Institute of Education

High-Stakes Accountability and Teacher Turnover: how do different school inspection judgements affect teachers' decisions to leave their school?

Sam Sims

Department of Quantitative Social Science

Working Paper No. 16-14

October 2016

Disclaimer

Any opinions expressed here are those of the author(s) and not those of the UCL Institute of Education. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

DoQSS Workings Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Department of Quantitative Social Science, UCL Institute of Education, University College London,

20 Bedford Way, London WC1H 0AL, UK

High-Stakes Accountability and Teacher Turnover: how do different school inspection judgements affect teachers' decisions to leave their school?

Sam Sims[1]

Abstract

High teacher turnover damages pupil attainment (Borg *et al.*, 2012; Ronfeldt *et al*., 2012). But while the effects of pupil and teacher characteristics on turnover are well documented, relatively little attention has been paid to the impact of the accountability system. This paper is the first to evaluate the effect on turnover of schools receiving different judgements from the English national schools inspectorate, Ofsted. Theoretically, the effects of inspection judgements are ambiguous. An 'Inadequate' rating may harm teachers' self-efficacy, increasing the chance of them leaving their current school. On the other hand, an 'Inadequate' rating provides a negative signal about the quality of teachers working in that school, decreasing the chance of them finding employment elsewhere. I use a difference in difference approach to estimate this empirically and find that an 'Inadequate' rating leads to an increase in turnover of 3.4 percentage points. By contrast, schools receiving an 'Outstanding' rating see no change in turnover. The results are robust to a number of specifications, sample restrictions and a placebo test.

JEL codes: I21, J44, J63, D82

Keywords: Teacher turnover, high-stakes accountability, school inspection, efficacy, signalling

Contact Details: Sam Sims (s.sims@ucl.ac.uk), UCL Institute of Education

Acknowledgements: I would like to thank John Jerrim, Lorraine Dearden, Gill Wyness and Bilal Nasim for the guidance and feedback they have provided on this paper.

[1] Department of Social Science, Institute of Education, University College London

1. Introduction

Policymakers in the US and UK have expressed concerns about the effect of high teacher turnover on pupil attainment (House of Commons Education Committee, 2012; House of Representatives Committee on Education and the Workforce, 2012). Indeed, some US scholars have gone as far as to claim that "teacher turnover may be the most significant problem facing K12 schools" (Richardson *et al*., 2008, p10). While teachers have similar levels of turnover to 'comparable' professions such as nursing (Harris & Adams, 2007), these averages hide wide variation between schools, with turnover tending to be concentrated in schools with high levels of deprivation and high levels of ethnic minority students (Allen *et al*., 2012; Borg *et al*., 2012; Ingersoll and May, 2012).

High teacher turnover in individual schools might have detrimental effects on pupils for several reasons (Brown & Wynn, 2009; Ronfeldt *et al*., 2012) including: lack of continuity or coordination between teachers; loss of hard-to-acquire knowledge about specific pupils; opportunity costs associated with recruiting and inducting replacements; and the potential for lower quality replacement teachers. Recently, empirical studies have shown that these concerns are justified, showing negative impacts of high turnover on pupil attainment (Borg *et al*., 2012; Ronfeldt *et al.,* 2012). However, it is also important to stress that turnover is not always detrimental: successful school turnaround programmes are often characterised by high initial teacher turnover (Dee, 2012).

Concerns about the effects of high turnover in some schools have led researchers to investigate the determinants of this phenomenon in some detail. Teachers are more likely to resign if they are young (Allen *et al.*, 2012), inexperienced (Keigher & Cross,

2010), highly qualified (Boyd *et al.*, 2005), less effective (Boyd *et al.*, 2008), have higher earnings potential outside of teaching (Guarino *et al.*, 2006), or are reaching retirement age (NFER, 2008). With respect to pupil characteristics, turnover tends to be higher if there are higher proportions of minority, special educational needs or deprived pupils in a school (Falch & Strom, 2005; Allen *et al.*, 2012). Working environment also has an effect, with turnover higher in schools in which other staff, including school leadership, are less supportive (Guarino *et al.*, 2006; Brown & Wynn, 2009; Simon & Johnson, 2013).

But while these demographic and school factors have been studied extensively, there is less evidence on the effect of the accountability system. Clotfelter *et al.* (2004) study the introduction of a high-stakes accountability system in North Carolina in which schools are given one of four grades and find that being labelled low performing increases teacher turnover. Boyd *et al.* (2008) study the introduction of a high stakes public examination for fourth grade students in New York State. They find that this decreases fourth grade teacher turnover but conclude that this reflects reallocation of more experienced (and therefore less likely to resign) teachers to that year group. Feng *et al.* (2010) study the school accountability system in Florida, which rates all schools from A to F. They exploit a reform of the ranking system which exogenously reallocated just over half of all schools to a different grade overnight. They employ a difference in difference (DD) estimator and find that a school being reallocated to a lower grade increases turnover, while being allocated to a higher grade reduces turnover. This study benefits from the strong exogeneity of the intervention but interpretation of the findings is hampered by the fact that reallocation occurred contemporaneously with the introduction of the wider A+ Reform Programme. Finally, Dizon-Ross (2014) uses a regression discontinuity

design to study a similar scoring system for schools in New York City and finds, contrary to the other published studies, that receiving a low grade decreases turnover by 3%.

