IFN Working Paper No. 1476, 2023

Free to Improve? The Impact of Free School Attendance in England

Marco Bertoni, Gabriel Heller-Sahlgren and Olmo Silva

Research Institute of Industrial Economics
P.O. Box 55665
SE-102 15 Stockholm, Sweden
info@ifn.se
www.ifn.se

FREE TO IMPROVE?

THE IMPACT OF FREE SCHOOL ATTENDANCE IN ENGLAND[*]

Marco Bertoni[†]; Gabriel Heller-Sahlgren[‡]; and Olmo Silva[§]

September 2023

Abstract: We investigate the impact of attending a free school in England – that is, a new start-up school that enjoys considerable autonomy while remaining in the state sector. We analyse the effects of two secondary free schools with different teaching philosophies: one follows a 'no excuse' paradigm, while the other one adopts a 'classical liberal', knowledge-rich approach. We establish causal effects exploiting admission lotteries and a distance-based regression discontinuity design. Both schools have a strong positive impact on student test scores on average. However, we also find heterogeneous effects: the 'no excuse' school mostly benefits boys, while the 'classical liberal' school mainly benefits White British and non-poor students. Both schools similarly reduce student absences and school mobility. Peer quality, teacher characteristics, and inspectorate ratings cannot fully explain the schools' effectiveness. Instead, a quantitative text analysis of the schools' 'vision and ethos' statements shows that the 'no excuse' and 'classical liberal' philosophies adopted by the two free schools clearly set them apart from the counterfactual schools where rejected applicants enrol, and likely explain their heterogeneous effects.

JEL Classification: I21; I24; I28.

Keywords: school autonomy; quasi-markets; free schools; achievement.

[*] We would like to thank the New Schools Network (NSN) for facilitating discussions with the two free schools that have shared data with us. Our findings represent our views and do not mirror or are influenced by the views of the NSN or the schools involved in our analysis. We are grateful to seminar participants at the University of New South Wales in Sydney, Ca' Foscari University of Venice, CEP Education Work-in-Progress meeting, CEP 2023 Annual Conference, CESIfo Effee 2022 Conference, EALE 2022 Annual Conference, IFS Education Seminar, SED 2022 Conference, SEHO 2022 Conference, SOLE 2023 Conference, Workshop on 'Human Capital, Labour Markets and Public Policy' in Naples, and Atila Abdulkadiroğlu, Gigi Foster, Björn Öckert and Marco Ovidi for comments and suggestions. We are also truly indebted to Filippo Boeri for helping us structure our text analysis. We are responsible for any error or omission. Office for National Statistics (ONS) Disclaimer: This work was produced using statistical data from ONS. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.
[†] University of Padova and IZA-Bonn.
[‡] Research Institute of Industrial Economics (IFN) and London School of Economics (LSE).
[§] LSE and IZA-Bonn. Corresponding author. Email at: o.silva@lse.ac.uk

1. Introduction

Over the past decades, policymakers in several countries – including Canada, Chile, England, Sweden, and the US – have promoted reforms that aim to improve school standards by creating quasi-markets in education. Common elements that underpin such market-oriented reforms include parental choice, school competition, operational autonomy, and entry of new providers.

In this paper, we focus on the latter element – i.e., the entry of new start-up schools – by analysing the impact of free schools in England. Introduced in 2010, free schools are newly established state-funded schools set up by teachers, parents, charities, and other non-governmental groups. Despite remaining part of the state sector, free schools enjoy significant autonomy and operate outside the control of the local government in terms of staff management and pay, length of school terms and school days, and pedagogical approaches.

Like charter schools in the US and Canada, and *friskolor* in Sweden, English free schools embody a policy agenda that seeks to liberalise the supply side of the education quasi-market – which is an understudied part of the English education reforms enacted over the past two decades.[1] Proponents of such reforms argue that entry of new providers enhances choice, stiffens competition, and promotes differentiation and innovation. But are start-up schools good at educating their own pupils?[2]

This paper seeks to answer this question through a quantitative case study of two free schools. Specifically, we focus on two high profile secondary free schools that were set up in the early days of the programme. The schools differ in terms of location – one is located in a mid-sized town in West Yorkshire and one in a relatively affluent London borough – as well as educational philosophies. Both institutions adopt fundamentally 'traditional' schooling and instructional models, but their high-level philosophies differ in many respects. While the West Yorkshire school adopts a pedagogical approach that is openly inspired by the 'no excuse' (NE) paradigm to which many US charter schools subscribe – characterised by a focus on high-expectations, routines, and discipline – the London school adopts a 'classical liberal' (CL), markedly knowledge-based and teacher-centred educational philosophy that draws its material and methods from both the humanities and the sciences.

