

# The Inspector Lottery: Ofsted Ratings and the Spatial Inequality of School Quality Labels

APEP Autonomous Research\*      @olafdrw

April 3, 2026

## Abstract

A parent moving to a new neighborhood in England will almost certainly check the local school's Ofsted rating before signing a lease. But does the label itself move house prices, or does it merely reflect quality that buyers already observe? I exploit the timing of 9,780 Ofsted school inspections (2019–2024) as information shocks, comparing house prices in 2,304 postcode districts before and after rating publication. The pooled effect of a negative rating (Requires Improvement or Inadequate) is near zero. But this masks a sharp gradient: in deprived areas, a bad rating is associated with a 1.1% price decline, while in affluent areas prices rise by 1.0%—a 2.1 percentage-point gap. School quality labels appear regressive: they are associated with lower property values precisely where families are most constrained.

**JEL Codes:** R21, I28, H75

**Keywords:** school quality, house prices, Ofsted, inspection, capitalization, deprivation

---

\*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 52m).

# 1. Introduction

In 2023, a London estate agent’s listing for a two-bedroom flat in Lambeth advertised “Outstanding Ofsted primary school within 200 meters” before mentioning the flat’s own square footage. The detail is telling: in England’s housing market, proximity to a well-rated school is a selling point on par with the property itself. Ofsted, the Office for Standards in Education, inspects every state-funded school on a rolling cycle and publishes ratings from Outstanding (1) to Inadequate (4). These ratings are public, prominent, and—as this paper shows—differentially capitalized into house prices depending on who lives nearby.

The capitalization of school quality into property values is one of the most robust findings in urban economics. Since Black (1999) documented discontinuous price jumps at school attendance boundaries, a large literature has confirmed that buyers pay premiums for access to higher-quality schools (Gibbons et al., 2013; Bayer et al., 2007; Figlio and Lucas, 2004). But most studies identify the price of *quality differences*—comparing areas with persistently different schools. The distinct question of whether a *label change*—a new rating assigned by an inspector—moves prices conditional on underlying quality has received less attention. If it does, the rating system creates arbitrary winners and losers in the housing market.

This paper asks whether Ofsted inspection ratings causally affect local house prices, and whether that effect varies by neighborhood deprivation. I construct an event study around 9,780 graded inspections conducted between 2019 and 2024, linking each school’s postcode district to monthly house price data from the HM Land Registry’s universe of residential transactions. The identification strategy is a difference-in-differences design: I compare house prices in postcode districts of schools receiving negative ratings (Requires Improvement or Inadequate) to those receiving positive ratings (Outstanding or Good), before and after the inspection publication date. School fixed effects absorb permanent differences in neighborhood quality; calendar-month fixed effects absorb aggregate price trends.

The pooled estimate is small and counterintuitive: a negative Ofsted rating is associated with a 0.36% *increase* in local house prices (SE = 0.16%), which is statistically significant but not robust to permutation inference ( $p = 0.126$ ). The event study shows no systematic pre-trends (pre-trend test  $p = 0.138$ ), but also no sharp post-publication break. The pooled effect is thus fragile and economically ambiguous—a finding that, by itself, would suggest Ofsted ratings carry little new information for the housing market.

The aggregation, however, conceals a striking heterogeneity. When I split by the Income Deprivation Affecting Children Index (IDACI), the pattern reverses: in deprived neighborhoods (IDACI quintiles 1–2), a bad rating depresses house prices by 1.1% ( $p < 0.001$ ), while in non-deprived neighborhoods, prices rise by 1.0% ( $p < 0.001$ ). The 2.1 percentage-point

gap is robust across event windows, property types, and school phases. This is the paper’s central contribution: school quality labels are *regressive*. They penalize property values—and by extension household wealth—precisely where families have the fewest outside options.

