combining the results from these charts, we can count the number of schools experiencing both high use and high loss of NQTs over a five-year period. We count loss of NQTs at the end of the academic year, both to other schools or to the profession as a whole. As shown in Table 1, we found fewer than 200 schools across the country who met these criteria. This figure sounds quite low, given how many schools there are in the sample, but it is about 10 times higher than we would expect through a chance process of NQT allocation and a chance process of NQT loss combined. A

chi-squared test of association indicates there is a significant relationship between schools that use a high number of NQTs and those that lose a high number of NQTs.

Table 1: Number of schools exhibiting unusually high use and loss of NQTs over five-year period, 2010-2014

		Use of NQTs	
		Not unusual	**Unusual**
Loss of NQTs	**Not unusual**	13642	1292
	Unusual loss to school but not profession	368	77
	Unusual loss to profession	786	122

Note: Only includes schools for which we have data for all years

Figure 6 shows the location of these schools across the country. The even distribution across regions is partly the result of our choice to run the analysis separately for each region in turn. As it happens, the North East has the greatest proportion and the South West has the lowest proportion of the high-use-high-loss schools. This does not mean that the North East's overall statistics are poor; rather that there is significant (special-cause) variation across schools within the region.

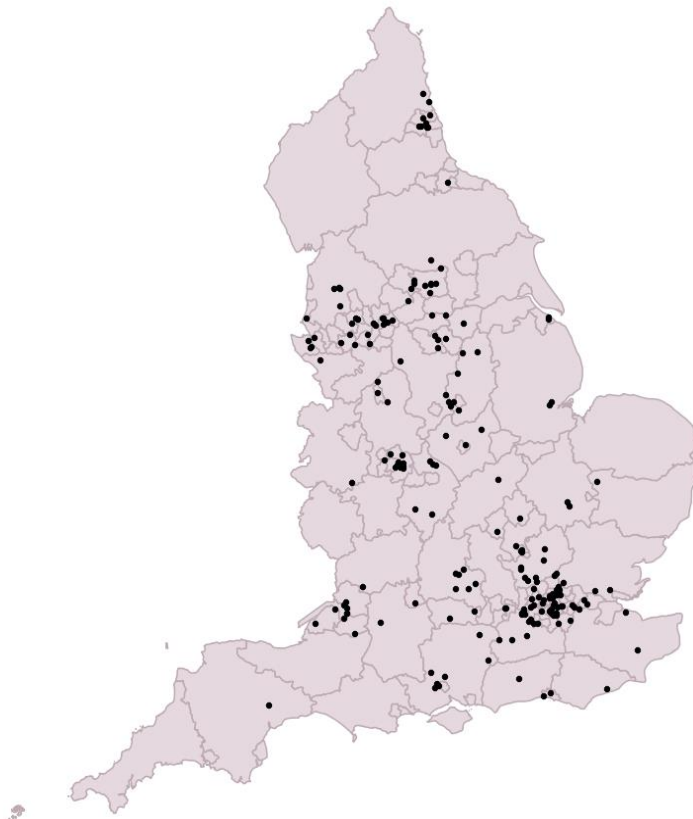*Figure 6: Location of schools that overuse NQTs, using pooled 2010-14 data*

Table 2 compares the characteristics of high-use-high-loss schools with other schools, in order to look for clues about what might be causing this behaviour. We might expect high-use-high-loss

schools to be characterised by poor inspection ratings, which might demoralise teaching staff or lead to high levels of voluntary teacher turnover, for example. Or we might expect high-use-high-loss schools to have very disadvantaged intakes, which might increase the challenges involved in teaching there. Table 2 shows that high-use-high-loss schools have: teachers with fewer years of experience in teaching, teachers who have been in their current job for longer, and a greater proportion of teachers who are NQTs. This is what we would expect given the way we have categorised these schools. High-use-high-loss schools also have a slightly higher proportion of pupils on free school meals (FSM), which is a standard proxy for disadvantage, and a slightly lower proportion of schools with Good or Outstanding ratings from Ofsted, the national schools inspectorate.

Table 2: Comparing the characteristics of schools that do, and do not, overuse NQTs

		High-use-high-loss schools	Other Schools	Standardised Difference
Pupils on Free School Meals (%)	Mean	17.82	15.85	0.172
	SD	11.08	12.42	
Number of Pupils	Mean	17.82	15.85	0.221
	SD	11.08	12.42	
Female Teachers (%)	Mean	0.82	0.82	0.030
	SD	0.13	0.13	
Teacher Experience (years)	Mean	9.70	12.89	0.994
	SD	3.05	3.37	
Years at current school	Mean	5.21	6.85	0.677
	SD	2.22	2.60	
Qualified Teachers (%)	Mean	0.96	0.97	0.138
	SD	0.06	0.06	
Ofsted: Outstanding	%	18.97	20.29	0.033
Good	%	63.79	66.96	0.067
Requires Improvement	%	14.66	11.44	0.095
Inadequate	%	2.59	1.31	0.093