[1] The impacts of choice, autonomy and competition have been investigated in the English context. See for example Gibbons et al. (2008); Eyles and Machin (2019) and Bertoni et al. (2020). Similarly, there is a large body of literature on these issues that focusses on US and Swedish schools. See amongst others Hoxby (2000), Cullen et al. (2006), and Böhlmark and Lindahl (2015).

[2] Unlike for England, there is a sizeable literature on the impact of US start-up charter schools (summarised in Cohodes and Parham, 2021). This includes Angrist et al. (2010), Abdulkadiroglu et al. (2011), Angrist at al. (2013), Dobbie and Fryer (2013), Fryer (2014) and Dynarski et al. (2018).

The schools also use different tie-breaking procedures for admission purposes. The NE school combines lotteries with fair banding – random draws among pupils with the same ability profile, as determined by a test at the point of application – to achieve a mixed intake. Conversely, the CL school uses lotteries for children grouped at different distances from the school premises, but also reserves some seats for pupils who live nearby, thus creating the possibility to use a distance-based regression discontinuity design (RDD) to study the impact of the school on achievement. We exploit the inner workings of the assignment mechanisms for causal inference, adapting to our context the insights from the recent 'Research Design Meets Market Design' (RDMD) methods introduced by Abdulkadiroğlu et al. (2017, 2021) to deal with the multi-staged nature of these mechanisms.

To carry out our analysis, we use pupil-level data for two cohorts of applicants to each of the two schools. We have access to schools' admission criteria as well as data covering pupil applications and enrolment. We match these data to pupil- and school-level background information as well as outcome variables from various administrative datasets. Finally, we collect novel data on the pedagogical approaches used by the two free schools, as well as the counterfactual schools attended by pupils not admitted by the free schools, by web-scraping information on the 'vision and ethos' statements posted on the schools' websites.[3] Our analysis in this domain is guided by two full-day visits at the free schools and by semi-structured interviews carried out with the schools' headteachers and founders.

Our results can be summarised as follows. First, we find that our research design leveraging lottery and RDD risks delivers a sample of successful applicants that is comparable to unsuccessful applicants in terms of a broad set of pre-determined individual-level covariates, including gender, ethnicity, free-school-meal eligibility (a proxy for income), and test scores at primary school. This corroborates the internal validity of our quasi-experimental setup.

Second, we estimate that, on average, the two free schools have a large, positive, and statistically significant impact on exam scores at the end of secondary education (comprising of five years of schooling). An additional year spent in either of the two free schools improves test scores by roughly 8% of a standard deviation. In other words, pupils who spend all five years of secondary education in the free schools perform about 40% of a standard deviation better than pupils in the counterfactual schools. This effect is similar across both cohorts and

[3] There are 52 counterfactuals for the NE free school, and 96 for the CL one. We also collect this information for all 102 secular, secondary free schools that had opened by the end of our analysis period, in order to compare our two free schools' educational approaches to those of other free schools. Every school in our analysis has a website and has posted a 'vision and ethos' statement.

schools. However, the results mask heterogenous effects depending on pupil background. The NE school primarily benefits boys and appears better at improving test scores among low-achieving pupils; conversely, the CL school primarily benefits White British and non-poor pupils and – to some extent – appears better at improving test scores among high-achievers. These patterns are expected given the schools' different pedagogical approaches and given the evidence from previous research on US charter schools. Among others, Angrist et al. (2010), Angrist at al. (2013), and Dobbie and Fryer (2013) find that a 'no excuse' approach is more effective for disadvantaged boys. On the other hand, Dynarski et al. (2018) find that schools in the for-profit National Heritage Academy chain are more effective among non-poor, non-urban students. While also employing traditional 'no excuse' practices, the network's inspiring principles resemble those adopted by the CL school – including a focus on student 'character' and a curriculum modelled on Plato's cardinal virtues (e.g., wisdom, self-control, gratitude, perseverance, courage, and respect).

We also find that both free schools impact non-test score outcomes (see Jackson, 2018), decreasing pupils' school mobility and unauthorised absences, but we only find limited evidence of increased use of disciplinary sanctions, such as temporary exclusions. Furthermore, the impact on absences and mobility is not as clearly heterogeneous as the effects on achievement – and not in ways that mirror the differential impact on end-of-secondary school test scores. This indicates that simple behavioural mechanisms that operate by increasing educational continuity (i.e., fewer absences and school changes) cannot fully account for our results.

Third, we compare the two free schools and their counterfactuals along several 'standard' school characteristics – such as peer quality, teacher characteristics, and inspectorate ratings – but find that differences along these dimensions cannot conclusively explain the free schools' effectiveness. For example, the NE school has a lower pupil/teacher ratio, and more qualified but less well-paid teachers, than the counterfactual schools – while the opposite is true for the CL one. Similarly, the NE school receives inspectorate ratings that are significantly better than those of counterfactuals across all domains – namely, management, pupil behaviour, teaching and learning, and achievement. On the other hand, the CL has worse inspectorate reports, except in the behavioural domain. To dig deeper into this issue, we study whether the characteristics of the two free schools differ from those of counterfactual schools in ways that are heterogeneous and mirror the differential effect on pupils' test scores. We broadly find that this is not the case, suggesting that the extensive range of 'standard' school characteristics we

investigate can neither account for the free schools' overall effectiveness nor for their heterogeneous impact among pupils from different backgrounds.