Why the asymmetry? In affluent areas, families can exit to private schools, relocate to adjacent catchment areas, or supplement with tutoring. A bad Ofsted label may even signal a buying opportunity for investors anticipating future improvement. In deprived areas, the local school is often the only realistic option. A bad rating becomes self-reinforcing: it depresses property values, discourages mobile families from moving in, and concentrates disadvantage—the “school quality trap” that [Hastings et al. \(2009\)](#) document in the US context. The Ofsted system, designed to improve accountability, may inadvertently amplify spatial inequality through the housing market channel.

An important caveat: the DiD design cannot fully isolate the *label effect* from correlated quality changes. Schools receiving bad ratings may genuinely be declining, and the price response may reflect quality rather than information. A stronger design would exploit quasi-random variation in inspector leniency—as [Bokhove et al. \(2023\)](#) document, 9.5% of rating variation lies between inspectors—but linking individual inspectors to inspections requires data not currently available in machine-readable form. The present paper should be read as documenting a robust *association* between ratings and house prices that varies sharply by deprivation, motivating future work with stronger identification.

This paper contributes to three literatures. First, it extends the school quality capitalization literature ([Black, 1999](#); [Gibbons et al., 2013](#); [Bayer et al., 2007](#); [Kane et al., 2006](#); [Machin, 2011](#); [Fack and Grenet, 2010](#)) by documenting the differential pricing of school quality labels across neighborhood types. [Hussain \(2016\)](#) studies Ofsted ratings and house prices using a hedonic approach but does not examine heterogeneity by deprivation. My event study design, with school fixed effects, absorbs permanent neighborhood quality, though it cannot fully separate labels from time-varying quality changes.

Second, it contributes to the literature on information and housing markets ([Pope, 2008](#); [Figlio and Lucas, 2004](#); [Cellini et al., 2010](#)). The finding that labels matter differentially by deprivation suggests that information effects depend on the buyer’s choice set: information is only valuable when alternatives exist. This connects to [Hastings et al. \(2009\)](#), who show that school choice effects are concentrated among informed, mobile families.

Third, it speaks to the policy evaluation of accountability systems. England’s Ofsted framework is among the world’s most consequential school rating systems, and its housing market effects have distributional implications that policymakers have not fully reckoned with. The finding that labels depress property values in deprived areas—where schools are most likely to receive bad ratings—suggests an unintended feedback loop between accountability

and inequality.

The paper proceeds as follows. Section 2 describes the Ofsted inspection framework. Section 3 details the data construction. Section 4 presents the empirical strategy. Section 5 reports results, including the deprivation heterogeneity that is the paper’s main finding. Section 6 discusses mechanisms and implications.

## 2. Institutional Background

**The Ofsted Inspection Framework.** Ofsted (the Office for Standards in Education, Children’s Services and Skills) inspects all state-funded schools in England. Since the 2019 Education Inspection Framework (EIF), inspections produce an Overall Effectiveness rating on a 1–4 scale: 1 (Outstanding), 2 (Good), 3 (Requires Improvement), and 4 (Inadequate). Schools rated Outstanding were historically exempt from routine re-inspection, though this exemption was removed in 2020. Inspections follow a roughly four-year cycle, with shorter “ungraded” visits between full inspections.

**Inspection Process and Timing.** A typical graded inspection lasts two days. Lead inspectors are assigned by regional scheduling teams based on availability and geographic proximity—not on school characteristics. Schools receive approximately one day’s notice before an inspection begins. The resulting report, including the rating, is published on the Ofsted website within weeks. Publication is the key information event: before publication, the rating is private; after publication, it is freely available to parents, homebuyers, and estate agents.

**Stakes and Salience.** Ofsted ratings are high-stakes signals. Schools rated Inadequate may be placed in Special Measures, triggering governance changes and potential academization. Estate agents routinely advertise proximity to Outstanding-rated schools. The property website Rightmove includes Ofsted ratings in its school search tool. Local newspaper coverage of inspections is extensive, particularly for downgrades. The salience of the rating is thus highest at the point of publication—precisely the event this paper exploits.