Having said this, arguably the most striking thing about this table is how small the difference in pupil disadvantage and Ofsted grades are. The standardised difference between the %FSM pupils in high-use, high-loss schools and other schools is 0.172. This relationship is unlikely to account for high-use, high-loss behaviour since, even looking at the unconditional association, a two standard deviation increase in %FSM pupils is associated with just a 0.2 percentage point increase in the probability of being a high-use-high-loss school. The standardised differences between the Ofsted grades are all below 0.1 and, looking at the unconditional association, moving from being rated Outstanding to being rated Inadequate is associated with just a 0.4 percentage point increase in the probability of

high-use-high-loss school. These findings are consistent with the results from papers which model teacher retention directly using survey data on working conditions linked with administrative data on turnover and attrition. Research using data on English schools finds that, conditional on the quality of working conditions in schools and teacher and pupil characteristics, Ofsted grades and pupil deprivation are not related to teacher job satisfaction or desire to move school (Sims, 2017). Six other studies using data from the US come to similar conclusions (Simon and Johnson, 2015). In summary, we interpret this evidence as showing that very high levels of NQT use and loss cannot be explained by schools being under pressure from either the accountability system or the demands of their specific pupil intake. It is instead far more likely that these schools have very poor working conditions which leads to poor retention (Kraft *et al.,* 2016). In analysis not shown here (available from the authors on request) we show that turnover of non-NQT early career teachers and new-to-school teachers is also higher in high-use-high-loss schools, which is also consistent with our interpretation.

4. The costs of high use and high loss of newly qualified teachers

High use and high loss of NQTs has important opportunity costs in terms of increasing teacher shortages and the need to spend additional money on training replacements. In this section we aim to quantify these costs.

Schools that overuse NQTs exacerbate staffing shortages by pushing teachers out of the profession. In times of sufficient teacher supply, this might not be a problem since it could be the least effective teachers that leave the profession. However, at a time of teacher shortages, loss from the profession is unambiguously costly. We identify 122 schools (0.5% of the number of schools in 2014) that both use and lose an unusually high proportion of NQTs from the profession between 2010 and 2014. In total they lose 577 NQTs over this period, which is 33% of the NQTs they employ. If they had lost NQTs at a rate more typical of other schools (11.5%) then they would have retained an additional 376 teachers who could then have progressed beyond their NQT year.[3] By comparing this to overall teacher shortages in 2014/15, which we calculate from the Initial Teacher Training shortfall was

---

[3] Another option for specifying the counterfactual would have been to estimate the additional teachers retained if all the 122 schools came just within the control limits. However in our framework, all variation within the control limits is assumed to represent common cause variation. Because common cause variation can be thought of as random draws from a stable distribution, we are justified in assuming that these schools would be distributed evenly around the average, making this the relevant counterfactual.

1708, we can see that the excess loss of NQTs from the profession accounted for by special-cause variation in these high use/loss schools is equivalent to around 22% of the current teacher shortage.

Even without teacher shortages, reducing this NQT loss is valuable because it means we can reduce the number of new teachers that need to be trained to sustain a given level of supply. The National Audit Office (2016) estimates that England spent £700m per year training teachers in 2014 which yielded 23,900 qualified teachers. On this basis, each new NQT cost £29,289 to train, a figure consistent with the training cost range quoted in Allen *et al.* (2016). This suggests that the public cost (excluding personal trainee costs) of overuse/loss NQTs in these schools between 2010 and 2014 amounts to around £11m.

5. Implications for policy

Given that teachers are in short supply in England, we might expect that teachers would be able to avoid schools in which working conditions are poor, or which have very poor retention rates. However, we observe 122 schools in our data that were able to persist in recruiting and losing high numbers of NQTs across a five year period. This suggests that NQTs do not have access to the information necessary to identify and avoid such schools. In practice, NQTs in England do not have access to data on school turnover rates and working conditions data is only available for a small sample of anonymised schools. Although newly qualified teachers could attempt to glean this information through visiting schools and talking to existing staff, qualitative research following teachers through the early stages of their careers suggests they often do not do this (Johnson and Birkeland, 2003b). That early career teachers do display preferences for good working conditions when choosing their second jobs (Kukler-Acavedo, 2009; Boyd *et al.,* 2010) is also consistent with the interpretation that NQTs suffer from a lack of information when choosing their first school.