Finally, we conclude our investigation by carrying out a quantitative text analysis on web-scraped 'vision and ethos' (VE) statements to compare the educational approaches used by the two free schools to those of the counterfactuals in our sample. To do so, we first create a taxonomy that characterises the approaches used by the two schools on the basis of: *i*- key words identified in the VE statements – including schools' 'core values' and 'drivers'; and *ii*- salient elements that emerged during our visits – including from our semi-structured interviews, discussions with students, and experience of class time. We then create a 'dictionary' that allows us to associate such key words to a wide range of synonyms and akin expressions in the VE statements of counterfactual schools. Finally, we create some synthetic measures that capture whether a key concept is more present at one of the two free schools than what would be expected by looking at the overall set of VE statements of the counterfactuals.

We find that the two free schools share similarities with counterfactual schools along several domains. For example, the NE school's focus on good manners, autonomy, and purpose in learning is not more heavily influencing its approach than among counterfactuals. Similarly, the CL school's focus on high standards, high aspirations, and hard work can be similarly found in the schools attended by its rejected applicants. However, what clearly stands out relative to the counterfactual schools are the two high-level philosophies that permeate every aspect of the schools' approach to teaching and learning. The NE school focusses on a 'no excuse, no shortcut' approach to education, based on strict routines, a culture of success, and hard work. The CL instead focusses on a 'classical liberal' education, centred around a knowledge-rich curriculum, teacher-led learning, and an openly competitive atmosphere. While exposure to the free schools' educational philosophies does not vary by pupil background, the philosophies' effectiveness is likely to differ in ways congruent with the heterogeneous effects we find on achievement. We therefore conclude that the differential impact of the two free schools is most likely explained by the way they structure teaching and learning.

Our findings and approaches are novel in the setting under investigation. First, from the methodological point of view, little research in developed-country contexts outside the US has exploited admissions lotteries and/or the insights of the RDMD agenda to study school effects on pupil achievement. Similarly, the use of text analysis to characterise schools' educational philosophies is new. In this respect, we share some similarities with Biasi and Ma (2023), who focus on the innovative content of taught courses at the university level. Second, although there is research on US charter schools (reviewed in Cohodes and Parham, 2021), Swedish *friskolor*

(e.g., Böhlmark and Lindahl, 2015), and English academies (see Eyles et al., 2017; Eyles and Machin, 2019; Bertoni et al., 2020), there is no causal evidence on the effectiveness of the original start-up English free schools. Machin et al. (2020) study the effectiveness of University Technical Colleges. Although these fall within the free-school sector, they only represent a late expansion of the initial policy and cover a subset of undersubscribed, quasi-vocational schools. Most other research on free schools analyses their effects on social segregation or is descriptive only (see Green et al. 2015; Andrews and Johnes, 2017; and Allen and Higham, 2018).

Our paper contributes to the literature on the effects of quasi-markets in education by providing credible causal evidence on whether bottom-up initiatives that expand school supply can boost pupil performance. The answer in our context is a 'qualified yes': the two schools we investigate have strong causal effects on pupil test scores. These findings provide an important piece of evidence for ongoing policy discussions on the organisation of public education systems: they suggest that the state does not necessarily need to set up and manage schools directly, as school services can be successfully provided by independent entities. In such a system, the role of the state is mostly to determine the 'rules of the game' in terms of common curricula and non-discriminatory admission criteria; to fund schools on the basis on their ability to cater for demand; and to monitor school operations and pupil outcomes.

Of course, a key concern in this regard is whether our findings can be extrapolated to the rest of the free-school sector. This is far from clear for at least two reasons. First, we only study the impact of two schools with quite distinct pedagogical approaches. Our quantitative text analysis applied to the other free schools that existed during the period of our investigation reveals that a non-negligible proportion adopted the pedagogical approaches embodied by the NE school, but very few embraced the principles of the CL school. Second, our findings show that the different approaches used by the two free schools produce effective teaching among a sub-set of pupils whose characteristics 'match' the schools' educational philosophies – stated differently, they represent pupil-to-school match-specific gains.[4]

Evidence on such match-specific gains can provide valuable insights when interpreted through the framework of an education quasi-market with differentiated suppliers. Indeed, the school-choice literature posits the existence of gains from differentiation because a heterogeneous education supply can improve the matching process between pupils' pedagogical needs and school offer – which is supposed to improve learning. Our evidence

[4] A counterpoint to this conclusion is the evidence in Fryer (2014), who shows that injecting charter school practices inspired by a 'no excuse', high-expectations, high-dosage pedagogical approaches in state schools in Chicago, Denver, and Houston can significantly improve education standards.

