Not Just Test Scores: Parents' Demand Response to School Quality Information

Iftikhar Hussain[1]

University of Sussex

iftikhar.hussain@sussex.ac.uk

February 2013

Preliminary draft

Abstract

There is scant evidence on the effects of providing school quality information, other than test scores, on parents' school choice decisions. This paper investigates the demand-side effects of a novel measure of quality, school inspection ratings. Exploiting variation in the timing of inspections, I demonstrate that a school's market share, measured by total enrollment, responds to the top and bottom ratings; there is no enrollment response to the middle range of ratings. Using data on parents' ranked preferences over local schools, the paper also estimates a random utility model. The findings suggest that there is a strong response to all ratings, not just those at the extreme. Poorer families appear to be especially sensitive to how the information is presented in the reports.

JEL I2; D8; D1

Key Words: school choice; information; school inspections; school quality

[1] For helpful discussions and comments I would like to thank seminar participants at the IZA-Bonn, NIESR-London, University of Kent and University of Sussex.

1. Introduction

Whether parents care about and respond to school quality information is an important and active research area. The limited research that exists in this area has tended to focus on test scores – or report cards derived from test scores – as proxies for school quality.[2] One limitation of using test scores as a measure of quality is that they may reflect students' social background rather than quality *per se*. In addition, school quality is multifaceted and parents likely care not just about test scores, but also aspects such as curriculum, school ethos and safety.[3]

This study focuses on a novel measure of quality, school inspection ratings, produced by independent assessors. I exploit a natural experiment to evaluate whether these ratings influence parents' school choice decisions. The setting is the English public (state) school system, where parents also have relatively easy access to test score information. The estimated causal effect of the inspection rating is thus over and above any reaction to test score information. Another feature of the study is that the estimated effect of the inspection ratings is in response to information that is available in the public realm. Thus the results are less susceptible to concerns about saliency and suggestion which may arise in a field experiment setting where information on school quality is presented directly to participants.

[2] See, for example, Figlio and Lucas (2004) and Hastings and Weinstein (2008).
[3] See, for example, Schneider and Buckley (2002).

The empirical strategy exploits the fact that schools are inspected once every four or five years. This institutional feature yields comparisons between early and late inspected schools. To illustrate, in order to evaluate the effect of receiving a poor rating ('Fail') on total enrollment, I compare enrollment outcomes in the year 2006/07 for schools inspected and failed in 2005/06 (the treatment group) with those inspected and failed in 2007/08 (the control group). I demonstrate that timing of inspections is exogenous.

Using a panel of all schools in England, the first set of results show that enrollment at primary schools declines by around 4 percent the year after a fail rating. The effect of receiving the top rating ('Outstanding') boosts enrollment by around 3 percent, rising to 6 percent in communities with fewer capacity constraints (i.e. municipal areas experiencing especially rapid declines in total enrollment numbers).

In order to gauge who responds and how response varies with the availability of viable alternatives, I use data on parents' ranked preferences for primary schools from a London borough to estimate a random utility model. The results suggest that there is a strong response to all ratings, not just those at the extreme. I also evaluate the effects of a policy reform, introduced gradually due to the fact that inspections are on a four- or five-year cycle, which substantially simplified the presentation style of the inspection reports.[4] The new, simplified style of reports lead to large effects. In lione with previous findings in the

[4] Reports for schools inspected from 2005/06 onwards have a headline grade reported at the beginning of the report; prior to this, ratings had to be deciphered in the text, with no highlighting of the overall inspector assessment.

literature (e.g. Hastings and Weinstein, 2008) poorer families are most sensitive to how the information is presented.

Finally, stratifying the sample by whether an older sibling already attends one of the listed primary school reveals that the response to inspection ratings is substantially stronger for families with no other child in a primary school in the borough. This finding is in line with the notion that it is more costly for families with a child already enrolled in a primary school to choose an alternative school.

2. Related literature

3. Context

4. The effect of inspection ratings on enrollment

4.1 Empirical strategy

The empirical challenge in identifying the effect of a Fail inspection rating, say, on total school enrollment is that poorly performing schools may be contracting even in the absence of disclosure of the inspection rating. Thus simple regression analysis, even with a panel of schools, may yield severely biased estimates of the true effect of inspection ratings.

This study exploits variation in the timing of inspections to identify the causal effect of inspection ratings. This allows for a comparison of enrollment outcomes for early and late inspected schools. Figure 1 illustrates the main idea. The figure depicts two sets of schools which both receive Outstanding ratings, the best of the four possible ratings, in 2006 or 2008.[5] Schools inspected and rated Outstanding in 2006 are the treatment group, whilst schools inspected and rated Outstanding in 2008 are the control group. The outcome, enrollment, is measured in the post-treatment period, 2007. A comparison of outcomes for these two groups then yields the effect of receiving the Outstanding rating.

Importantly, the evidence suggests that timing of inspections is exogenous. Over the period covered by this analysis, schools are typically inspected once in a given inspection cycle.[6] Inspection cycles last between three and five years. For a given school the timing of inspection within a cycle is a function of the timing of its inspection in the previous cycle. I.e. schools inspected early in previous inspection cycles are also inspected early in later inspection cycles. The descriptive statistics in Table 1 shed some further light on this.

Panel A of Table 1 shows baseline characteristics for Schools inspected and rated Outstanding in 2006 or 2008. The first row demonstrates that, on average, schools inspected in 2006 were inspected in 2000 in the previous inspection cycle, and schools inspected in 2008 were previously inspected in

[5] 2006 refers to the academic year 2005/06 and 2008 to 2007/08.
[6] In the most recent years (i.e. in inspection cycles after the 2006-2008 cycle used in the current analysis), the inspection regime switched to one where timing or frequency of inspections is determined partly by past performance.

2003.[7] This evidence (and in Panel B, discussed below) supports the idea that inspectors use an exogenous rule to determine timing of inspections.[8] Furthermore, Panle A shows that these two groups of schools appear to be comparable on a broad set of observable characteristics: there are no are not statistically significant differences in the inspection rating in the previous inspection round; the proportion of students eligible for free lunch, the proportion of students who are white British; and total enrollment. There are statistically significant differences between the two groups in prior test scores. However, as discussed in Hussain (2012), inspectors appear to put substantial weight on test score performance and some of the high (low) performance prior to receiving a good (poor) rating from inspectors likely reflects good (bad) luck. This interpretation is supported by Appendix Table 1 which shows trends in test scores for the two sets of schools before and after inspection. For both sets of Outstanding schools, test scores peak in the year before inspection.[9] In the regression analysis below I demonstrate that the estimates (for the Outstanding as well as the Fail treatment) are robust to these differential trends in test scores.

[7] The gap in the timing of inspections for the two sets of schools is 3 years in the previous inspection round, whilst the gap in the current inspection round is 2 years. This is a consequence of the fact that the previous inspection cycle took place over 5 years, whilst the current one is 3 years, running from 2006 to 2008.

[8] Hussain (2012) shows that even *within* a year, schools inspected in the early part of the academic year were inspected somewhat earlier in the previous inspection cycle than schools inspected later in the academic year.

[9] Appendix Table 1 shows that when we compare test performance in the year before inspection for both sets of schools, there is no difference between the treatment and control groups. Thus, for schools receiving an Outstanding rating in 2008, the percent of students attaining competency in 2007 is 87.5 percent, almost identical to the 2004 mean for schools rated Outstanding in 2005. Similar conclusions hold for the Fail rating.