Publishing information on schools turnover rates would be one way of tackling this information asymmetry, but doing so would require careful consideration of the implications and could potentially be misleading, as discussed earlier in Section 3. However, funnel plots can provide simple and sufficient information for trainee teachers by listing schools that are outside control limits using readily available, public data.[4] The information they provide is therefore more reliable than that of league tables because it is designed to discount random variation resulting from common-causes. The information they provide is also reliable in the sense that it would be very hard for schools to behave strategically in order to game these measures. Using information on observed teacher

---

[4] In our framework, being close to the control limits is not informative, since it only indicates that we are unable to rule out common cause variation explaining all of the deviation from average rates of use or loss.

movements is incentive compatible: teachers who like working in a school will tend to stay, while those who do not like working there will not stay. Schools would also have scope to remove genuinely incapable NQTs and remain within their control limits, as long as they did a good job of retaining other trainees. Surveying teachers on the quality of working conditions in schools, by contrast, may not be incentive compatible. For example, teachers may come under pressure from school leaders to provide positive responses, in order not to exacerbate staffing shortages and workload at the school in future years.

One potential concern with this policy is that this new accountability metric will only exacerbate staffing difficulties in struggling schools (see Clotfelter *et al.*, 2004). Indeed, for the policy to work, classifying a school as being beyond the control limits needs to potentially deter early-career teachers from working there, since this is the mechanism by which it incentivises school leaders to improve working conditions. In order to mitigate the risk that the policy contributes to staffing difficulties in schools, we propose that two safeguards are introduced. First, funnel plots should be based on at least two years of pooled data, so that if a school begins to struggle with working conditions and retention in a single year, it can rely on data from the previous year to keep it within control limits while it addresses these problems. Second, in the same way that schools which have been rebrokered (reopened under new management) are allowed a three year "grace period" before their next Ofsted inspection, we propose that any schools which are rebrokered have their NQT retention rates wiped clean and their first year of data ignored, meaning that it will be at least three years before they accumulate enough pooled data to be declared inside or outside of their control limits. Schools should also be supported to improve working conditions through evidence-based leadership programmes during this period (see, for example: Knibbs *et al*., 2017).

6. Conclusion

Teacher shortages are a widespread challenge in public school systems and declining retention is an important contributor to these shortages. Recent research has gone beyond looking at cross-sectional measures of retention to investigate schools that display chronic staffing instability (Holme *et al.,* 2017; Allensworth *et al.,* 2009). We have contributed to this literature by showing how funnel plots can be used to identify schools which both use and lose an unusually high number of NQTs. Using these methods, we have identified 122 schools which lose NQTs at a rate three times the national average and between them are responsible for losing 577 NQTs from the profession over a five year period. As well as the significant personal and emotional costs this imposes on teachers

(Johnson and Birkeland, 2003a; Hong, 2012; Buchanan, 2010), these schools exacerbate system-wide teacher shortages and represent wasted public investment on teacher training.

We argue that funnel plots provide a simple, reliable and low cost way to highlight unsustainable use/loss of NQTs. In particular, we argue that this is superior to publishing league tables on the grounds that they are designed to discount random, common-cause variation across institutions. Publishing lists of schools which are inside or outside of control limits would help improve NQTs access to information about the quality of working conditions in schools and in doing so help them chose schools in which they are more likely to have a sustainable career in teaching. Further research would be needed in order to monitor and evaluate how teachers respond to the new information and whether or not it improves retention of NQTs. Monitoring changes in the retention of NQTs across cohorts would also allow the control limits on the funnel plots to be calibrated empirically, in line with Shewhart's (1931) recommended approach.

REFERENCES

Allensworth, E., Ponisciak, S. and Mazzeo, C. (2009), *The schools teachers leave: Teacher mobility in Chicago Public Schools*, Chicago, IL: University of Chicago Consortium on Chicago School Research.

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

Knibbs, S., Mollidor, C., Lindley, L., Allen, R. and Sims, S. (2017), *High Potential Middle Leaders (Secondary) programme: an evaluation*. London: National College for Teaching & Leadership.

Allen, R., Burgess, S. and Mayo, J. (2017), 'The teacher labour market, teacher turnover and disadvantaged schools: new evidence for England', *Education Economics*, pp. 1-20.

Allen, R. and Sims, S. (2018), 'Do Pupils From Low-Income Families Get Low-Quality Teachers? Indirect Evidence from English Schools', Forthcoming in Oxford Review of Education.

Allen, R. and Sims, S. (forthcoming), *The Teacher Gap,* London: Routledge.

Atteberry, A., Loeb, S. and Wyckoff, J. (2016), 'Teacher Churning: Reassignment Rates and Implications for Student Achievement', *Educational Evaluation and Policy Analysis*, 20(10), pp. 1–28.

Bevan, G. and Wilson, D. (2013), 'Does "naming and shaming" work for schools and hospitals? Lessons from natural experiments following devolution in England and Wales', *Public Money & Management*, 33(4), pp. 245–252.

Boyd, D., Grossman, P., Ing, M., Lankford, H., Loeb, S. and Wyckoff, J. (2010), 'The Influence of School Administrators on Teacher Retention Decisions', *American Educational Research Journal,* 48(2), pp. 303–333.