This paper is the first to evaluate the effects of schools inspection judgements on turnover in English schools. Ofsted inspection judgements are similar to the policy framework studied by Clotfelter *et al.* (2004) in that all schools are periodically given one of four judgements, in this case: Outstanding, Good, Requires Improvement, or Inadequate. I estimate the effect on turnover of two different treatments: moving from a Good to an Outstanding rating and moving from a Requires Improvement to an Inadequate rating. Theoretically, the effect of receiving these judgements is ambiguous. On the one hand, receiving an 'Outstanding' rating will enhance teacher's sense of efficacy and this has been shown to increase commitment and retention (Meyer *et al.*, 2002; McNatt & Judge, 2008; Klassen & Chiu, 2011). On the other hand, an 'Outstanding' rating provides a valuable labour market signal about the quality of teachers in a school (Staiger & Rockoff, 2010) and this has been shown to increase the chances of leaving a job (Schwab, 1991; Trevor *et al.*, 1997). The converse will be true of receiving an Inadequate rating.

In order to estimate this empirically, I assemble a matched, balanced panel of English primary and secondary schools between 2010 and 2013. My preferred specification employs a DD estimator and incorporates a rich set of control variables including teacher and pupil demographics and school fixed effects. I find that schools being reclassified from 'Good' to 'Outstanding' do not see any change in turnover. However, schools being reclassified from 'Requires Improvement' to 'Inadequate' see an in turnover of 3.4 percentage points, statistically significant at the 1% level. This is equivalent to 25% of average turnover across all school-year observations. I

also show that one third of the teachers that leave schools reclassified from 'Requires Improvement' to 'Inadequate' go on to teach in other schools. Moreover, a third of these 'movers' end up in relatively ineffective schools, where they are least likely to improve their teaching (Supovitz *et al.*, 2009; Jackson & Bruegmann, 2009; Ronfeldt, 2012).

The rest of the paper is structured as follows. Section 2 of the paper describes the school accountability system in England and provides details of Ofsted inspection process, section 3 describes and justifies my empirical approach, section 4 describes the data, section 5 presents my empirical results, section 6 discusses potential mechanisms linking inspection judgements and teacher turnover, section 7 concludes.

2. School Inspections and Accountability in England

The OECD defines school inspection as a "mandated, formal process of external evaluation" which "involves one or more trained inspectors who evaluate quality based on a standard procedure" and which "aims to hold schools accountable" (OECD, 2015, p479). Twenty eight of thirty six OECD nations require such inspections, although there is wide variation in how the inspections are structured and how the information is used (OECD, 2015).

England's inspectorate, the Office for Standards in Education, Children' Services and Skills (Ofsted), was established as an independent organisation in 1992. Schools are given no more than three days notice before an inspection occurs and inspectors spend around two days in the school collecting information. This results in an overall judgement of either 'Outstanding', 'Good', 'Requires Improvement' or 'Inadequate', along with suggestions for how the school can improve. All schools are inspected at

least once every five years but schools given a Requires Improvement judgement are inspected within three years and schools given an inadequate judgement are monitored on an ongoing basis (Ofsted, 2010). Results of the inspection are made publicly available through Ofsted's website. School inspections are high-stakes and an Inadequate judgement can lead to job losses (Jones & Tymms, 2014; Eyles & Machin, 2015).

3. Empirical Approach

Estimating the effect of different Ofsted judgements on teacher turnover is challenging for a number of reasons. A simple cross-section regression (model 1) of teacher turnover for a particular school (Y_s) on a treatment dummy for that school (D_s) and a range of school characteristics (X_s) such as the proportion of pupils at the school with special educational needs, is subject to a number of threats.

(1) $Y_s = \beta_0 + \beta_1 D_s + \beta_2 X_s + u_s$

First, the direction of causality will be unclear, since high turnover could be driving school failure, rather than an 'Inadequate' judgement driving turnover. Second, previous studies have linked a range of variables to teacher turnover which are hard to measure validly and reliably, such as collegiality of the staff (Guarino *et al.,* 2006) and the quality of school leadership (Brown & Wynn, 2009). Attempting to control for these variables explicitly may result in measurement error, which would also bias the estimates. Third, given that the recent literature has identified a series of new and unexpected determinants of turnover (Simon & Johnson, 2013), it is hard to rule out further research identifying yet more relevant variables. This highlights the risk of

omitted variable bias. Fourth, wider macro-economic trends or changes in the policy environment may be driving differences in turnover across time (Hutchings, 2011).

My empirical approach is designed to help overcome these problems. I employ nearest neighbour matching prior to analysis to ensure that there is common support across treatment and control group schools. I then make two key changes to the cross-section model. First I exploit the longitudinal nature of my data to incorporate school fixed effects. This allows me to control for all unobserved, time-invariant, school-level factors. I argue that this captures difficult to measure or unmeasured variables such as support structures and quality of the leadership, which are unlikely to change within the short (one year) window of observation. This is shown in model (2) through the incorporation of a school fixed effect (Z_s) and the subscript t being added to the Y, D, X and u variables to show that each of them is now being measured for a particular year.

(2) $Y_{st} = \beta_0 + \beta_1 D_{st} + \beta_2 X_{st} + \beta_3 Z_s + u_{st}$

Second, I adopt a DD estimator which allows me to be resolve the temporal ordering of cause and effect and to control for any time-varying unobserved factors which are common across schools. I argue that this captures the effects of wider macro-economic trends or changes in the national education policy framework. This is shown in model (3) through the inclusion of a vector of dummies for all but one of the years between 2010 and 2013 (W_{st}) and an interaction between the treatment variable (D_s) and a dummy (T_{st}) which is 'on' in the year after an inspection.