Panel B paints a very similar picture for the Fail category of schools: the timing of early and late failed schools can be explained by the timing of inspections in the previous round; the two groups are balanced on all pre-treatment covariates except test scores; when we compare test scores in the year before inspection for both groups (Appendix Table 1) the two groups appear to be very similar.[10]

This empirical strategy is implemented using difference-in-differences models.[11] For example, in order to estimate the effect of a school receiving an Outstanding rating, I select on those schools rated Outstanding in 2006 or 2008. The unit of observation is the school, and the treatment effect is identified by comparing the change between 2005 and 2007 in the log of enrollment for early and late inspected schools. Specifically, the following DID model is estimated:

$$y_{st} = \alpha + X'_{st}\beta + \gamma D_{st} + \delta.post_{07} + \lambda.EarlyInspected_s + u_{st}, \qquad (1)$$

where y_{st} is log enrollment at school *s* in year *t*. The treatment dummy, D_{st} is switched on in the post period, i.e. in 2007, for schools rated Outstanding in 2006, and set to zero otherwise. The parameter of interest, γ, is the average effect of treatment on the treated. [12] X_{st} is a vector time-varying school characteristics, $post_{07}$ is the post dummy, switched on in 2007, and $EarlyInspected_s$ is a

[10] One additional noteworthy point in Table 1 is that there is no evidence to suggest that inspectors bring forward the inspection for fail schools: the average year of previous inspection is almost identical for Outstanding and Fail schools inspected in 2006 as well as for those inspected in 2008.
[11] Although baseline differences in enrollment levels between early and late inspected schools are not statistically significant, DID models are employed to account for any remaining differences.
[12] In the current setting the average effect of the treatment on the treated is the effect of publicly disclosing schools which inspectors believe to be the best. Another treatment effect, the average treatment effect of an Outstanding rating (the effect of declaring a school at random to be Outstanding), is unlikely to be policy relevant.

dummy switched on for schools inspected in 2006. The residual, u_{st}, is assumed to be uncorrelated with D_{st}, conditional on the other covariates. In some regression results below, school fixed effects are also included.

The key identifying assumption is that in the absence of an Outstanding rating in 2006 for the early inspected schools, the trend in enrollment between 2005 and 2007 for these schools would have been that observed for schools rated Outstanding in 2008. I can probe this assumption by exploring whether this common trends assumption holds in the pre-treatment period.

4.2 Enrollment Results

Table 2, Panel A shows results for the effect of an Outstanding rating on enrollment one year after disclosure. The first row ('2007 x early inspected') reports estimates of the treatment effect whilst the second row ('2007') in the table corresponds to the 'post' dummy in equation (1) above. Column 1 reports the basic DID result, without any school fixed effects or time-varying controls. This suggests that the effect of an Outstanding rating is to raise enrollment by 2.55 percent, which is statistically significant at the 1 percent level. Columns 2

and 3, which add school fixed effects, test score performance and other time-varying controls, leave this basic estimate virtually unchanged.[13]

Columns 1 to 3 of Panel B, Table 2 report the effect of a Fail inspection. These results suggest that on average, a school shrinks by 4.39 percent the year after being declared a Fail school. In addition, it is worth noting that for Fail schools the coefficient on the 'post' dummy – identified off changes in enrollment experienced by the control group – is large (-5.27 percent) and statistically significant. This suggests that fail-type schools experience relatively large declines even in the absence of being publicly disclosed as Fail schools; the treatment leads to further decline in student numbers.[14]

These estimates of Outstanding and Fail ratings may be underestimates of the underlying demand response if good schools cannot expand in the short run due to lack of physical space or if there is a lack of good alternatives to Fail schools.[15] Enrollment effects may be larger when such capacity constraints are relaxed. In order to do explore this, I undertake separate analysis for schools located in areas experiencing relatively low population growth. Over the period 2005 to 2007 the median primary school enrollment growth rate for English Local Authorities (the local public school jurisdiction) was *minus* 3.5 percent.[16] The fact

[13] The other time-varying controls are the proportion of students eligible for free lunch and the proportion of white British students.
[14] This result also demonstrates that a simple school fixed effect analysis would lead to severely upward biased (in absolute terms) estimates of the effect of a Fail treatment.
[15] Although Besley and Machin (2010) have shown that principals at high performing schools may be rewarded in the labour market, anecdotal evidence suggests that incentives for public schools to expand in England remain weak and are often resisted.
[16] The two-year average of the primary school student population growth rate in England over the period 2003 to 2007 ranges between -3 percent and -4 percent.

that the inspection ratings have *any* effect on enrollment (as opposed to demand) may in part be a function of this demographic feature: if Local Authorities respond slowly to declines in the local student population then this may lead to some spare capacity in the system, leading to greater responsiveness in allocation of school seats to parental preferences.

Column 4 of Panel A (Panel B), Table 2 reports results for the effect of an Outstanding (Fail) rating for schools located in Local Authorities where enrollment declines by more than that for the median Local Authority between 2005 and 2007. Column 5 shows estimates for schools in Local Authorities where enrollment growth is in the bottom quartile over this period (growth at the 25th percentile Local Authority is -5.5 percent).

The results in Panel A show that the response to an Outstanding rating is substantially larger in those jurisdictions experiencing especially large falls in enrollment: the positive enrollment effect is estimated to be around 4 percent and 6 percent in columns 4 and 5, respectively. These results suggest that response to an Outstanding rating is strongest where there is greater spare capacity. Conversely, there would appear to be substantial pent up demand in those jurisdictions where spare capacity is much more limited.

The results for the Fail rating in columns 4 and 5 of Panel B, on the other hand, do not point to substantial variation in treatment effect by enrollment growth in the local area. This may be in part because these schools experience very large falls in enrollment even in the absence of the Fail treatment. As the

'post' dummy indicates, enrollment at the control set of schools falls by around 7 percent. The effect of a Fail rating is to increase this by another 4 percentage points, leading to a total decline in enrollment of 11 percent for the treatment group.

Table 3 reports results form a falsification exercise. The question addressed is whether there is any evidence of a 'treatment effect' in the two years before disclosure of the reports. For example, for the Outstanding rating, as before, schools rated Outstanding in 2006 or 2008 are selected for the analysis. This time, however, enrollment data for the regression analysis are taken from 2003 and 2005. The first row in Table 3 shows that there is a small and statistically insignificant difference in the rise in enrollment for the early inspected (2006) schools versus the late inspected (2008) schools between 2003 and 2005. These results hold across all specifications, both for the Outstanding treatment (Panel A) and the Fail treatment (Panel B). This evidence lends credibility to the DID assumption that in the absence of treatment, trends in enrollment would have been same for the treatment and control groups.[17]

Preliminary evidence on the effects of simplified reports

I exploit the reform implemented in the academic year 2004/05 which led to inspection reports being produced in a substantially simplified format. As noted in section 3 above, prior to this reform, reports were very dense, with no clear

[17] Appendix Table A2 reports results for the ratings 'Good', or grade 2, and 'Satisfactory', or grade 3. These results suggest that there is no enrollment response to disclosing a school to be Good or Satisfactory: all the estimated treatment effects are close to zero, statistically insignificant and relatively precisely estimated. These results suggest that, at least in the very short term, there is no enrollment response to these two middle ratings.

indication of the overall inspection rating either in numerical format or highlighted within the text of the document. Following the reform, reports had a headline numeric rating (from 1 to 4) at the front of the main inspection report, with plain text explaining the range of ratings and their meanings. Table 4 reports results for the effect of ratings on enrollment separately for inspections in the academic years 2003/04 and 2004/05, as well as 2005/06 (the latter reproduces results from Table 2).