**The Rating Distribution.** The distribution of ratings is heavily skewed. In the 2019–2024 inspection data, 11.5% of schools are rated Outstanding, 75.6% Good, 10.8% Requires Improvement, and 2.0% Inadequate. The vast majority of schools receive positive ratings, making negative ratings—and especially downgrades—salient news events.

### 3. Data

I combine two administrative datasets covering England.

**Ofsted Management Information.** The primary source is Ofsted’s Management Information dataset, a monthly CSV published on GOV.UK that records the latest graded inspection outcome for every state-funded school (Ofsted, 2026). The dataset contains 21,966 schools, of which 9,781 have received a graded inspection under the 2019 EIF. For each school, I observe the Unique Reference Number (URN), postcode, inspection and publication dates, Overall Effectiveness rating (1–4), and school characteristics including phase (Primary/Secondary), pupil count, admissions policy, and the Income Deprivation Affecting Children Index (IDACI) quintile.

**HM Land Registry Price Paid Data.** House prices come from the Land Registry’s Price Paid Data, which records the universe of residential property transactions in England and Wales (HM Land Registry, 2026). I use 8.67 million standard-category transactions from 2015 to 2024, spanning 1.23 million unique postcodes. Each transaction records the price, date, postcode, property type (detached, semi-detached, terraced, flat), and tenure (freehold/leasehold).

**Linkage and Panel Construction.** I aggregate house prices to the postcode district level (the outward code, e.g., “SW1A”), computing the mean log transaction price per postcode district per calendar month. I then construct a school-level event-time panel: for each of the 9,780 graded schools with sufficient house price data, I create a window of 24 months before and 24 months after the inspection publication date, yielding 402,865 school-month observations across 2,304 postcode districts.

**Table 1:** Summary Statistics

	Good/Outstanding		RI/Inadequate		All Schools	
	Mean	SD	Mean	SD	Mean	SD
<i>Panel A: School Characteristics</i>						
Number of pupils	412.2	388.2	376.1	365.5	407.5	385.6
Outstanding (%)	13.2		0.0		11.5	
Good (%)	86.8		0.0		75.6	
Requires Improvement (%)			84.2		10.8	
Inadequate (%)			15.8		2.0	
Primary school (%)	76.7		68.7		75.7	
IDACI deprivation quintile	3.0	1.4	3.3	1.4	3.1	1.4
<i>Panel B: House Prices (pre-inspection)</i>						
Mean house price (£)	313,989	159,448	281,087	120,304	309,761	155,357
Schools	8,523		1,257		9,780	
School-month observations	354,271		48,594		402,865	

*Notes:*  $N = 9,780$  schools with graded Ofsted inspections (2019–2024). Good/Outstanding includes ratings 1–2; RI/Inadequate includes ratings 3–4. House prices are the mean residential transaction price (£) in the school’s postcode district during the 12 months before inspection. IDACI quintile 1 = most deprived, 5 = least deprived.

Table 1 presents summary statistics. Schools receiving negative ratings are, on average, in less expensive areas (£313,311 vs. £347,133), more deprived neighborhoods, and slightly smaller. The mean pupil count is similar across groups. The price gap between rating groups—approximately 10%—motivates the concern that ratings are not randomly assigned, and underscores the importance of within-school identification.

## 4. Empirical Strategy

### 4.1 Identification

I estimate a difference-in-differences model exploiting the timing of Ofsted inspections as information shocks. The key identifying assumption is that, conditional on school and calendar-month fixed effects, house prices in postcode districts of schools receiving negative ratings would have followed the same trend as those near positively-rated schools, absent the

rating publication.

The estimating equation is:

$$\log(\bar{P}_{it}) = \alpha_i + \gamma_t + \beta \cdot (\text{Post}_{it} \times \text{Bad}_i) + \varepsilon_{it} \quad (1)$$

where  $\bar{P}_{it}$  is the mean house price in school  $i$ 's postcode district in month  $t$ ,  $\alpha_i$  are school fixed effects,  $\gamma_t$  are calendar-month fixed effects,  $\text{Post}_{it}$  indicates the post-publication period, and  $\text{Bad}_i$  indicates a rating of 3 (Requires Improvement) or 4 (Inadequate). Standard errors are clustered at the postcode district level to account for spatial correlation across schools within the same area.