3) $Y_{st} = \beta_0 + \beta_1(D_s * T_{st}) + \beta_2 X_{st} + \beta_3 Z_s + \beta_4 W_{st} + u_{st}$

The key assumption required to give β_1 a causal interpretation is that there are common trends between treatment and control groups. There are several theoretical reasons to think this assumptions is justified in this case. State school teachers operate in the same labour market, meaning they are unlikely to experience different effects from macroeconomic fluctuations, and they all face a common, nationally-determined policy framework, meaning policy changes are unlikely to affect them differently. Ruling out alternative explanations is also aided (Steiner *et al.*, 2009) by the fact that the intervention occurs at a known point in time (I observe the date of inspection judgement), is implemented with immediate effect (the judgement is publicly announced within fifteen days of inspection), and because there is a strong theoretical basis for determining expected time lags in the response of the dependent variable (most teachers that move school do so at the end of the academic year). This allows me to look for changes in turnover within a tightly defined period of one year or less after the inspection takes place, reducing the chances of contemporaneous but unrelated factors driving any observed differences in turnover.

I am also able to provide empirical evidence to support my argument that β_1 can be given a causal interpretation. It could be argued that poorly performing schools are both more likely to be judged Inadequate and more likely to be restructured as part of a policy initiative to turn the school around. School quality might therefore be the reason for both an Inadequate judgement and high turnover, rather than the inspection judgement affecting turnover independently. However in section 5 I restrict my sample to schools that have not been restructured and show that the

main results do not change. Similarly, it is plausible that some of the schools labelled inadequate see a change in pay policy after receiving the judgement and this is independently driving the change in turnover. However the same sample restriction also rules out this interpretation, since by excluding all schools that become academies I am left only with schools who have to follow the national pay scales. In section 5 I also conduct a placebo test which shows that treatment and control schools do move along common trends in the year immediately prior to treatment.

In summary, my empirical approach has several strengths. First, it allows me to control for unobserved time-invariant characteristics and unobserved time-varying characteristics which are common across schools. Second, by taking into account all teachers in the school, I am able to avoid problems resulting from the strategic reallocation of teachers across grades that occurs in Boyd *et al.* (2008). Third, because the intervention does not occur as part of a wider package of reforms I am better able to rule out alternative explanations for my observed results, which is an issue for Feng *et al.* (2010). Fourth, although I cannot rule out the presence of unobserved, time-varying uncommon factors, I am able to provide both theoretical and empirical justification for a causal interpretation of my results.

4. Data

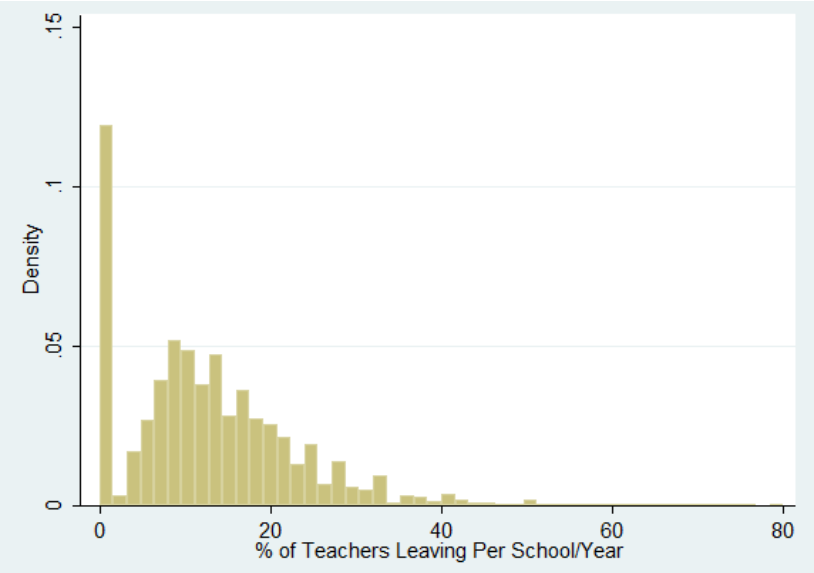
I draw on three datasets for this research. The School Workforce Census (SWC) contains individual level data on all staff in regular employment in English schools. From this dataset I take information on teacher's age, gender, ethnicity, contract type and time in role. I use unique school reference numbers to link this with a panel of Ofsted judgements and governance changes for all schools in England between 2006 and 2013. I also link to the National Pupil Database from which I draw a rich set of school-level educational and demographic variables including free school meal

status, area deprivation, and incidence of special educational needs, as well as indicators for region, primary/secondary and gender mix. I drop all years of data prior to 2010. This leaves me with 18,326 (primary and secondary) schools in my dataset.

I define turnover as the proportion of directly employed qualified teachers that exit a school in a given academic year. My data gives me two ways of measuring this. The first, which I call the exits measure, is defined by the proportion of teachers working in a school in year y-1 that are not still working in that school in year y. This is my preferred measure and is used in all models unless otherwise stated. However the exits measure is censored for 2010 because I do not have data on teachers for 2009. The second measure, which I call the arrivals measure, uses teacher's time in role to calculate the proportion of teachers who have joined a school since the beginning of the school year. This measure is potentially less valid, since schools may change the total number of teachers employed from one year to the next, meaning newly arrived teachers do not reflect other teachers leaving. However, the arrivals measure is not censored in 2010, giving me a larger sample. I therefore use the arrivals measure only when estimating models using subsamples of my data, in order to sustain sample size.