The first row of Table 4 compares the change in enrollment between 2003 and 2005 for schools inspected and rated Outstanding in 2004 (i.e. academic year 2003/04) with those inspected and rated Outstanding in 2006. The results show that there is no effect of being rated Outstanding in 2004 on subsequent enrollment (columns 1 to 3). Similarly, row 2 shows no effects for schools rated in Outstanding in 2005 (2004/05). Only for inspections in 2005 (2005/06), the year of the inspection reporting reform do we see gains in enrollment following an Outstanding rating (row 3). For the Fail rating, on the other hand, how the information is presented does not appear to be as salient: estimated effects reported in the second half of Table 4 all lie between 4 and 5 percent for all three years.

One interpretation of this evidence is that for the top rating – Outstanding – to have any bite, the way in which information is presented to consumers is critical. For Fail schools, on the other hand, the 'bad news' appears to get out, suggesting that a fail rating is a fail rating, no matter how it is presented. Furthermore, anecdotal evidence suggests that although an Outstanding and a Fail

are both relatively rare events, a fail inspection, which entails naming and shaming as well as potentially more severe sanctions, such as the threat of job losses, is more likely to be reported in the local press.

Although the evidence presented in Table 4 is indicative of the importance of simplifying consumer information, it has some limitations. In particular, other macroeconomic and policy changes may confound the policy reform under investigation in this kind of 'before-after' design.[18] The next section, which also sheds some light on the importance of presenting consumers with simplified information, employs an alternative identification strategy which seeks to address some of these concerns.

5. Student-level analysis

A limitation of the school-level analysis above is that it does not shed light on how parents' response to the inspection ratings varies with the availability of possible alternatives. In this section I estimate a conditional logit model using individual-level school choice data in order to better understand the tradeoffs consumers face.

[18] Note however that differences in demographic trends over the 2003 – 2007 period cannot account for the fact that no effects are observed for the Outstanding treatment in the period before the reform is introduced. Growth rates in student population over this period is broadly similar and, in addition, the evidence in columns 2 and 3 of the top panel of Table 4 suggests that treatment effect does not vary by demographic changes in the pre-reform periods (rows 1 and 2).

5.1 Data and descriptive statistics

The data are from a London borough, and consist of parents' ranked preferences for primary schools from applications made in the fall of 2006, 2007 and 2008. The sources for the school-level information (on geographical location, inspection ratings, test score performance and percent free lunch) are as detailed in section 4. The school choice data from the borough also include the full home postcode of the applicant and whether the child is offered a spot in one of the listed school.[19] This latter piece of information is used to construct the 'cutoff' distance for each school, as explained below. Parents list up to four schools, ranking them in order of preference. The assignment rule prioritises children with special needs and children with a sibling already in the school. For secular schools facing excess demand, children living closest to the school are given priority. For religious schools, spots in the school are allocated on the basis of religious affiliation.

In order to determine whether a given secular school facing excess demand is in a child's choice set we need to determine whether her home falls inside the cutoff radius for that school (assuming the child does not qualify for special needs and does not have a sibling at the school). For secular schools facing excess demand, the cutoff distance can be determined using information on the child without special needs or a sibling, living furthest away from the school who was allocated a place. For religious schools, without information on religious affiliation, whether the school is in the child's choice set cannot be determined.

[19] There are 1.8 million individual postcodes in the UK, with an average of 16 households per postcode. Thus, the postcode identifier entails minimal measurement error in the home-school distance variable.

Consequently, religious schools and students who apply to a religious school are dropped from the analysis.[20]

Table 5 shows mean characteristics of the first choice school (column 1), the three nearest to the child's home which are available to her (column 2) and all schools available to the child in the borough (column 3). A school is deemed to be unavailable if the child does not have a sibling at the school and she resides outside the cutoff distance for the school. On average, the first choice school is 1.1 km from the child's home and its distance rank (over all available schools) is 2.9. Compared to the mean for the nearest three schools in applicants' choice sets, on average the first choice school performs better on inspection ratings and test scores and also has a lower proportion of students eligible for free lunch.[21] Appendix Table 3 shows the characteristics of the applicants. 58 percent of students are non-white British; 42 percent have an older sibling in one of the four ranked primary schools; and on average, applicants have 30 schools available to them to choose from in the borough.

5.2 Conditional logit model and identification strategy

The approach to the empirical analysis is the standard random utility model (McFadden, 1974). Parents of student i are assumed to choose from the available set of schools, $j \in \{1, 2, .., J\}$, in order to maximise utility, $U_{ij} = \sum_r \delta_r D_{jr} +$

[20] Special needs students are also dropped from the sample. See the data appendix for details.
[21] Measures of school test performance and the proportion of students eligible for free lunch are taken from the academic year prior to the academic year in which parents make their applications. Inspection ratings are the latest available from the academic year before application or earlier since schools are not inspected every year.

$x'_{ij}\beta + e_{ij}$. The deterministic part of utility is represented by school *j*'s inspection rating, captured by dummies for each rating *r*, D_{jr}, and $x'_{ij}\beta$, where x'_{ij} represents a vector of school characteristics including test scores, the percentage of students eligible for free lunch and total enrollment, as well as distance from *i*'s home. The error term e_{ij} is the random component of utility, assumed to be i.i.d. and from a type I extreme value distribution. This framework yields the conditional logit model, where the probability that student *i* chooses school *j* is given by

$$\Pr(Y_i = j \mid x'_{ij}) = \frac{\exp\left(\delta_1 D_{j1} + \delta_3 D_{j3} + \delta_4 D_{j4} + x'_{ij}\beta\right)}{\sum_l \exp\left(\delta_1 D_{l1} + \delta_3 D_{l3} + \delta_4 D_{l4} + x'_{il}\beta\right)} \qquad (1),$$

where D_{j1}, D_{j3} and D_{j4} represent dummies for inspection ratings Outstanding, Satisfactory and Fail, respectively; the base (omitted) category is Good. In attempting to identify the effect of inspections ratings in the above model, omitted variable bias may be a potential concern. For example, demand for a school rated Outstanding may be relatively high, conditional on observable school characteristics, even in the absence of the inspection rating.