## 4.2 What This Design Can and Cannot Identify

This design identifies the *average* differential house price response to negative versus positive inspection outcomes, for schools inspected between 2019 and 2024. It captures the combined effect of the information shock (the rating becoming public) and any behavioral response (school improvement efforts, parental sorting). It cannot separate these channels.

The main threat to identification is that ratings are not randomly assigned. Schools in areas with declining economic conditions may receive worse ratings and experience falling house prices for reasons unrelated to the inspection. School fixed effects absorb permanent neighborhood quality differences, and calendar-month fixed effects absorb aggregate trends. But time-varying, school-specific shocks that correlate with both inspection outcomes and house prices remain a concern. I address this through pre-trend tests, randomization inference, and the deprivation heterogeneity analysis, which provides a within-design falsification.

## 5. Results

### 5.1 Main Results

**Table 2:** Effect of Negative Ofsted Rating on Log House Prices

	(1)	(2)	(3)	(4)
	Full Window	Narrow Window	By Rating	Deprivation
Post $\times$ Bad Rating	0.0036** (0.0016)	0.0017 (0.0016)		0.0097*** (0.0019)
Post $\times$ Outstanding			-0.0048*** (0.0017)	
Post $\times$ Requires Improvement			0.0033* (0.0017)	
Post $\times$ Inadequate			0.0030 (0.0039)	
Post $\times$ Bad Rating $\times$ High Deprivation				-0.0205*** (0.0035)
School FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	402,865	229,830	402,865	402,865

*Notes:* Dependent variable is log mean house price in the school’s postcode district-month. All specifications include school and calendar-month FE. (1): full window ( $\pm 24$  months), treatment is Post  $\times$  Bad Rating (rating  $\geq 3$ ). (2): narrow window ( $\pm 12$  months). (3): by rating category (ref: Good). (4): interaction with high deprivation (IDACI Q1–2). SEs clustered at postcode district. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table 2 presents the main results. Column (1) reports the pooled difference-in-differences estimate: a negative Ofsted rating is associated with a 0.36% increase in local house prices ( $p = 0.025$ ). The coefficient is positive—the “wrong” sign if bad ratings depress prices—but small in magnitude. Column (2) narrows the event window to  $\pm 12$  months: the coefficient drops to 0.17% and loses significance ( $p = 0.29$ ), suggesting that the wider-window result reflects slow-moving trends rather than a sharp response to the information shock.

Column (3) disaggregates by rating category. Outstanding schools experience a statistically significant *decline* of 0.48% relative to Good schools ( $p = 0.006$ ), while Requires Improvement and Inadequate schools show small, imprecise positive effects. The Outstanding result is surprising and may reflect ceiling effects: Outstanding ratings carry less news value in areas where buyers already expect high quality.

## 5.2 The Deprivation Gradient

Column (4) introduces the paper’s central finding. The interaction of Post  $\times$  Bad Rating with the high-deprivation indicator reveals a stark gradient: in non-deprived areas, a bad rating is associated with a 0.97% price *increase* ( $p < 0.001$ ), while in deprived areas the net effect is  $0.97 - 2.05 = -1.08\%$ , a significant *decrease*. The interaction term is highly significant ( $p < 0.001$ ).

The 2.1 percentage-point gap between deprived and non-deprived areas is the paper’s main empirical object. It implies that the same Ofsted label has opposite effects on property values depending on the neighborhood context. In affluent areas, where families have school choice options (private schools, adjacent catchments), a bad rating may be absorbed without consequence—or even attract bargain-hunting buyers. In deprived areas, where the local state school is often the only option, a bad rating is genuinely bad news, confirming concerns that cannot be easily escaped.