Figure 1 shows the distribution of turnover rates for each school-by-year, revealing a positive skew. A large number of observations have a turnover close to zero, half of the observations lie between 6% and 18% and around one-in-ten observations are above 25%. Table 1 shows that the mean turnover rate is 13.7% and this is fairly stable across years. When calculated using the arrivals measure this drops to 12.6%, suggesting this measure underestimates turnover.

Figure 1: Histogram of Turnover for Each School-Year



One option for defining my treatment variable is to look at the average effect across all upward movements and all downward movements. However this treatment is never experienced by a real world school and is therefore not of primary policy interest. There are four Ofsted grades, which means there are twelve possible moves, six up and six down. Estimating twelve different treatment effects would however make the analysis unwieldly, so I focus instead on the two extreme moves, which is where previous studies suggest the effects on turnover are strongest (Feng *et al.,* 2010). I therefore create two treatment dummies. The first takes the value 1 for any school which is rated 'Inadequate' having been rated 'Requires Improvement' in the prior period, and takes the value zero otherwise. In the rest of the paper I refer to this as being regraded Inadequate. The second treatment dummy takes the value 1 for any school which is rated 'Outstanding' having been rated 'Good' in the prior period and 0 otherwise. In the rest of the paper I refer to this as being regraded Outstanding. Table 1 shows how many schools are treated in each year. Across the four years for which I have data, 5% (944) of schools in my initial dataset are

regraded Outstanding at some point, while 2% (409) are regraded Inadequate. I also create a set of year dummies for each year between 2009 and 2013, and then drop the first of these to act as the reference case. Finally, I create a set of dummies that indicate the year in which a school was regraded to either Inadequate or Outstanding.

Table 1: Treatment and Turnover in Each Year

	2010	2011	2012	2013	Average
Turnover (exits measure)	na	13.3%	13.5%	14.4%	13.7%
Turnover (arrivals measure)	12.4%	11%	12.7%	14.3%	12.6%
					Pooled
Schools regraded Outstanding	311 (1.7%)	165 (0.90%)	193 (1.05%)	325 (1.77%)	994 (5.42%)
Schools regraded Inadequate	62 (0.34%)	153 (0.83%)	90 (0.49%)	104 (0.57%)	409 (2.23%)

I implement nearest neighbour matching using the psmatch2 programme in STATA (Leuven & Sianesi, 2014) and enforce common support. In order to generate valid propensity scores, it is necessary to condition on those variables which determine assignment to treatment. Inspection judgements are based on a mixture of attainment data, in-school observation and consultation with parents (Ofsted, 2010). I match on a number of characteristics that are likely to influence Ofsted rating: the proportion of pupils receiving free school meals, the proportion of pupils with special educational needs, the proportion of pupils with English as a second language, school stage and exam results measured as average point scores. This results in 2,893, or 16% observations being dropped from my dataset. The balance test from the matching process is reported in Table 2 and 3 in the Annex. It shows that the matching process achieves a substantial reduction in bias leaving no variables with a bias of more than +/- 2.5%. The t-tests show that for no variables can we reject the

null hypothesis that treatment and control groups are equal. In results not shown, I check my main findings for robustness to using kernel instead of nearest neighbour matching but find that this makes no notable difference to the number of observations dropped or the main results.

Table 4 in the Annex shows the number of missing observations for each variable. Most of the variables employed in my models have less than 3% of observations missing but two have much higher levels. In order to account for this missing data I use multiple imputation to create 20 imputations of my dataset using chained equations. Missing data for each of the variables listed in Table 4 is imputed using data for the same set of variables from another year. All results reported are therefore averages across these 20 imputed datasets, allowing me to sustain sample size while ensuring unbiased standard errors for my parameter estimates (Cheema, 2014).

5. Results

The results from my analysis are presented in Table 5. Model 1 uses nearest neighbour matching and controls for all observable characteristics. It shows that schools being regraded Outstanding see a 11 percentage point increase in turnover in the year following treatment, statistically significant at the 1% level. Schools being regraded Inadequate see a substantial increase in turnover of 23 percentage points, again significant at the 1% level.

Model 2 incorporates fixed effects. The results show that when schools are regraded Outstanding this leads to a very small (0.1 percentage point) reduction in turnover but this is not statistically significant, event at the 10% level. Schools being regraded Inadequate see a substantial increase in turnover of 3.4 percentage points,

significant at the 1% level. Controlling for unobserved, time-invariant factors therefore explains away the change in turnover for schools being regraded Outstanding but not for schools being regraded Inadequate. Model 3 further expands the range of factors controlled for by using a DD approach to control for time-varying common factors, making it my preferred specification. The results from model 3 show that schools being regraded Outstanding do not see a statistically significant change in turnover relative to treatment schools. Schools being regraded Inadequate again see a 3.4 percentage point increase in turnover relative to control group schools in the year following treatment, statistically significant at the 1% level. This is a materially important difference, equivalent to 25% of average turnover across all school-year observations.

The key assumption required to give β_1 in model 3 a causal interpretation is that there are common trends between treatment and control groups. This is necessary in order to isolate the treatment as the cause of the divergence in turnover rates between treatment and control schools. In order to test this I run a model that includes a placebo treatment introduced in the year before the genuine treatment. As the results in Table 6 show, the effect of the placebo treatment is statistically insignificant, even at the 10% level, both for schools being regraded Outstanding and for schools being regraded Inadequate. Turnover in treatment and control groups therefore moves along a common trend, at least in the period prior to treatment.