suggests that these match effects can materialise in the context of start-up schools that bring to the market new and distinctive approaches. An obvious drawback of such diversification is the likely increased segregation of pupils with different background in different schools. This sorting can be problematic if it does not represent active parental choices based on expected match-specific benefits that schools can impact on children – and, in this respect, a lack of information may make it difficult for parents to take advantage of the opportunities provided by education quasi-markets. These considerations lead us to conclude that for free schools to work as effectively as possible, and become a 'tide that lifts all boats', it is important to pursue policies that make the matching process between pupils and schools as efficient as possible. In the context of a large-scale experiment in Chile, Arteaga at al. (2022) show that provision of personalised information about the likelihood of admission to specific schools can affect families' search efforts and significantly improve matching efficiency. In our context, personalised information about the likely benefits of schools with differing educational approaches for students with different characteristics could lead to a better matching process, thus improving overall education standards.

2. Institutional setting

2.1 Education stages and main features of the English education system

Compulsory education in England is divided into primary and secondary schooling. The former covers pupils aged 5–11, while the latter covers pupils aged 11–18. The schooling system is further organised around Key Stages (KS). Pupils normally enter school at the Foundation Stage (ages 4–5 or grade 0), then move on to KS1 (ages 5–7; grades 1–2), and progress to KS2 (ages 7–11; grades 3–6), which marks the end of primary school. After KS2, pupils move to secondary school (ages 11–12; in grade 7) where they progress through KS3 (ages 11–14; grades 7–9), and KS4 (ages 15–16; grades 10–11). Compulsory schooling ends when pupils are 16 years old, which generally coincides with the end of KS4.

During their education, pupils sit several assessments. At KS1, pupils are assessed in English and Mathematics. KS1 exams are externally set but internally marked by teachers. At KS2, pupils take standardised national tests in English, Mathematics, and Science, which are externally set and assessed. At KS4, pupils sit academic (GCSEs) and/or vocational (NVQs) tests in a range of subjects, with English, Mathematics, and Science being compulsory. These tests are externally set and assessed. School-average performance in these tests is published in league tables, alongside other characteristics, such as school size and pupil composition. These tables are salient in the media and are routinely used by parents to inform their school choices.

Further information on school quality is disseminated through the publication of ratings by the school inspectorate, Ofsted. Ofsted visits schools every three to five years and the inspections result in publicly available reports rating schools from 'outstanding' to 'inadequate', both overall and in different domains (such as teaching, management, and pupil behaviour). Although Ofsted is a government department, its inspections are carried out, and its reports published, independently of the government.

2.2. Quasi-markets for education and autonomous schools in England

Quasi-markets for education have been in place in England since the 1988 Educational Reform Act, which abolished the automatic assignment of children to schools based on residency and gave parents the right to choose schools. It also linked school funding more closely to the number of pupils enrolled, thus providing incentives for schools to compete for pupils. Under this system, admission to state education is based on parental preferences, constrained by the fact that popular schools are often oversubscribed. When this occurs, various criteria are used to prioritise applicants, usually favouring those who live nearby, those with special educational needs or in care of the Local Authority (LA) in question, and those with siblings in the school. Certain types of schools can prioritise applicants using other criteria (e.g., faith schools are allowed to select pupils on basis of religion). Finally, a small proportion of secondary schools, so-called 'grammar schools', select pupils using entrance exams. Depending on the LA in which families live, they can apply to between three and six schools. To allocate pupils to their preferred schools, LAs run constrained versions of the Gale-Shapley student-optimal stable mechanism, also known as a Deferred Acceptance (DA) algorithm.

State-funded school types originally included community schools, voluntary-controlled, foundation, and voluntary-aided schools. Community and voluntary-controlled schools are managed by the LA, which employs the staff, owns the buildings, and handles admissions. Voluntary-aided and foundation schools enjoy more autonomy from the LA, which nevertheless retains powers of oversight. In all these cases, funding comes from the LA using money provided by central government through general taxation.[5]

The transition towards an education quasi-market accelerated in the 2000s with the introduction of academies. Early academies were promoted by the Labour government with the aim of replacing failing traditional LA-controlled schools. These 'sponsored' academies were mostly managed by a government-approved sponsor – usually a charity or a business group –

[5] On top of this, there is also a small fee-charging private school sector, enrolling roughly 5% of pupils in total.

that would take over the school and turn it around by leveraging increased autonomy in terms of staff employment, pay and working conditions, and pedagogical ethos. The programme was revised and expanded when the Liberal-Conservative government enacted the 2010 Academies Act, aimed at fostering differentiation, innovation, and competition in the education system. This legislation introduced a new type of 'converter' academies. Converter academies are state-funded schools that have been rated 'good' or 'outstanding' by Ofsted, which makes them eligible for academy status without the need to associate with a sponsor. Unlike other forms of state-financed schools, academies are funded directly by the central government without any LA acting as an intermediary.