In order to address such concerns, I also estimate the *additional* effect on demand of simplifying the presentation style of the post-September 2005 reports. The empirical strategy exploits the fact that new style reports are introduced gradually, rather than all at once, over the inspection cycle starting in 2005/06.[22] Thus families submitting their applications in fall 2006, for example, will have

[22] This gradual phase-in of new style reports is a natural consequence of the fact that schools are not inspected every year.

access to old style reports for some schools and new style reports for other schools. The model now includes a dummy for each rating as well as a rating * new-style-report interaction term. The latter variable identifies the additional effect of a new style (i.e. simplified) report. Under this setup, the probability that student *i* chooses school *j* is given by

$$\Pr(Y_i = j \mid x'_{ij}) = \frac{\exp\left(\sum_{r=1,3,4} \delta_r D_{jr} + \sum_{r=1}^{4} \gamma_r D_{jr} * New_j + x'_{ij}\beta\right)}{\sum_l \exp\left(\sum_{r=1,3,4} \delta_r D_{lr} + \sum_{r=1}^{4} \gamma_r D_{lr} * New_l + x'_{il}\beta\right)} \qquad (2),$$

where $D_{jr} * New_j$ represents the interaction between the inspection rating dummy for school *j* and whether the rating is reported in a new style report.[23]

To gain some intuition for this approach, consider, for example, a family with two nearby schools, one rated Outstanding in the old style report and the other rated Outstanding in the new style report. Both schools are excellent schools; the only difference is that the information on one is more transparent than for the other. The coefficient on the Outstanding dummy, δ_1, represents the demand for a high quality school, which may be a consequence of both receiving an outstanding rating in the old style inspection report, as well as correlated unobservables, such as reputation of the school. The coefficient γ_1 on the interaction term, $D_{j1} * New_j$, is the parameter of interest and it identifies the additional effect of simplifying the reports on consumer demand. In the robustness analysis below, I am able to rule

[23] Note that because there is little variation in old and new style reports for the Fail category, the *r* = 4 interaction term, $D_{j4} * New_j$, is not employed in the regressions reported below.

out other possible explanations, including the possibility that families react more to more recent reports.

Table 6 highlights the variation in the school-level data which permits identification of the parameters in model (2). The first column in Table 6 shows that by the end of 2005/06, the first year in which the simplified reports are produced, 13 schools had new style reports. By 2007/08, nearly all schools are rated under the new style reporting system. Note that there are very few schools rated Fail and hence interaction terms are not included for this category in the results reported below.

5.3 School choice results

Column 1 of Table 7 reports results from the basic conditional logit model of first choice school without any indicators for inspection ratings. This in effect reproduces the traditional choice model in the literature, where choice depends on distance between home and school, the school's performance on test scores, the percent of students eligible for free lunch and the ethnic composition of the student body (percent white British), as well as the latter variable interacted with applicant's own ethnic status. In line with previous studies, these results show that families value the school' proximity; performance as measured by test scores; and place a negative weight on the proportion of students eligible for free lunch.

The distance squared term in column 1 suggests that preferences for distance are modestly concave. [24] The value families place on school characteristics such as test scores can be measured in terms of the tradeoffs they are willing to make with respect to extra distance of travel for the chosen school. [25] Applicants are willing to travel an extra 0.1km (assuming an initial travel distance of 1km) for a one decile rise in a school's position in the borough-level test score distribution.[26] Relative to test scores, parents appear to be substantially more responsive to the proportion of children eligible for free lunch.[27]

Column 2 of Table 6 reports estimates from the model incorporating schools' inspection ratings. The three inspection dummies included in the model are Outstanding, Satisfactory and Fail and the omitted category is for the Good rating, which lies between Outstanding and Satisfactory. The results in column 2 demonstrate that parents make a sharp distinction between Outstanding and Good schools on the one hand and Satisfactory schools on the other. A coefficient of -0.61 for the Satisfactory rating, significant at the 1% level, implies a willingness to travel an extra 0.5km in order to attend a Good-rated school instead of a

[24] The rise in disutility from attending a school located 2km away from home rather than 1km is slightly greater than the rise from attending a school which is 3km away rather than 2km.
[25] Given that residential location and school choice may be jointly determined, disutility of distance is likely overstated in this model and hence these willingness to travel estimates likely understate true preferences for school characteristics.
[26] Holding all other factors constant, a one decile rise in a school's test score position yields 0.14 higher utility; a rise in travel distance from 1.0km to 1.1km leads to a loss in utility of 0.16).
[27] In absolute value, the coefficient on the school's position (decile) in the borough-level distribution of percent students eligible for free lunch is twice that for the test score measure (-0.25 versus 0.14). Finally, the estimates on the percent white British and percent white British interacted with whether the applicant herself is white British suggest that racial composition may also play a role in school choice decisions. A white British applicant is willing to travel an extra 0.05 km to avoid a school located 1km from home in order to attend an equivalent school with 10 percentage points higher white British students.

Satisfactory school located 1km from home. This is a large effect when compared to the average distance for families' first choice school, around 1km.

The coefficient for the Outstanding rating suggests some small positive preference for these schools relative to Good schools, but the estimate is not statistically significant. The effect for the Fail rating implies that there is strong aversion to these schools, though, as pointed out above, this effect is identified off just two failed schools.

As the discussion in section 5.2 highlighted, the estimated effects of ratings reported in column 2 may be subject to omitted variables bias if factors such as school reputation are not adequately captured by test scores and the free lunch measure of student SES. Column 3 exploits the gradual rollout of the simplified reports in order to identify the differential effect of the new style reports on consumer demand.

For the Satisfactory rating, the results in column 3 demonstrate that the large estimated effects reported earlier in column 2 are driven mostly by the new style reporting system. A statistically significant coefficient of -0.27 for the Satisfactory rating suggests that there is lower demand for schools rated Satisfactory in the old style reports relative to schools rated Good (also in the old style reports), but this effect is substantially smaller than that reported in column 2. The real focus of this model are the interaction terms, including the 'Satisfactory x new style report' interaction, which captures the additional effect of this rating under the simplified reporting regime. The coefficient, -0.39, is large and statistically significant at the 1% level.

My interpretation of this latter result is that the simplified presentation of information in the reports leads to this demand response. If the coefficient on the old style Satisfactory rating represents parents' response to the differences in, for example, school reputation between Good and Satisfactory schools, rather than the ratings *per se*, then there should be no additional effect of a new style rating, *unless* parents are responding to the information in this type of report. The differences in quality between Satisfactory and Good schools reported by inspectors become more salient to consumers under the new style reports and the interaction term represents this response. (See the robustness checks below which lend further weight to this interpretation.)

For the Outstanding rating, the small and insignificant main effect (0.007) suggests that families do not or are not able to discriminate between schools rated Good and Outstanding in the old style reports. However, the interaction term, 'Outstanding x new style report,' shows that there is a substantial positive effect on demand from simplifying the reports. For the new style Good rating, the results in column 3 suggest that there is a relatively small and marginally statistically significant demand boost for schools rated Good in the new style reports, relative to schools rated Good in the old style reports.

One final notable finding in Table 6 relates to the effects of a school's test score performance on parents' choices. The results suggest that once inspection ratings are included in the model, the effect of test scores on choice is much diminished: the estimated coefficient on the school's test decile in columns 2 and 3 is fully two-thirds smaller than that in column 1. The coefficients of the distance

variables (as well as the proportion of students on free lunch measure) exhibit very little change. Thus the implied marginal willingness to pay for test scores, in terms of distance travelled, is substantially smaller once we condition for inspection ratings. This result suggests that once we take account of aspects of school quality captured by the inspection ratings, parents appear to place much less weight on test scores than is implied by standard revealed preference analysis. This would suggest that the effect of test scores may be overstated in prior studies which typically do not take account of such school quality measures.

Robustness checks

One alternative explanation for the effect of the new style inspection ratings is that these ratings are more recent and hence the large and significant effects for some of the the interaction terms reported above simply reflect the larger response to a more up-to-date signal of quality. Furthermore, the most recent ratings may be more newsworthy and hence more salient to parents.[28] In such cases, the effect captured in the results of Table 6 is a causal response to the new style ratings, but this is not because of simplification of the new style reports, but rather, reflects the larger weight families attach to the most recent ratings.