Table 5: Results

	Model 1	Model 2	Model 3
Schools regraded Outstanding	0.113*** (0.000915) N=15,122	-0.00122 (0.00518) N=15,241	0.00122 (0.00518) N=15,241
Schools regraded Inadequate	0.23*** (0.00119) N=15,371	0.0339*** (0.00897) N=15,371	0.0339*** (0.00897) N=15,371
Pupil and Teacher Characteristics	✓	✓	✓
School Fixed Effects		✓	✓
Diff-in-diff			✓

Notes: *** indicates significance at 1% level; ** at 5% level and * at 10% level.

Table 6 also includes a number of further robustness checks using restricted samples. I also show results using the arrivals measure of turnover in cases where no effect is found with the exits measure, in order to check that smaller sample size is not driving the result. I first restrict the sample by removing all schools which have experienced restructuring during the observation window. I use restructuring to refer to a number of processes, namely: opening, closing, splitting, merging with another school, becoming an academy and increasing or decreasing the number of forms of entry. All of these processes may have an influence on turnover by changing the staffing requirements of a school. This robustness check is also important because my fixed effect models treat schools that are converted to academies as being the same school before and after the conversion. In order to check whether these sorts of phenomena are driving my results I therefore run model 3 excluding all schools (just over 15% of my sample) which experienced any restructuring. The results for schools regraded Outstanding are unchanged, with the coefficient on the treatment

variable remaining statistically insignificant, even at the 10% level. Using my preferred exits measure of turnover the results for schools regraded Inadequate are also virtually unchanged.

I also split the sample into primary (81%) and secondary (19%) schools. The results remain unchanged for schools regraded Outstanding. For schools regraded Inadequate, the coefficient for primary schools remains essentially unchanged and the results remain statistically significant at the 5% level. The coefficient for secondary schools also remains unchanged but is no longer statistically significant. This loss of statistical significance is very likely due to the very low number of secondary schools (21) in my dataset that are regraded Inadequate.

Table 6: Robustness Checks

		Placebo Test	Excluding Restructured	Primary Schools	Secondary Schools
Schools regraded Outstanding	**Exits**	-0.00624 (0.0121) N=14,192	0.000457 (0.00875) N=12,999	0.00211 (0.00993) N=12,831	0.00314 (0.0194) N=2,482
	Arrivals	-0.00556 (0.00831) N=14,624	0.00355 (0.00728) N=13,000	0.00394 (0.00780) N=12,832	0.00833 (0.0155) N=2,482
Schools regraded Inadequate	**Exits**	0.00442 (0.00714) N=15,207	0.0348*** (0.0134) N=12,848	0.0342** (0.0133) N=12,656	0.0334 (0.0239) N=2,414
	Arrivals	-0.00281 (0.00566) N=15,294	Not Shown	Not shown	-0.032 (0.0246) N=2,414

Notes: *** indicates significance at 1% level; ** at 5% level and * at 10% level.

The Ofsted inspectorate aims to improve schools through inspection (Matthews & Sammons, 2004) and Hussain (2014) shows that schools regraded inadequate do indeed improve in the short run. However, by focusing on the effects of inspections on teacher mobility this research raises questions about the effect of inspections on

the wider school system. The overall effect of inspection on pupil attainment and equity will depend on 1) which teachers exit a school after it is regraded 2) whether those teachers move to another school (movers) or leave the profession entirely (leavers) and 3) which schools the movers go to. I now discuss each of these in turn, drawing on descriptive evidence from my dataset and findings from the US literature.

If the best teachers exit a school regraded Inadequate then this will likely cause the school to perform worse in the future. My dataset does not allow me to evaluate teacher quality since teachers cannot be linked to individual pupils and even comparison of departmental performance is hindered by sorting of pupils across subjects. Research from the USA, where teachers can be linked to pupils, shows that teacher effectiveness is robustly related to turnover, with less effective teachers more likely to leave their school (Hanushek *et al.*, 2005; Krieg, 2006; Feng & Sass, 2011; Goldhaber *et al.*, 2011). If the schools losing relatively ineffective teachers can recruit average quality teachers in their place then it follows that the average quality of teaching in the school will increase. This is consistent with the findings from Hussain (2014) that schools receiving an Inadequate rating improve their performance in the short run.

However, if these relatively ineffective teachers move to other schools then this may just result in a zero-sum reallocation of teachers across schools. Table 7 compares the move or leave decision of below-retirement-age teachers exiting schools regraded Inadequate with teachers exiting all other schools. It shows that just under a third (30%) of those exiting schools regraded Inadequate are moving on to other schools. Moreover, below-retirement-age teachers exiting schools regraded Inadequate are actually less likely to leave the profession than those exiting from other schools. This is consistent with evidence from the US (Feng & Sass, 2011;

West & Chingos, 2009) which shows that, while the least effective teachers are the most likely to exit schools, they are no more likely to leave the profession altogether.