The 2010 Academies Act also created the legal basis for 'free schools' – the type of school analysed in this paper. Unlike academies, free schools are newly established state-funded schools, run not-for-profit by parents, charities, teachers, faith groups, businesses, and other groups. The rationale for their introduction was to expand the supply-side of education by allowing new actors to promote bottom-up, start-up initiatives, thus increasing choice and competition in the system. Once established, free schools are identical to academies from a legal standpoint and therefore operate in a similar fashion. Like academies, free schools enjoy considerable freedoms in terms of teachers' pay and conditions, the length of terms and school days, curriculum choices, pedagogical philosophy, ethos, and admission criteria.[6] To set up a free school, proponents need to either demonstrate the need for additional school places in a local area, or the need for 'high quality' places. As a result, the establishment of free schools is not only related to demographic needs. The governance of free schools is overseen by a trust, a not-for-profit company that formally employs the staff and is responsible for school performance. Trusts can run a group of free schools and/or academies – or include both – and are then labelled multi-academy trusts (MATs). Like academies, free schools are funded directly by the central government rather than via the LA.

While initially representing a small-scale 'revolution', the free-school programme grew quite rapidly. From only consisting of 24 schools in 2011, the sector expanded to include nearly 270 schools by the end of the Liberal-Conservative government in 2015 – an eleven-fold expansion in four years.[7] Since then, the expansion of the programme has flattened out: in 2021, there were around 600 free schools, a doubling of the sector in more than six years, with a higher incidence in the London area.

[6] Like all state-funded schools, free schools must abide by the non-discriminatory rules set by the national admission code.
[7] There are approximately 15,000 primary schools and 3,200 secondary schools in England.

The actors involved in setting up schools also changed over time. Initially, most free schools were set up by parents and teachers, who were perceived by many as 'disrupting innovators' and 'pioneering educators'. In later years, an increasing number of free schools have instead been set up and run by MATs.[8]

Consistent with the aims of the policy, the free-school sector is heterogeneous and includes later sub-categories, such as University Technical Colleges, which are hybrid schools providing a blend of general and vocational education (often specialising in engineering subjects), and Studio Schools, which have close links with businesses and use a project-based learning approach. We do not study these types of free schools in this paper.

Instead, this paper analyses the impact on achievement of two secondary free schools that were examples of the early days of the programme. The schools opened in the first and second years of the programme, following the enactment of the 2010 Academies Act. We now describe these schools' key features.

2.3. The 'no excuse' school[9]

The first school we analyse (Dixons Trinity Academy) is located in Bradford, a deprived mid-sized town in West Yorkshire with the second-highest incidence of South Asian British residents. The school opened in September 2012 as part of the second cohort of free schools, which included just under 50 primary and secondary free schools throughout England. The school was one of only four free schools operating in the area at the time of its establishment. The school initially admitted 112 pupils per cohort and has always been highly oversubscribed.[10] It is supported by the Dixons Trust, a MAT operating in Bradford and Leeds that focusses on 'making a difference' for young people affected by socio-economic disadvantage.

The school adopts an unashamedly 'no excuse' (NE) approach. Quoting from their webpage, its philosophy reads: 'We promise to do whatever it takes to ensure that every student (…) achieves their full potential. We have extremely high expectations, and *just as there are no shortcuts, there are no excuses.* We (…) have a sentence that states the lasting impression we want to leave on the world. Our academy sentence is: "The academy ensured that all students succeeded at university, or a real alternative, thrived in a top job and had a great life."'

[8] See https://www.theguardian.com/education/2021/feb/09/michael-gove-free-schools-at-10-successful-policy-since-the-war-or-mistake

[9] Although the two schools have agreed for their identity to be disclosed, we will refer to them using their pedagogical approach throughout the paper. This is more informative than the actual schools' names.

[10] The school mostly advertised itself through word-of-mouth and posts on social media, such as Facebook. Prior to opening, the founders also used a consultation period to raise the school profile in the local community.

The school was the first free school to be rated outstanding by Ofsted in 2014. The inspectorate's report highlights the high-aspiration, disciplined, and result-driven philosophy championed by the school.

We visited the school in late June 2022 and met with both the headteacher and the CEO of the Dixons Trust (in this case, the school founder). Our meeting centred around a semi-structured interview and the questionnaire we prepared is presented in Appendix Exhibit 1. During our discussions, we avoided as much as possible to reveal the details of our results on effectiveness and especially the patterns of heterogeneity we found. We also avoided prompting to minimise answers that would *ex-post* rationalise our *a priori* and findings. The interview lasted for approximately 1.5 hours and was followed by an 'experiential' tour of the school, including lunch in the canteen with the pupils and some discussions with the Year-7 and Year-8 head boy and head girl.