Buchanan, J. (2010), 'May I be excused? Why teachers leave the profession', *Asia Pacific Journal of Education*, 30(2), pp. 199–211.

Burgess, S. (2015), *Human Capital and Education: The State of the Art in the Economics of Education*, COEURE: Cooperation on European Research in Economics.

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

Clotfelter, C., Ladd, H., Vigdor, J. and Diaz, R. (2004), 'Do school accountability systems make it more difficult for low-performing schools to attract and retain high quality teachers?', *Journal of Policy Analysis and Management*, 23(2), pp. 251–271.

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

Dolton, P. and van der Klaauw, W. (1999), 'The turnover of teachers: a competing risks explanation', *Review of Economics and Statistics*, 81(3), pp. 543-552.

Dolton, P., Tremayne, A. and Chung, T. P. (2003), '*The economic cycle and teacher supply',* Paris: OECD.

Falch, T., Johansen, K. and Strøm, B. (2009), 'Teacher shortages and the business cycle', *Labour Economics*, *16*(6), pp. 648-658.

Gini, R. and Forni, S. (2009), 'Funnel plots for institutional comparisons', United Kingdom Stata Users' Group Meetings 2009 10, Stata Users Group.

Hamre, B. and 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), pp. 949-967.

Hanushek, E. and Rivkin, S. (2012), 'The Distribution of Teacher Quality and Implications for Policy', *Annual Review of Economics*, 4, pp. 131-157.

Henry, G., Fortner, C. and Bastian, K. (2012), *'*The Effects of Experience and Attrition for Novice High-School Science and Mathematics Teachers*', Science*, *335*(6072), pp. 1118–1121.

Hobson, A. J., Malderez, A., Tracey, L., Homer, M. S., Ashby, P., Mitchell, N., ... and Tomlinson, P. D. (2009), Becoming a Teacher: Teachers' experiences of initial teacher training, Induction and early professional development (Final report). *Nottingham: Department for Children, Schools and Families (DCSF)*.

Holme, J., Jabbar, H., Germain, E. and Dinning, J. (2017), 'Rethinking Teacher Turnover: longitudinal measures of instability in schools', *Educational Researcher*, 0013189X17735813.

Hong, J. (2012), 'Why do some beginning teachers leave the school, and others stay? Understanding teacher resilience through psychological lenses', *Teachers and Teaching*, 18(4), pp. 417-440.

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).

Johnson, S. and Birkeland, S. (2003a), 'Pursuing a "sense of success": New teachers explain their career decisions', *American Educational Research Journal*, 40, pp. 581-617.

Johnson, S. and Birkeland, S. (2003b), 'The Schools That Teachers Choose'. *Educational Leadership*, 60(8), pp. 20-24.

Kraft, M. (2017), 'Teacher Effects on Complex Cognitive Skills and Social-Emotional Competencies', *Journal of Human Resources*, 0916-8265R3.

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

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.

MAC (2016), *Partial review of the Shortage Occupation List*, Migration Advisory Committee.

National Audit Office (2016), *Training New Teachers,* London: NAO.

OECD (2014), *TALIS 2013 Results: An International Perspective on Teaching and Learning*, Paris: OECD.

Oulton, J.A. (2006) The Global Nursing Shortage: An Overview of Issues and Actions, *Policy, Politics and Nursing Practice,* 7(3) pp. 34S-39S.

Mohammed, M. A., Cheng, K. K., Rouse, A. and Marshall, T. (2001), 'Bristol, Shipman, and clinical governance: Shewhart's forgotten lessons', *The Lancet*, 357(9254), pp. 463-467.

Papay, J. and Kraft, M. (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, pp. 105–119.

Ronfeldt, M., Loeb, S. and Wyckoff, J. (2012), 'How Teacher Turnover Harms Student Achievement', *American Educational Research Journal*, 50(1), pp. 4–36.

Shewhart, W. (1925), 'The application of statistics as an aid in maintaining quality of a manufactured product', *Journal of the American Statistical Association*, 20(152), pp. 546-548.

Shewhart, W. (1931), *Economic control of quality of manufactured product*. ASQ Quality Press.

Sims, S. (2018), *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.

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

Sims, S. (2017), *TALIS 2013: Working Conditions, Teacher Job Satisfaction and Retention*, Department for Education Statistical Working Paper.

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

Spiegelhalter, D. (2005), 'Funnel plots for comparing institutional performance', *Statistics in Medicine,* 24, pp. 1185–1202.

Wiliam, D. (2016), *Leadership for teacher learning: Creating a culture where all teachers improve so that all students succeed*, Learning Sciences International.

Wiswall, M. (2013), 'The dynamics of teacher quality*', Journal of Public Economics*, *100*, pp. 61-78.