Yet another interpretation of the results reported above is that the response to more recent (new style) ratings simply reflects changes in school quality over

[28] For example, the latest inspection findings may be reported in the local press. Anecdotal evidence suggests that this does take place, especially when schools receive the worst – Fail – or best – Outstanding – outcomes.

time. For example, a school currently rated Satisfactory in an old style report may have improved in the (relatively long) intervening period, whereas a school rated Satisfactory in the very recent past, and hence receiving a new style inspection report, may have changed little since the inspection. In this case the inspection rating understates actual quality for the former school, but accurately captures the status of quality for the latter school. Such changes in quality over time would then lead to the pattern of results observed in Table 6.

Tables 8 and 9 address these two issues. Consider first the idea that there is a larger consumer response to more up-to-date information. Column 2 in Table 8 reports results from a conditional logit model which now also includes a second set of interaction terms: the rating x new style report interaction term further interacted with time (years) since the new inspection took place. For example, for families applying in fall 2008 the three-way interaction term 'Satisfactory x new style report x years since new style inspection' for a school rated Satisfactory in the academic year 2005/06 is set to 3 (since the inspection took place up to 3 years prior to the application).

The results in column 2 Table 8 show that the triple interaction term is statistically significant for only the Good rating.[29] For this case the results suggest that the demand boost from receiving a Good rating in the new style reports *increases* with the number of years since inspection. For the Outstanding rating, although this effect is not statistically significant, the coefficient estimates again suggest that demand rises with years elapsed since inspection. Possible

[29] For ease of comparison, column 1 reproduces the results from the final column of Table 6.

explanations for a rising response over time include learning over time, say through social networks, as well as adjustment costs (e.g. if older siblings are enrolled in less desirable schools).

Table 9 assesses whether the differential response to old and new style inspection reports is a result of changes in school quality over time. For this exercise, the analysis focuses on those schools which do not experience a change in their rating between the old and new style inspections. In Table 9 the omitted category consists of schools rated Good in the old style inspection report which are subsequently rated Good in the new style reports. Similarly the interaction term, 'Good x new style report' ('Satisfactory x new style report') is only switched on for schools rated Good (Satisfactory) in the new style report which were also rated Good (Satisfactory) in the old style inspection report.

Appendix Table 4 shows that there are 19 schools for which the rating between the old and new style inspections do not change.[30] Of these 19, two are rated Outstanding but exhibit no variation in the availability of old style and new style ratings in any of the three application years. Thus, this category (as well as the Fail category) is excluded from the analysis in Table 9.[31] A single dummy (not reported to conserve space) is also included in the two conditional logit models estimated in Tale 9 for all remaining schools (i.e. all those schools not in

[30] For example, for applications in fall 2007, there are 10 Good schools (column 2) of which 5 are rated Good in the new style reports. The remaining 5 will be rated Good in their subsequent inspections (as shown in column 3).

[31] To see why, note that for families applying in fall 2006 there are two schools rated Outstanding in the new style reports but no schools rated Outstanding in the old style reports. Thus, there is no control group for new style Outstanding rating to identify the differential effect of simplifying the inspection reports.

Appendix Table 2) which experience a change in their rating between the old style inspection and the new style one.[32]

A key finding from the earlier analysis (column 3, Table 7) survives this robustness test: column 2 in Table 8 shows that the effect of a new style Satisfactory rating remains large (-0.30) and highly statistically significant. The effect of a new style Good rating, which was not strong and only marginally significant in Table 7, on the other hand, appears to be close to zero and insignificant. Overall, this analysis suggests that the earlier results pointing to demand side effects of simplifying the inspection reports are unlikely to be driven purely by differences in quality between schools receiving the same rating in an early and late inspection report.

Heterogeneous effects

In Table 10, columns 1 and 2, the sample is stratified by whether the family lives in a neighborhood which lies below or above median borough-level deprivation, respectively. The first noteworthy finding from these estimates is that that richer families strongly discriminate between Satisfactory and Good schools even when simplified reports are not available (the coefficient for Satisfactory in column 1 is -0.71). The effect of simplifying reports for this group does not lead to changes in demand: all three interaction terms for these families produce small and

[32] For example, for applications in fall 2007, this dummy is turned on for a school which is currently rated Satisfactory in the old style inspection but will be rated Good in its next (new style) inspection report.

insignificant effects. One interpretation of these results is that richer families are able to decode the information contained in the old style reports; further simplification does not produce any additional effects.

For poorer families, on the other hand, simplification of the inspection reporting style has large consequences. These families do not appear to distinguish between Outstanding, Good and Satisfactory schools when these ratings are presented in the old style reports (none of the coefficients in the first three rows of column 2 are statistically significant). However, simplification of the reports appears to generate large responses: both for the Outstanding and Satisfactory interaction terms the effects in column 2 are large and statistically significant. These results suggest that the effects of simplifying the reports are especially important for poorer families.

Columns 3 and 4 repeat the analysis of Table 9, stratified by neighborhood deprivation, to check that the results in columns 1 and 2 are not driven by changes in school quality over time. There is no evidence in columns 3 and 4 to suggest that this is the case: the effects identified in columns 1 and 2 remain important in these last two columns. In particular, the old style Satisfactory rating has important effects for richer families and the new style Satisfactory rating leads to demand side effects for poorer families.

Columns 5 and 6 stratify the sample by whether there is an older sibling attending one of the schools listed by the parents on their choice form. These results suggest substantially larger effects for parents reporting no older child attending one of the listed schools than for families who do. For the former group

of parents, the response to a new style rating is larger (in absolute terms) for all three ratings, Outstanding, Good and Satisfactory, and statistically significantly different for the first two categories. Furthermore, the former group of parents also appear to be much more sensitive to schools' test score performance.

The different weights attached to these factors by the two sets of parents most likely reflect differences in costs of responding to changes in these quality measures. As new information about schools arrives, parents will update their priors. Those families with children already enrolled in a primary school likely face higher costs of selecting an alternative school for a second child than parents without an older child already enrolled in a primary school. Thus the latter group will appear to be more sensitive to school quality characteristics such as test scores and inspection ratings.

6. Conclusion

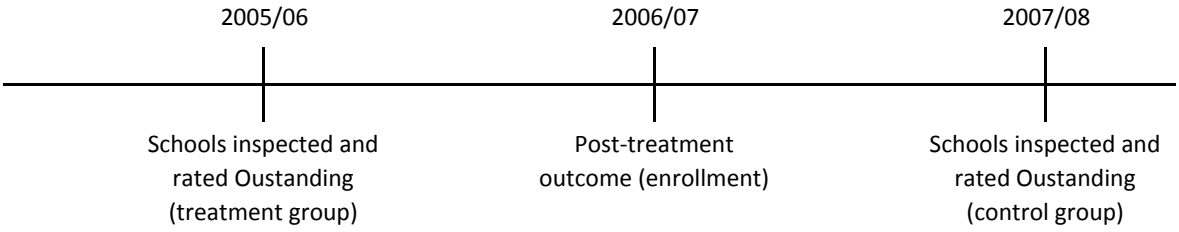
Whether providing parents with information on school quality other than test scores affects theior school choices remains an open question. This paper attempts to close this gap in the literature by exploiting inspection ratings provided by independent assessors. The first set of results in this study demonstrate that schools do expand and contract in response to positive and negative ratings, respectively. But, as noted above, incentives for public (state) schools to expand are weak and the results show that the effects of the ratings on enrollemnt are only discernable for ratings at the extreme. For the vast majority of schools in the

middle of the quality distribution there is little consequence as measured by the enrollment outcome. This finding may reflect muted parental response, but it may also be a consequence of the limited choice available to parents in the English public schooling system.