Table 7: Destinations of Teachers that Exit their Schools

	Movers	Leavers
Schools not regraded Inadequate	23.58%	76.42%
Schools regraded Inadequate	30.47%	69.53%

Notes: estimated on full dataset, prior to matching and imputation

The third and final determinant of how these turnover patterns will affect pupils is the destination of movers. Teacher quality is not a fixed attribute and is affected by the school in which a teacher works (Kraft & Papay, 2014). More specifically, teacher effectiveness is influenced by the quality of school-specific teacher training (Harris & Sass, 2011) and the quality of other teachers in a school from whom they can learn (Jackson & Bruegmann, 2009; Supovitz *et al.*, 2009; Ronfeldt, 2012). If the movers go to more effective schools then the quality of their teaching will likely improve as a result. But if the movers go to ineffective schools then they will not benefit in this way. In order to investigate this I look at the Ofsted grades for the destination schools of teachers moving from schools regraded Inadequate. Table 8 shows that while just under two thirds (65%) of these teachers move to a Good or Outstanding school, the remaining third move to a Requires Improvement or Inadequate school. They are disproportionately likely (5%) to end up in Inadequate schools relative to the overall proportion of Inadequate schools (2%) and relative to teachers moving from schools regraded Outstanding (3%). Again, this is consistent with evidence from the US which shows that more effective teachers tend to move to more effective schools (Loeb *et al.*, 2012) and less effective teachers tend to move to less effective schools (Feng & Sass, 2011). 35% of teachers moving from schools

regraded Inadequate therefore end up in low performing schools in which they are less likely to improve their teaching. In summary, it appears that a group of relatively ineffective teachers are churning round the English school system.

Table 8: Distribution of Inspection Results for Destination Schools

	Inadequate	Requires Improvement	Good	Outstanding
All schools, 2010-2013	1.87%	30.48%	51.69%	15.96
Destination of movers from schools regraded Inadequate	4.52%	30.32%	50.97%	14.19%
Destination of movers from schools regraded Outstanding	2.78%	31.25%	41.67%	24.31%

Notes: estimated on full dataset, prior to matching and imputation

6. Discussion

Evidence of an underlying mechanism can strengthen a claim that the statistical relationship between a dependent and independent variable constitutes a causal relationship and clarify the external validity of a finding (Clarke *et al.,* 2014). In the introduction I suggested that an Inadequate Ofsted inspection might increase turnover by damaging teachers self-efficacy, but may also reduce turnover by sending a negative signal about teacher quality. In this section I discuss these potential mechanisms in more depth and use them to aid interpretation of my empirical results and the existing literature.

Economists have long stressed the difficulties faced by managers attempting to judge worker quality when recruiting outside their own firm (Lazear & Oyer, 2004). This theoretical account is also consistent with empirical evidence from teaching showing that, while school leaders can distinguish the very best and very worst

teachers in their own schools (Jacob & Lefgren, 2008), they are poor at judging the quality of 'outsider' teachers and even low-reliability signals are therefore very valuable to school leaders looking to improve the quality of the teachers they recruit (Staiger & Rockoff, 2010). Empirical research confirms that signals can increase turnover. Schwab (1991) shows that credible, externally visible signals of quality increase chances of turnover among university faculty members. Drawing on Schwab's reasoning, Trevor *et al.* (1997) show that promotions increase the chance of turnover by sending a positive signal about worker quality. An Inadequate rating may therefore make it harder to find employment elsewhere due to the negative signalling effect.

A second way in which inspection judgements might affect turnover is through self-perceptions. Self-Determination Theory (Deci & Ryan, 2008) provides a theoretical account of how people need to see themselves as competent and will attempt to change their working environment if it thwarts this basic need (Deci & Ryan, 2000). Teachers may therefore value being part of an Outstanding school because it boosts their self-efficacy. Empirical Studies have also confirmed this in a range of settings (Meyer *et al.*, 2002; Klassen & Chiu, 2011; McNatt & Judge, 2008). An Inadequate rating will therefore encourage teachers to exit a school in which they cannot see themselves becoming a good teacher.

This description of mechanism is useful for two reasons. First, it can explain why schools regraded Inadequate in my study see an increase in turnover while schools regraded Outstanding see no change. It is well known that English schools are suffering from teacher shortages, particularly in certain areas of the country and in certain subjects (Howarth, 2015; NAO, 2016). A shortage of teachers will not affect schools ability to take into account positive signals attached to teachers from schools

regraded Outstanding, meaning the signalling and efficacy effects both operate and, along with any other mechanisms in operation, appear to balance each other out. By contrast, the need to recruit a sufficient number of teachers may require schools to disregard negative signals attached to teachers from schools regraded Inadequate, meaning the signalling effect will lose its power and the efficacy effect, along with any other mechanisms operating, leads to an increase in turnover. This is consistent with my findings.

Second, the mechanism helps reconcile some apparently conflicting findings in the literature. My results for schools given the lowest grade are consistent with Feng *et al.* (2010), Clotfelter *et al.* (2004) and Dizon-Ross (2014). However, while I find that schools regraded Outstanding see no change in staff turnover, Feng *et al.* (2010) find that schools in Florida receiving the highest rating see a reduction in turnover. This difference can also be explained by reference to the mechanisms described. To see how, remember that the exogenous change in the accountability system that serves as the treatment in their study reallocated 51% of schools to a new grade overnight. This instability in the grading system will likely have reduced the value of the signal provided, which can explain why the efficacy effect, along with any other mechanisms operating, dominated the signalling effect in Florida.

7. Conclusion

This study is the first to evaluate the effect of different Ofsted inspection judgements on teacher turnover in English schools. I find that schools regraded Inadequate see a 3.4 percentage point increase in turnover, while schools regraded Outstanding see no change in turnover. I also find that one third of the teachers that leave schools regraded Inadequate move on to teach in other schools and a third of these 'movers' find jobs in relatively ineffective schools, where they are less likely to improve their

teaching. A group of ineffective teachers may therefore be churning round the English school system. I suggest that the ability of these teachers to find new jobs may be due to a shortage of teachers in England which forces schools to disregard the negative signal attached to teachers from schools regraded Inadequate. If this interpretation is correct, then school inspections, which are an important component of the accountability system in England, will have a muffled impact on the overall quality of teaching, as long as these shortages persist.