The school presents itself as a new, purpose-built construction in a deprived and markedly diverse part of Bradford. Banners hung outside the school gate highlight the school philosophy encapsulated in the 'no excuses' and 'in this academy only excellence will do' sentences – see Appendix Exhibit 2.

Our interview highlighted similar features. The school approach to study is guided by three core values – hard-work, trust, and fairness – and pupils are told to follow three drivers: mastery, autonomy, and purpose. A key metaphor used by the school to describe learning and organise activities is that of 'climbing the mountain', highlighting how achievements require hard-work, devotion, and a long-term perspective. In line with this, the three schoolhouses are named after three different mountains (Makalu, Aconcagua, and Pelvoux) with the following three mottos: 'no excuses', 'no shortcuts', and '100% every day'. Lists on the school corridors rank the 'best climbers' of the week, based on behaviour and progression. The three schoolhouses compete with one another and receive rewards and acknowledgments for their success. However, these rewards are deliberately small and low key as the aim of the school is to inspire passion for learning and long-term goals, rather than short-term recompenses. The headteacher and CEO clarified that these pedagogical choices were influenced by research trips taken to visit US charter schools. In particular, the US Knowledge Is Power Program (KIPP) network was mentioned as one of the inspiring models.

Both the headteacher and the CEO emphasised the importance of strict routines that guide school life. Understanding and adhering to such routines is key for both pupils and teachers, and formal induction processes are in place. Prospective pupils are invited to a meeting at the school before they apply and then again upon offer to clarify the school practices and principles.

Furthermore, the first week of enrolment is not devoted to formal learning, but to forming an understanding of the culture and the ethos of the school. Throughout their education, pupils who 'forget' values and practices go through a process of 're-induction' via individual or group sessions, and the school has a dedicated re-induction room.[11] Nonetheless, both interviewees emphasised that the militaristic, 'masculine' and harsh approaches taken by some US charters find no place at the school. Instead, they described their approach as 'maternal'.

Importantly, both leaders emphasised the importance of commitment to routines and core values also for staff members. Like pupils, new recruits go through a process of induction – and a process of re-induction when their practices deviate from expectations. To make sure that the pedagogical philosophy is not diluted, the school uses no supply or temporary teachers. Instead, gaps left by staff absences, periods of sick leave, and even maternity/paternity breaks are covered by colleagues, who put in the extra time to preserve the integrity of the school's educational approach. When discussing the teaching approach deployed by the school, we understood this is not focussed on drill or rote learning, nor follows a set of strict guidelines provided by the central management of the trust or senior staff at the school. Teachers are empowered to carry out some 'experimentation' if they wish to do so, although this should not disrupt or undermine the core routines and key values of the school.

Finally, to appeal to a community that might have professional-oriented aspirations for their children, the curriculum is not narrowed down to a limited number of academic topics. However, there is a clear ambition to 'do the key things well' (meaning the core subjects) – and to send as many children as possible to university.

2.4. The 'classical liberal' school

The second school in our study (West London Free School) is located in a West London inner borough. This spans both some wealthy areas and fairly-deprived neighbourhoods, which have a predominantly White population, including White British as well as many European migrants (the largest groups being French, Polish, and Portuguese). The school opened in September 2011 and was part of the very first batch of 7 secondary free schools. It was the only free school operating in the borough for a long time (and during the time of our analysis). The school had a capacity of 120 pupils per cohort and was oversubscribed from the beginning.[12]

[11] Strict routines and behavioural practices guide every aspect of school life, including walking down the corridors, queueing for food in the canteen, and tidying up afterwards.

[12] To raise the visibility of the school and attract parental demand, the founders held open days, visited the local primary schools, and 'banged the drum' in the media (one of the founders is a well-known journalist).

The school was founded by a group of high-profile promoters believing in a traditional, knowledge-based approach to education, and was supported by many teachers and parents. Quoting from the school's webpage, it adopts a 'classical liberal' (CL) approach to education, that is 'a (…) rich education that draws its material and methods from the best and most important work in both the humanities and the sciences. The aim (…) is to prepare children to participate in (…) debates about contemporary issues, as well as the universal questions that have been troubling mankind (…). *We want pupils to leave our school with the confidence that comes from possessing a treasure trove of essential knowledge*, and a deep and lasting respect for reason, evidence, civility, honesty, kindness and the value of hard work and self-discipline.'

The school was inspected by Ofsted in 2013 and 2017 and, in both instances, it was rated as good with outstanding features in the domains of pupils' behaviour and safety. The reports confirm the 'classical liberal philosophy' advertised by the school.

We visited the school in early July 2022 and met with the lead headteacher. Our semi-structured interview was centred around the same questionnaire we used for the NE school and lasted just above one hour. As for the NE school, we avoided prompting and disclosing too many details of our evidence to avoid influencing our interviewee's answers. We also had an opportunity to briefly talk to the joint headteacher and were taken for a tour of the school. As students were away on fieldtrips, we did not have an opportunity to talk to the pupils. Therefore, we visited the school again in September 2022 and observed lessons.