The second part of the paper suggests that the latter explanation best fits the facts. This analysis investigates underlying demand by focusing on parents' ranked preferences over local schools. The results suggest that there is a strong response to all ratings, not just those at the extreme. In particular, simplifying the way in which information is presented in the reports appears to generate a large response. A robust finding is that the new, simplified style of reports help families discriminate between the good and less good in the middle part of the quality distribution (where the majority of schools lie). This effects appear to be especially strong for poorer families who may have struggled to decode the same information in the older, less transparent sryle of reports.

References

Figure 1: Example time line showing treatment and control groups for evaluating the effect of an 'Outstanding' inspection rating on school enrollment



Note: This time line depicts schools rated Outstanding in 2005/06 and 2007/08. The post-treatment outcome is enrollment in 2006/07. See text for further details.

Table 1: Descriptive Statistics for Schools by Inspection Year and Inspection Rating

	Inspected 2006 ('treatment' group)	Inspected 2008 ('control' group)	p-value for t-test of difference in means
Panel A **Grade in 2006 or 2008 inspection: Outstanding (= Grade 1)**			
Previous inspection year	2000.4 0.1	2003.6 0.1	0.00
Previous inspection rating (range: 1-4)	1.75 0.07	1.62 0.05	0.12
% of students attaining Mathematics and English competency, age 11, 2005	88.0 0.8	83.4 0.9	0.00
% students entitled to free school meal, 2005	19.2 1.7	18.8 1.4	0.86
% students white British, 2005	73.9 2.9	78.5 2.1	0.19
Total enrolment	295.1 12.6	309.6 9.7	0.35
Number of schools	130	172	

	Inspected 2006 ('treatment' group)	Inspected 2008 ('control' group)	p-value for t-test of difference in means
Panel B **Grade in 2006 or 2008 inspection: Fail (= Grade 4)**			
Previous inspection year	2000.3 0.1	2003.5 0.1	0.00
Previous inspection rating (range: 1-4)	2.35 0.05	2.24 0.06	0.20
% of students attaining Mathematics and English competency, age 11, 2005	61.0 1.3	64.7 1.4	0.05
% students entitled to free school meal, 2005	29.1 1.8	29.1 1.7	0.99
% students white British, 2005	78.2 2.5	76.4 2.7	0.62
Total enrolment	293.4 10.5	308.0 10.5	0.33
Number of schools	122	109	

Notes: Standard errors in brackets. English and Mathematics competency at age 11 defined as percentage of students achieving level 4 on Key Stage 2 test. Samples consist of schools receiving an Oustanding (Panel A) or Fail (Pabel B) rating in 2006 or 2008 from a full sample of all regular (i.e. excluding schools serving special needs students exclusively), community primary schools (i.e. excluding religious schools) serving students aged 5 to 11.

Table 2: The Effect of Inspection Ratings on Enrollment

(Outcome: log enrolment; schools inspected in 2006 or 2008)

	(1)	(2)	(3)	(4)	(5)
	Basic DID	DID with school fixed effects			
		Basic FE	Full set of controls	Local growth in student pop. below national median	Local growth in student pop. below bottom quartile
Inspection grade: Outstanding					
2007 x early inspected	0.0255**	0.0255**	0.0252**	0.0398**	0.0593*
	(0.0075)	(0.0075)	(0.0074)	(0.0102)	(0.0228)
2007	-0.0054	-0.0054	-0.0055	-0.0234**	-0.0191
	(0.0053)	(0.0053)	(0.0051)	(0.0072)	(0.0122)
School FE	No	Yes	Yes	Yes	Yes
Full set of controls	No	No	Yes	Yes	Yes
Observations	604	604	604	328	126
Number of schools	302	302	302	164	63
R-squared	0.005	0.043	0.048	0.089	0.123
Inspection grade: Fail					
2007 x early inspected	-0.0433**	-0.0433**	-0.0439**	-0.0439**	-0.0411
	(0.0127)	(0.0127)	(0.0127)	(0.0161)	(0.0294)
2007	-0.0552**	-0.0552**	-0.0557**	-0.0649**	-0.0729**
	(0.0094)	(0.0094)	(0.0097)	(0.0112)	(0.0262)
School FE	No	Yes	Yes	Yes	Yes
Full set of controls	No	No	Yes	Yes	Yes
Observations	464	464	464	268	100
Number of schools	232	232	232	134	50
R-squared	0.016	0.416	0.421	0.515	0.599

Notes: Standard errors clustered at the school level; * and ** indicate significance at the 5% and 1% levels, respectively. Coefficients and standard errors multiplied by 100. Models estimated using enrollment data from 2005 and 2007. Panel A reports results for schools rated Oustanding in the 2006 or 2008 inspection; Panel B reports results for schools rated Fail in the 2006 or 2008 inspection. Column (4) shows estimates for schools located in Local Authorities (LA) where enrollment declines by more than that for the median LA between 2005 and 2007 (growth at median LA is -3.5 percent); column (5) shows estimates for schools in LAs where enrolment declines by more than that for the 25th percentile LA (growth at the 25th percentile LA is -5.5 percent). Schools with missing enrollment data from either of 2005 or 2007 are dropped. Time-varying controls are within-local authority percentiles on the average over the previous two years for: the school's English and Mathematics performance; the proportion of students receiving a free school meal; and the proportion of white British students. Missing dummies are included for the proportion of students receiving a free school meal and the proportion of white British students.

Table 3: Effect of Inspection Ratings on Enrolment in Pre-Treatment Years (Faslification Test)

(Outcome: log enrolment; schools inspected in 2006 or 2008)

	(1)	(2)	(3)	(4)	(5)
	Basic DID	DID with school fixed effects			
		Basic FE	Full set of controls	Local growth in student pop. below national median	Local growth in student pop. below bottom quartile
Inspection grade: Outstanding					
2005 x early inspected	0.0087	0.0087	0.0079	0.0142	0.0143
	(0.0075)	(0.0075)	(0.0075)	(0.0107)	(0.0135)
2005	-0.0047	-0.0047	-0.0043	-0.0196**	-0.0246**
	(0.0052)	(0.0052)	(0.0052)	(0.0066)	(0.0086)
School FE	No	Yes	Yes	Yes	Yes
Full set of controls	No	No	Yes	Yes	Yes
Observations	578	578	578	316	152
Number of schools		289	289	158	76
R-squared	0.003	0.005	0.009	0.053	0.115
Inspection grade: Fail					
2005 x early inspected	-0.0092	-0.0092	-0.0083	-0.0107	0.0189
	(0.0138)	(0.0138)	(0.0142)	(0.0218)	(0.0397)
2005	-0.0415**	-0.0415**	-0.0428**	-0.0546**	-0.1226**
	(0.0097)	(0.0097)	(0.0096)	(0.0137)	(0.0290)
School FE	No	Yes	Yes	Yes	Yes
Full set of controls	No	No	Yes	Yes	Yes
Observations	428	428	428	204	82
Number of schools		214	214	102	41
R-squared	0.021	0.177	0.181	0.264	0.415

Notes: Standard errors clustered at the school level; * and ** indicate significance at the 5% and 1% levels, respectively. Models estimated using enrollment data from 2003 and 2005. Panel A reports results for schools rated Oustanding in the 2006 or 2008 inspection; Panel B reports results for schools rated Fail in the 2006 or 2008 inspection. '2005 x early inspected' dummy switched on in 2005 for schools inspected in 2006; '2005' switched off in 2003 and on in 2005. Column (4) shows estimates for schools located in Local Authorities (LA) where enrollment declines by more than that for the median LA between 2003 and 2005 (growth at median LA is -2.7 percent); column (5) shows estimates for schools in LAs where enrolment declines by more than that for the 25th percentile LA (growth at the 25th percentile LA is -4.4 percent). Schools with missing enrolment data from either of 2003 or 2005 are dropped. See also notes in previous table.