Bibliography

Allen, R., Burgess, S. & Mayo, J. (2012). *The teacher labour market, teacher turnover and* disadvantaged schools: new evidence for England. CMPO working paper No. 12/294.

Hussain, I. (2014). Subjective performance evaluation in the public sector: evidence from school inspections. *Journal of Human Resources*, *50*(1), 189–221.

Borg, J. R., Borg, M. O. & Stranahan, H. A. (2012). Closing the achievement gap between high-poverty schools and low-poverty schools. *Research in Business and Economics Journal*, *5*, 1-24.

Boyd, D., Lankford, H., Loeb, S. & Wyckoff, J. (2008). The Impact of Assessment and Accountability on Teacher Recruitment and Retention Are There Unintended Consequences? *Public Finance Review*, *36*(1), 88-111.

Boyd, D., Grossman, P., Lankford, H., Loeb, S. & Wyckoff, J. (2008). *Who leaves? Teacher attrition and student achievement.* NBER Working Paper No. 14022.

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

Brown, K. & Wynn, S. (2009). Finding, Supporting, and Keeping: The Role of the Principal in Teacher Retention Issues. *Leadership and Policy in Schools*, 8(1), 37-63.

Cheema, J. R. (2014). A Review of Missing Data Handling Methods in Education Research. *Review of Educational Research*, 84(4), 487–508.

Clarke, B., Gillies, D., Illari, P., Russo, F. & Williamson, J. (2014). Mechanisms and the Evidence Hierarchy. *Topoi,* 33(2), 339–360.

Clotfelter, C. T., Ladd, H. F., Vigdor, J. L. & Diaz, R. A. (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), 251-271.

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/Psychologie Canadienne*, 49(3), 182-185.

Dee, T. (2012). *School turnarounds: Evidence from the 2009 stimulus.* NBER Working Paper No. 17990.

Dizon-Ross, R. (2014). *How do school accountability reforms affect teachers? Evidence from New York City*. Working Paper.

Falch, T. & Strøm, B. (2005). Teacher turnover and non-pecuniary factors. *Economics of Education Review*, 24(6), 611-631.

Feng, L. & Sass, T. (2011). *Teacher quality and teacher mobility*. Andrew Young School of Policy Studies Research Paper Series 12-08.

Feng, L., Figlio, D. N. & Sass, T. (2010). *School accountability and teacher mobility.* NBER Working Paper No. 16070.

Goldhaber, D., Gross, B. & Player, D. (2011) Teacher career paths, teacher quality, and persistence in the classroom: Are public schools keeping their best? *Journal of Policy Analysis and Management,* 30(1), 57-87.
Guarino, C. M., Santibanez, L. & Daley, G. A. (2006). Teacher recruitment and retention: A review of the recent empirical literature. *Review of Educational Research*, 76(2), 173-208.
Hanushek, E. A., Kain, J. F., OBrien, D. M. & Rivkin, S. G. (2005). *The Market for Teacher Quality*. NBER Working Paper No. 11154.
Harris, D. N. & Adams, S. J. (2007). Understanding the level and causes of teacher turnover: A comparison with other professions. *Economics of Education Review*, 26(3), 325-337.
Harris, D. N. & Sass, T. R. (2011). Teacher training, teacher quality and student achievement." *Journal of Public Economics,* 95(7), 798-812.
House of Commons Education Committee (2012). *Great teachers: attracting, training and retaining the best: Government Response to the Committee's Ninth Report of Session 2010–12*.
House of Representatives Committee on Education and the Workforce (2012). *Hearing on the encouraging innovation and effective teachers act.*
Howarth, E. (2015). The 2015 NGA/TES Survey of Governors and Trustees. *Governing Matters,* September/October.
Hutchings, M. (2011). *What impact does the wider economic situation have on teachers' career decisions?: a literature review.* Department for Education Research Report RR136.
Ingersoll, R. M. & May, H. (2012). The magnitude, destinations, and determinants of mathematics and science teacher turnover. *Educational Evaluation and Policy Analysis*, 34(4), 435-464.
Jacob, B. A. & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26(1), 101-136.
Jackson, C. K. & Bruegmann, E. (2009). Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers. *American Economic Journal: Applied Economics*, 1(4), 85-108.
Jones, K. L., and P. B. Tymms. 2014. Ofsted's Role in Promoting School Improvement: The Mechanisms of the School Inspection System in England. *Oxford Review of Education* 40(3): 315–330.
Keigher, A. & Cross, F. (2010). *Teacher Attrition and Mobility: Results from the 2008-09 Teacher Follow-Up Survey.* National Center for Education Statistics 2010-353.
Klassen, R. M. & Chiu, M. M. (2011). The occupational commitment and intention to quit of practicing and pre-service teachers: Influence of self-efficacy, job stress, and teaching context. *Contemporary Educational Psychology*, 36(2), 114-129.
Kraft, M. A. & Papay, J. P. (2014). Can professional environments in schools promote teacher development? Explaining heterogeneity in returns to teaching experience. *Educational Evaluation and Policy Analysis*, 36(4), 476-500.
Krieg, J. M. (2006). Teacher Quality and Attrition. *Economics of Education Review,* 25(1), 13–27.
Lazear, E. P. & Oyer, P. (2004). Internal and external labor markets: a personnel economics approach. *Labour Economics*, 11(5), 527-554.
Loeb, S., Kalogrides, D. & Béteille, T. (2012). Effective Schools: Teacher Hiring, Assignment, Development, and Retention. *Education Finance and Policy*, 7(3), 269-304.