The school is housed in a mid-19th century residence (see Appendix Exhibit 2) with a modern annex at the back. The historical nature of the building provides an appropriate background to the classical liberal ethos of the school. Although according to the headteacher, the classical liberal label is 'somewhat dated', everything in the school's visual impact betrays this classics-inspired philosophy.

The headteacher described the learning approach of the school as centred around a 'knowledge-rich curriculum' and a 'specify and excel' (i.e., narrow the curriculum, but deepen the level of understanding) philosophy. Its three key values – kindness, hard-work, and high standards – overlap to some extent with those of the NE school. However, in the case of the CL school, the notions of high standards and aspirations are more clearly oriented towards academic excellence and the study of classic and core subjects. Indeed, Latin is compulsory for all students up to the third year of secondary education; the study of arts is focussed on the classics throughout history; music is a well-renowned specialism of the school; and areas that are considered not challenging enough (such as 'drama' and 'home economics') are not offered

at the school.[13] The school also places a big emphasis on debating and public speaking, and on providing students with the skills that equip them to join discussions on contemporary issues as well as universal questions about society.

The classical liberal foundations are further emphasised in the names of the four schoolhouses – Athenians, Corinthians, Olympians, and Spartans – and all teaching rooms are named after classic and influential scientists or thinkers (e.g., Euclid, Pythagoras, and Turing). Like at the NE school, the schoolhouses compete with one another in terms of pupil achievements and behaviour. But unlike the NE school, the CL school openly values competition and aims to foster a competitive atmosphere.

As for the NE school, the CL one places great emphasis on discipline, routines, and outstanding behaviour. Indeed, the headteacher clarified that 'liberal does not mean lax' – and portrayed the school as 'non-permissive'. Although we were not made aware of specific processes for pupils' inductions, it is clear that pupils are expected to adhere to a set of well-codified behavioural rules – including how to walk in the corridors, how to address teachers, and how and when to talk to one another.

The latter point is particularly relevant during classroom time, which is always centred around teacher-led lessons. Although the school promotes debating, lessons are not the time for honing this skill: teachers talk, while pupils listen and respond to teacher questions. Most discussions are thus between the teacher and individual pupils. Instead, there is little room for pupil-to-pupil 'chatting' in the school's approach to lessons, apart from during very short periods (about one minute) following a prompt by the teacher. As the headteacher emphasised, classes are physically designed in a way that fosters teacher-to-pupil interactions: all desks and chairs are oriented in the same direction, making sure pupils face the teachers and whiteboard.

In terms of its staff, the headteacher reported that the school has had a 'healthy amount of turnover', meaning that staff members who did not endorse and commit to the school ethos left (rather than stay and dilute the core pedagogical values). The headteacher also emphasised the advantage of originally setting up as a brand-new free school in terms of selecting only teachers and personnel that shared the classical-liberal orientations of the original founders. In short, it was clear from our visits that teachers are immersed in a well-defined culture and provided with quite clear guidance about what it means to work at the CL school.

[13] The headteacher reported that one of the reasons why the school does not achieve 'outstanding' Ofsted rating relates to its limited subject offer, which the inspectorate sees as a limitation from the point of view of inclusivity. There were no suggestions the school would change its approach simply to improve its ranking.

3. Admission rules and identification

3.1. General considerations

Pupils applying to a free school are unlikely to be representative of the pupil population, as self-selection is likely based on parental expectations about the gains from free-school attendance, which are unobservable to the econometrician. As a result, simple comparisons of pupil outcomes across applicants and non-applicants are more informative about selection patterns than the causal effects of attending free schools. Furthermore, even within the group of free-school applicants – who manifest similar preferences for such schools – comparing outcomes among successful and unsuccessful applicants may yield biased estimates, since the admission criteria can induce a correlation between applicants' ability and their likelihood of being admitted. This is obvious when schools select pupils on the basis of prior ability, but it also applies to other forms of admission priorities, such as proximity to residence.

Most standard approaches developed in the literature to solve such issues are unfeasible or unlikely to work in our setting. Since free schools are newly established schools, difference-in-differences and grandfathering strategies (see Abdulkadiroğlu et al., 2017) cannot be used, while the central role played by unobservable parental expectations and preferences weakens the credibility of selection-on-observables (matching) approaches to causal inference.

Instead, as both schools under investigation have been highly oversubscribed since opening, we leverage data on applications and detailed information about the way in which the schools prioritise applicants in case of oversubscription to design an identification strategy. To solve issues related to the multi-staged nature of the assignment mechanisms, and to identify pupils who face an identical 'risk' of being admitted to the schools, we borrow from recent advances in the RDMD literature developed by Abdulkadiroğlu et al. (2017) and Abdulkadiroğlu et al. (2021).