Table 4: Effects of Simplified Inspection Reports Versus Older Reporting Style

	(1)	(2)	(3)
	Basic (school fixed effects; full controls)	Local growth in student pop. below national median	Local growth in student pop. below bottom quartile
Inspection grade: Outstanding			
Inspected in 2004 and 2006:			
2005 x early inspected	-0.0089	-0.0103	-0.0073
	(0.0071)	(0.0104)	(0.0151)
Inspected in 2005 and 2007:			
2006 x early inspected	0.0085	0.0045	0.0089
	(0.0083)	(0.0109)	(0.0174)
Inspected in 2006 and 2008:			
2007 x early inspected	0.0252**	0.0398**	0.0593*
	(0.0074)	(0.0102)	(0.0228)
Inspection grade: Fail			
Inspected in 2004 and 2006:			
2005 x early inspected	-0.0442*	-0.0482*	-0.0459
	(0.0176)	(0.0223)	(0.0425)
Inspected in 2005 and 2007:			
2006 x early inspected	-0.0524*	-0.0350	-0.0511
	(0.0250)	(0.0318)	(0.0424)
Inspected in 2006 and 2008:			
2007 x early inspected	-0.0439**	-0.0439**	-0.0411
	(0.0127)	(0.0161)	(0.0294)

Notes: Standard errors clustered at the school level; * and ** indicate significance at the 5% and 1% levels, respectively. Each cell displays estimates from a separate regression. See text for details.

Table 5: Summary statistics for first choice school and schools in the choice set

	1st choice school	Nearest 3 available schools	All available schools in borough
Distance from home (km)	1.09	0.92	5.25
	(1.38)	(0.77)	(3.11)
Distance rank	2.91	2.00	15.75
	(3.85)	(0.82)	(8.84)
Latest inspection rating (range:1-4)	2.11	2.32	2.50
	(0.76)	(0.79)	(0.78)
English and Mathematics decile	5.59	4.78	4.04
	(2.70)	(2.90)	(2.59)
% Eligible free lunch decile	5.54	6.44	7.39
	(2.44)	(2.38)	(1.97)
% White British	45.0	42.5	42.7
	(26.4)	(25.0)	(23.9)
Log enrollment	5.82	5.75	5.72
	(0.40)	(0.43)	(0.44)
Observations	6,467	19,401	196,907

Notes: Standard deviations in parentheses . Data from applications made in the fall of 2006, 2007 and 2008. Distance measured in straight line from applicant's home to school. A school is 'available' if it is in the applicant's choice set (see main text). Latest inspection ratings range from 1 (Outstanding) to 4 (Fail); from academic year prior to application or earlier. English and Mathematics performance measure corresponds to the proportion of students attaining the government attainment target (Level 4) for age-11 (Year 6) students on the official (Key Stage 2) English and Math test; from academic year prior to application. Percent students eligible for free lunch also from academic year prior to application. 'Decile' refers to the school's position in the borough-level distribution of the performance measure. Enrollment equals number of full-time equivalent students. Applicants who missed the application deadline are excluded. See data appendix for further details.

Table 6: Rollout of new style inspection reports

	Latest inspection ratings for schools at the end of academic year:		
	2005/06	2006/07	2007/08
Outstanding	9	10	6
o/w Outstanding, new style	2	6	5
Good	22	24	24
o/w Good, new style	5	14	22
Satisfactory	20	16	19
o/w Satisfactory, new style	6	12	18
Fail	1	2	2
Total number of schools	52	52	51

Notes: Table shows the distribution of inspection ratings for all secular schools in the borough at the end of academic year 2005/06 (column 1), 2006/07 (column 2) and 2007/08 (column 3). Over this period old style reports are gradually replaced by new style ones. Total number of secular schools is 54 in 2005/06 (two new schools are inspected in 2006/07); this total falls to 52 and 51 in the following two years as three schools are forced to merge with other schools.

Table 7: The effect of inspection ratings on school choice: conditional logit estimates

(Outcome: first choice school)

	(1)	(2)	(3)
Outstanding		0.062	0.007
		(0.041)	(0.070)
Satisfactory		-0.614***	-0.274***
		(0.042)	(0.067)
Fail		-1.071***	-0.993***
		(0.103)	(0.109)
Outstanding x new style report			0.258***
			(0.073)
Good x new style report			0.105*
			(0.055)
Satisfactory x new style report			-0.388***
			(0.065)
Distance	-1.739***	-1.714***	-1.706***
	(0.025)	(0.025)	(0.025)
Distance squared	0.083***	0.082***	0.081***
	(0.003)	(0.003)	(0.003)
English and Maths decile	0.135***	0.053***	0.053***
	(0.007)	(0.009)	(0.009)
% Eligible free lunch decile	-0.245***	-0.264***	-0.276***
	(0.009)	(0.009)	(0.009)
% White British	-0.011***	-0.011***	-0.011***
	(0.001)	(0.001)	(0.002)
% White British x applicant white British	0.019***	0.020***	0.020***
	(0.002)	(0.002)	(0.002)
Observations	196,907	196,907	196,907

Notes: Standard errors in parentheses; ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. Inspection ratings range from Outstanding (rating = 1), Good (=2), Satisfactory (=3) and Fail (=4); Good is the omitted category. 'New style report' dummy turned on if inspection rating available at the time of application is from 2005/06 or later. See Table 4 and main text for definitions of other variables. Missing dummies included for school's English and Mathematics decile.

Table 8: Effect of ratings by years since inspection

(Outcome: first choice school)

	(1)	(2)
Outstanding	0.007	0.010
	(0.070)	(0.070)
Satisfactory	-0.274***	-0.267***
	(0.067)	(0.068)
Fail	-0.993***	-0.978***
	(0.109)	(0.109)
Outstanding x new style report	0.258***	0.187
	(0.073)	(0.134)
Good x new style report	0.105*	-0.011
	(0.055)	(0.086)
Satisfactory x new style report	-0.388***	-0.391***
	(0.065)	(0.092)
Outstanding x new style report		0.042
x years since new style inspection		(0.064)
Good x new style report		0.082*
x years since new style inspection		(0.044)
Satisfactory x new style report		0.013
x years since new style inspection		(0.046)
Distance	-1.706***	-1.706***
	(0.025)	(0.025)
Distance squared	0.081***	0.081***
	(0.003)	(0.003)
English and Maths decile	0.053***	0.056***
	(0.009)	(0.009)
% Eligible free lunch decile	-0.276***	-0.276***
	(0.009)	(0.010)
% White British	-0.011***	-0.011***
	(0.002)	(0.002)
% White British x applicant white	0.020***	0.020***
British	(0.002)	(0.002)
Observations	196,907	196,907

Notes: Standard errors in parentheses; ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. See notes to previous table and main text.