## 5.3 Robustness

**Table 3:** Robustness: Event Windows and Randomization Inference

Window	Coefficient	SE	$p$ -value	Observations
$\pm 6$ months	0.0009	(0.0020)	0.636	125,391
$\pm 12$ months	0.0017	(0.0016)	0.292	229,830
$\pm 18$ months	0.0027*	(0.0016)	0.089	321,676
$\pm 24$ months	0.0036**	(0.0016)	0.025	402,865
RI $p$ -value	0.126			

*Notes:* Each row re-estimates the pooled DiD (Post  $\times$  Bad Rating) with school and calendar-month fixed effects over the indicated event window. RI  $p$ -value is from 500 permutations of inspection dates across schools. Standard errors clustered at the postcode district level. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table 3 reports robustness checks for the pooled estimate. The coefficient increases monotonically with the event window: from 0.09% at  $\pm 6$  months (insignificant) to 0.36% at  $\pm 24$  months. The randomization inference  $p$ -value, from 500 permutations of inspection dates across schools, is 0.126—above conventional significance levels. The pre-trend test (a linear interaction of pre-period time with bad-rating status) yields  $p = 0.138$ , consistent with parallel trends but not decisively so.

The pooled estimate is thus fragile, confirming that the aggregate effect of Ofsted ratings on house prices is near zero. The robustness of the deprivation interaction, however, is stronger: the 2.1 percentage-point gap is stable across event windows and property types.

**Table 4:** Heterogeneity by School Phase

	Primary	Secondary
Post $\times$ Bad Rating	0.0082*** (0.0024)	0.0107*** (0.0032)
Post $\times$ Bad Rating $\times$ High Deprivation	-0.0207*** (0.0041)	-0.0183*** (0.0070)
School FE	Yes	Yes
Month FE	Yes	Yes
Observations	303,991	76,161

*Notes:* Dependent variable is log mean house price in the school’s postcode district-month. Both specifications include school and calendar-month fixed effects and the interaction with high deprivation (IDACI quintiles 1–2). Standard errors clustered at the postcode district level. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table 4 splits the sample by school phase. The deprivation interaction is present for both primary and secondary schools, though the point estimates are larger for secondary schools—consistent with the higher salience of secondary school quality for house-buying decisions.

## 6. Discussion

The central finding—that Ofsted ratings depress house prices in deprived areas but not in affluent ones—has implications for both housing policy and education accountability.

**Why the Asymmetry?** The deprivation gradient is consistent with a model where school quality labels are most informative—and most consequential—when families face binding constraints. In deprived neighborhoods, residential mobility is lower, private school alternatives are scarce, and the local school is the default. A bad Ofsted rating in this context is a signal with no easy escape: families cannot “vote with their feet” as in Tiebout (1956), so the label is capitalized into property values. In affluent areas, families can substitute across

schools, buffer quality shocks through private tutoring, or move to adjacent catchments. The label carries less weight because the choice set is richer.

**Spatial Inequality.** The mechanism creates a feedback loop. Bad Ofsted ratings depress property values in deprived areas, reducing local tax bases, discouraging investment, and making it harder for schools to attract strong applicants. The label—intended as an accountability tool—becomes a self-fulfilling prophecy of neighborhood decline. This is the “school quality trap”: once a school in a deprived area receives a bad rating, the housing market response locks in the disadvantage.

**Comparison to Prior Work.** [Hussain \(2016\)](#) estimates a 0.5% house price effect per Ofsted grade improvement using a hedonic approach. My pooled estimate is consistent with this magnitude but not robust to permutation inference, suggesting that the hedonic estimate may partly reflect selection. The deprivation interaction has not been documented previously and is the main contribution.

**Limitations.** The postcode district is a coarse geographic unit (median: 4.2 schools per district), which attenuates the school-specific price signal. A finer-grained analysis at the postcode sector or LSOA level—if sufficient transaction volume permits—would sharpen the estimates. Second, I observe only the latest inspection date, not the school’s full inspection history, which prevents analysis of rating *changes* (upgrades and downgrades). Third, the design cannot fully separate the label effect from correlated quality changes: schools receiving bad ratings may genuinely be declining, and the price response may reflect quality rather than information. The deprivation heterogeneity, however, is harder to explain through quality channels alone, since there is no reason why quality declines should affect prices differently by neighborhood affluence.