Leuven, E., & Sianesi, B. (2014). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. *Statistical Software Components*.
Eyles, A. & Machin, S. (2015). *The Introduction of Academy Schools to Englands Education*. IZA Discussion Papers, No. 9276.
Matthews, P & Sammons, P. (2004) *Improvement through inspection: an evaluation of the impact of Ofsted's work.* Institute of Education.
McNatt, D. B. & Judge, T. A. (2008). Self-efficacy intervention, job attitudes, and turnover: A field experiment with employees in role transition. *Human Relations*, 61(6), 783-810.
Meyer, J. P., Stanley, D. J., Herscovitch, L. & Topolnytsky, L. (2002). Affective, continuance, and normative commitment to the organization: A meta-analysis of antecedents, correlates, and consequences. *Journal of Vocational Behavior*, 61(1), 20-52.
NAO (2016) *Training New Teachers*. National Audit Office.
NFER (2008). *Teacher Resignation and Recruitment Survey (Report No. 40)*. Slough: NFER.
OECD (2015) Education at a Glance 2015.
Ofsted (2010). The Framework for School Inspection.

Richardson, B. K., Alexander, A. & Castleberry, T. (2008). Examining teacher turnover in low-performing, multi-cultural schools: Relationships among emotional labor, communication symmetry, and intent to leave. *Communication Research Reports*, 25(1), 10-22.

Ronfeldt, M. (2012). Where should student teachers learn to teach? Effects of field placement school characteristics on teacher retention and effectiveness. *Educational Evaluation and Policy Analysis*, 34(1), 3-26.

Ronfeldt, M., Loeb, S. & Wyckoff, J. (2012). How teacher turnover harms student achievement. *American Educational Research Journal*, 50(1), 4-36.

Schwab, D. P. (1991). Contextual variables in employee performance-turnover relationships. *Academy of Management Journal*, 34(4), 966-975.

Simon, N. S. & Johnson, S. M. (2013). *Teacher Turnover in High-Poverty Schools: What We Know and Can Do*. Project on the Next Generation of Teachers Working Paper.

Staiger, D. O. & Rockoff, J. E. (2010). Searching for effective teachers with imperfect information. *The Journal of Economic Perspectives*, 24(3), 97-117.

Steiner, P. M., Wroblewski, A., & Cook, T. D. (2009). Randomized experiments and quasi-experimental designs in educational research, in Ryan, K., & Cousins, J. B. (Eds.). *The SAGE international handbook of educational evaluation*. Sage, 75-95.

Supovitz, J., Sirinides, P. & May, H. (2009). How principals and peers influence teaching and learning. *Educational Administration Quarterly,* 46(1), 31-56.

Trevor, Charlie O., Gerhart, B. & Boudreau, J. W. (1997). Voluntary Turnover and Job Performance: Curvilinearity and the Moderating Influences of Salary Growth and Promotions. *Journal of Applied Psychology,* 82(1), 44–61.

West, M. R., Chingos, M. M. (2009). Teacher Effectiveness, Mobility, and Attrition in Florida, in: M. G. Springer (Ed.) *Performance Incentives: Their Growing Impact on American K-12 Education*. Brookings Institution Press, 251-271.

Annex

Table 2: Balance Tests after Matching for Schools Regraded Outstanding

Variable	Mean Characteristics				P Value
	Treated	Control	Bias (%)	Reduction in Bias (%)	
KS2 average point score	13.485	13.765	-2	54.4	0.662
KS4 Average capped point score	16.538	16.214	0.4	97.8%	0.919
% pupils free school meals	12.011	12.193	-1.4	95.9	0.732
% pupils special educational needs	4.7121	5.159	-2.5	-181.5	0.589
% pupils English additional language	11.88	12.298	-2.1	-67.7	0.658
School stage (primary/secondary)	0.07948	0.08149	-0.7	96.2	0.869

Table 3: Balance Tests after Matching for Schools Regraded Inadequate

Variable	Mean Characteristics			Reduction	P
	Treated	Control	Bias (%)	in Bias (%)	Value
KS2 average point score	17.671	17.886	-1.6	94.2	0.810
KS4 Average capped point score	13.771	14.101	-0.4	98.1	0.941
% pupils free school meals	23.452	23.344	0.8	98.6	0.919
% pupils special educational needs	2.0961	1.973	0.9	95.7	0.803
% pupils English additional language	15.624	15.362	-1.2	92.8	0.90
School stage (primary/secondary)	0.05147	0.05637	-1.7	93.8	0.757

Table 4: Missing Data for each of the Imputed Variables

Imputed Variable	Missing Before Imputation
Number Qualified FTE Teachers	2.33%
% Pupils Free School Meals	3.18%
% Special Educational Needs Statement	10.9%
% Special Educational Needs No Statement	2.86%
% White British	1.04%
% English as an Additional Language	6.05%
% Teachers Female	2.33%
% Teachers on Fixed Contract	2.33%
% Teachers on Permanent Contract	2.33%
% Teachers on Temporary Contract	2.33%
Average Age Teachers	2.33%
% Teachers White	2.33%