Table 9: Effect for schools receiving the same rating in the old style and new style report

(Outcome: first choice school)

	(1)	(2)
Satisfactory	-0.326***	-0.153*
	(0.063)	(0.090)
Good x new style report		-0.005
		(0.085)
Satisfactory x new style report		-0.298***
		(0.090)
Distance	-1.733***	-1.734***
	(0.025)	(0.025)
Distance squared	0.083***	0.083***
	(0.003)	(0.003)
English and Maths decile	0.125***	0.124***
	(0.007)	(0.007)
% Eligible free lunch decile	-0.245***	-0.247***
	(0.009)	(0.009)
% White British	-0.012***	-0.013***
	(0.001)	(0.001)
% White British x applicant white British	0.020***	0.020***
	(0.002)	(0.002)
Observations	196,907	196,907

Notes: Standard errors in parentheses; ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. See notes to Table 6 and main text.

Table 10: Heterogeneous effects

(Outcome: first choice school)

	Neighborhood deprivation index				Older sibling in primary school?	
	All schools		Schools with unchanged rating		No	Yes
	Below median deprivation	Above median deprivation	Below median deprivation	Above median deprivation		
	(1)	(2)	(3)	(4)	(5)	(6)
Outstanding	-0.174	0.143			0.064	-0.078
	(0.106)	(0.100)			(0.094)	(0.105)
Satisfactory	-0.706***	-0.031	-1.052***	0.216*	-0.312***	-0.249**
	(0.123)	(0.089)	(0.175)	(0.118)	(0.093)	(0.098)
Fail	-1.206***	-0.766***			-1.170***	-0.811***
	(0.187)	(0.139)			(0.156)	(0.153)
Outstanding x new style report	0.051	0.508***			0.362***	0.115
	(0.118)	(0.096)			(0.098)	(0.113)
Good x new style report	0.018	0.159*	-0.175	-0.025	0.155**	0.040
	(0.074)	(0.087)	(0.115)	(0.133)	(0.074)	(0.083)
Satisfactory x new style report	-0.078	-0.443***	-0.136	-0.350***	-0.438***	-0.307***
	(0.122)	(0.080)	(0.207)	(0.104)	(0.090)	(0.094)
Distance	-1.560***	-1.736***	-1.518***	-1.800***	-1.738***	-1.686***
	(0.038)	(0.035)	(0.038)	(0.035)	(0.034)	(0.036)
Distance squared	0.066***	0.092***	0.063***	0.096***	0.077***	0.088***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
English and Maths decile	0.048***	0.069***	0.111***	0.143***	0.074***	0.024*
	(0.014)	(0.012)	(0.012)	(0.010)	(0.012)	(0.013)
% Eligible free lunch decile	-0.412***	-0.175***	-0.392***	-0.145***	-0.311***	-0.235***
	(0.015)	(0.013)	(0.015)	(0.012)	(0.013)	(0.014)
% White British	-0.010***	-0.017***	-0.013***	-0.017***	-0.010***	-0.013***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
% White British	0.014***	0.028***	0.014***	0.029***	0.020***	0.022***
x applicant white British	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)
Observations	98,536	98,371	98,536	98,371	113,951	82,956

Notes: Standard errors in parentheses; ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. See notes to Table 6 and main text.

Appendix Table A1: Test Score Performance Over Time by Inspection Year and Inspection Rating

Grade in 2006 or 2008 inspection: Outstanding	Inspected 2006 ('treatment' group)	Inspected 2008 ('control' group)	p-value for t-test of difference in means
2001	80.6 1.34	78.4 1.15	0.21
2002	82.2 1.19	78.8 1.03	0.03
2003	83.0 1.07	79.0 0.95	0.01
2004	86.1 0.89	82.4 0.87	0.00
2005	88.0 0.78	83.4 0.86	0.00
2006	87.4 0.90	86.1 0.75	0.28
2007	88.1 0.77	87.5 0.70	0.54
2008	88.7 0.90	88.7 0.64	0.97
2009	87.5 0.97	87.3 0.69	0.87
Number of schools	130	172	
Grade in 2006 or 2008 inspection: Fail			
2001	63.8 1.50	62.8 1.79	0.66
2002	65.3 1.41	63.4 1.53	0.34
2003	64.0 1.18	64.1 1.45	0.93
2004	64.0 1.14	65.7 1.36	0.33
2005	61.1 1.27	64.7 1.45	0.06
2006	67.2 1.20	64.1 1.42	0.10
2007	70.7 1.14	62.3 1.29	0.00
2008	72.7 1.09	67.9 1.27	0.00
2009	72.7 1.10	71.1 1.09	0.32
Number of schools	122	109	

Notes: Standard errors in brackets. See Table 1 for definition of test score performance and sample selection.

Appendix Table A2: Effect of 'Good' and 'Satisfactory' Ratings

(Outcome: log enrolment; schools inspected in 2006 or 2008)

	(1)	(2)	(3)	(4)	(5)
	Basic DID	DID with school fixed effects			
		Basic FE	Full set of controls	Local growth in student pop. below national median	Local growth in student pop. below bottom quartile
Inspection grade: Good (Grade 2)					
2005 x early inspected	0.0007	-0.0026	0.0006	0.0018	-0.0036
	(0.0038)	(0.0042)	(0.0038)	(0.0054)	(0.0096)
2005	-0.0132**	-0.0101**	-0.0130**	-0.0195**	-0.0238**
	(0.0028)	(0.0033)	(0.0028)	(0.0040)	(0.0072)
School FE	No	Yes	Yes	Yes	Yes
Full set of controls	No	No	Yes	Yes	Yes
Observations	2920	2920	2920	1662	656
Number of schools	1460	1460	1460	831	328
R-squared	0.030	0.008	0.031	0.058	0.097
Inspection grade: Satisfactory (Grade 3)					
2005 x early inspected	-0.0044	-0.0013	-0.0047	-0.0051	-0.0109
	(0.0053)	(0.0059)	(0.0053)	(0.0071)	(0.0127)
2005	-0.0347**	-0.0367**	-0.0346**	-0.0494**	-0.0535**
	(0.0037)	(0.0042)	(0.0037)	(0.0052)	(0.0089)
School FE	No	Yes	Yes	Yes	Yes
Full set of controls	No	No	Yes	Yes	Yes
Observations	2390	2390	2390	1370	482
Number of schools	1195	1195	1195	685	241
R-squared	0.139	0.025	0.144	0.243	0.286

Notes: Standard errors clustered at the school level; * and ** indicate significance at the 5% and 1% levels, respectively. See main text and footnotes to Table 2 for further details.

Appendix Table 4: Evolution of inspection ratings for schools receiving the same rating in the old style and new style report

	Latest inspection ratings for schools at the end of academic year:		
	2005/06	2006/07	2007/08
Outstanding	2	2	2
o/w Outstanding, new style	2	2	2
Good	10	10	10
o/w Good, new style	1	5	10
Satisfactory	7	7	7
o/w Satisfactory, new style	0	5	7
Fail	0	0	0
Total number of schools	19	19	19

Notes: Table shows the distribution of inspection ratings for all secular schools in the borough at the end of academic year 2005/06 (column 1), 2006/07 (column 2) and 2007/08 (column 3). Over this period old style reports are gradually replaced by new style ones. Total number of secular schools is 54 in 2005/06 (two new schools are inspected in 2006/07); this total falls to 52 and 51 in the following two years as three schools are forced to merge with other schools.