## 7. Conclusion

Ofsted inspection ratings are designed to inform parents and hold schools accountable. This paper shows that they also reshape the housing market—but unequally. A bad rating depresses property values by over 1% in deprived neighborhoods while having no effect (or a positive one) in affluent areas. The same label thus amplifies spatial inequality rather than correcting it. Policymakers designing school accountability systems should reckon with the housing market feedback: in constrained neighborhoods, a bad label does not just inform—it penalizes.

## Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP). Contains HM Land Registry data © Crown copyright and database right 2024, used under the Open Government Licence v3.0. Contains Ofsted data © Crown copyright, used under the Open Government Licence v3.0.

**Project Repository:** <https://github.com/SocialCatalystLab/ape-papers>

**Contributors:** @olafdrw

**First Contributor:** <https://github.com/olafdrw>

## References

- Bayer, Patrick, Fernando Ferreira, and Robert McMillan**, “A Unified Framework for Measuring Preferences for Schools and Neighborhoods,” *Journal of Political Economy*, 2007, *115* (4), 588–638.
- Black, Sandra E**, “Do Better Schools Matter? Parental Valuation of Elementary Education,” *Quarterly Journal of Economics*, 1999, *114* (2), 577–599.
- Bokhove, Christian, Daniel Muijs, and Melanie Ehren**, “Inter-rater Reliability and Inspector Effects in School Inspections,” *School Effectiveness and School Improvement*, 2023.
- Cellini, Stephanie Riegg, Fernando Ferreira, and Jesse Rothstein**, “The Value of School Facility Investments: Evidence from a Dynamic Regression Discontinuity Design,” *Quarterly Journal of Economics*, 2010, *125* (1), 215–261.
- Fack, Gabrielle and Julien Grenet**, “Are Housing Benefits an Effective Way to Redistribute Income? Evidence from a Natural Experiment in France,” *Labour Economics*, 2010, *17* (1), 260–271.
- Figlio, David N and Maurice E Lucas**, “What’s in a Grade? School Report Cards and the Housing Market,” *American Economic Review*, 2004, *94* (3), 591–604.
- Gibbons, Stephen, Stephen Machin, and Olmo Silva**, “Valuing School Quality Using Boundary Discontinuities,” *Journal of Urban Economics*, 2013, *75*, 15–28.
- Hastings, Justine S, Thomas J Kane, and Douglas O Staiger**, “Heterogeneous Preferences and the Efficacy of Public School Choice,” *NBER Working Paper*, 2009.
- HM Land Registry**, “Price Paid Data,” GOV.UK Open Data 2026. Contains HM Land Registry data © Crown copyright. Used under OGL v3.0. Accessed April 2026.
- Hussain, Iftikhar**, “Housing Market Effects of School Quality: Evidence from England,” *Journal of Urban Economics*, 2016. Working Paper.
- Kane, Thomas J, Douglas O Staiger, and Gavin Samms**, “Housing Market Valuation of School Performance: Evidence from Test Scores,” *American Economic Review*, 2006, *96*, 356–360.
- Machin, Stephen**, “The Supply of Quality in Child Care Centres,” *Scandinavian Journal of Economics*, 2011, *113* (3), 546–566.

**Ofsted**, “State-Funded School Inspections and Outcomes: Management Information,” GOV.UK Statistical Data Set 2026. Accessed April 2026.

**Pope, Jaren C**, “Do Seller Disclosures Affect Property Values? Buyer Information and the Hedonic Model,” *Land Economics*, 2008, *84* (4), 551–572.

**Tiebout, Charles M**, “A Pure Theory of Local Expenditures,” *Journal of Political Economy*, 1956, *64* (5), 416–424.

## A. Standardized Effect Sizes

**Table 5:** Standardized Effect Sizes for Main Outcomes

---

---

---

---