Before detailing the admission rules for each free school, it is worth noting that both schools managed their admissions independently in their first year of operation, while they joined the LA assignment-mechanism from the second year onwards.[14] As a result, we face an additional empirical challenge when dealing with the second cohort, as discussed below.

[14] Free schools are allowed to act as their own admission authority in the first year of operation, but they must join the LA process from the second year of operation onwards. However, the timing of the application to and offers from free schools always overlapped with that of LA assigning procedures.

3.2. Admission rules and identification - NE school

In case of oversubscription, pupils looked after by the local authority (LAC) and pupils with special educational needs (SEN) are given priority for school places, and pupils with siblings who already study at the school are next in line. All other applicants fall within the same priority group, within which seats are allocated using so-called fair banding. Unlike in a serial dictatorship, where the best applicants are prioritised, the fair-banding process is used to guarantee that the ability distribution in the pool of admitted pupils mirrors the one observed among applicants. This system is implemented by means of a random ballot within ability bands, determined by the results on a non-verbal reasoning test completed by all applicants at the time of application. Applicants are divided into nine bands – called 'stanines' – and the thresholds to define each band are defined *ex-ante*. If there are more applicants than places available for the different bands, offers are made randomly within each band so that the number of pupils admitted mirrors the ability distribution among applicants.[15] Within each band, applicants with the highest lottery number are given priority. Importantly, the lottery number is unconditionally and randomly drawn for each pupil at the time of application.

In practice, unless they are classified as LAC or SEN, or have siblings at the school, applicants are at the same risk of being admitted within their band, and the lottery ensures that assignment within band is random.[16] As a result, the inclusion of stanine-fixed effects (cohort-specific) in our models implies that we only compare pupils who share the same (random) likelihood of admission when studying the effects of attending the school.

While this is sufficient to draw causal conclusions for the first cohort of applicants, when the school managed its admission process autonomously, this is not the case for the second cohort of applicants. For this cohort, we still observe all applications, and the school uses the same admission criteria. However, the school was then part of the broader assignment mechanism in the LA, which uses a DA algorithm to allocate pupils. Using this algorithm, applicants are considered for a seat at a school only if they are not offered a seat from a school for which they have a stronger preference. As a consequence, only applicants who did not receive an offer from a school which they prefer compared to the NE school may receive an

[15] Pooling across cohorts, we find the following stanine distribution: 8% of applicants fall in band 1; 12% in band 2; 16% in band 3; 17% in band 4; 15% in band 5; 14,5% in band 6; 6% in band 7; 6% in band 8; and 4% in band 9. Offers were made proportionally and there are no marked differences across cohorts in this respect. All stanines were oversubscribed.

[16] No pupil was admitted based on the LAC and SEN criteria in our data, or, for obvious reasons, on the basis of the sibling rule in the first cohort. Only a handful of children were given priority due to the sibling rule in the second cohort. As our identification strategy hinges on the combination of banding and random allocation, these pupils are excluded from our analysis.

offer from the NE school. As a result, the set of pupils who receive an offer is endogenously 'truncated' depending on their priority at other schools that they rank above the NE school.

Still, this preference-based truncation is not a function of applicants' lottery numbers in the NE school's admissions process. The NE school assigns these numbers to each applicant irrespective of his or her preferences for the NE or other schools, which the school does not observe, and before returning the data to the LA, which then runs the DA matching algorithm. At the same time, higher lottery numbers give applicants higher admission chances at the NE school within stanines. As a result, for the second cohort, we bypass the problem of endogenous truncation by using applicants' lottery numbers as an instrumental variable (IV) for the probability of receiving an offer at the school, conditional on stanine-fixed effects.

3.3. Admission rules and identification - CL school

In case of oversubscription at the CL school, priority is given in the following order: LAC and SEN children, pupils with a musical aptitude (10% of places), and siblings of children who are already attending the school.[17] Next, 50% of the remaining seats are offered to pupils on the basis of residential proximity. Once these pupils are assigned, two-thirds of the remaining seats are allocated using a lottery among pupils who live within a three-mile radius from the school, and one-third of the remaining seats are offered using another lottery among pupils who live between three and five miles from the school. Any remaining places are then assigned via a lottery among pupils living beyond five miles from the school.

In our set-up, we focus on applicants who fall within either the distance or the 'lotteries-within-distance' categories. First, for pupils living within three miles from the school, the risk of being admitted to the school is determined by a distance-based regression discontinuity design (RDD) as well as by a lottery among those who are not admitted through the distance mechanism. Thus, pupils living within three miles of the school are in principle at risk of being admitted via both the distance and the lottery categories, and we need to account for the fact that the risk of entering the lottery is conditional on pupils' outcome in the distance category. Second, for pupils in the 3–5-mile distance category, the risk of being admitted to the school is only determined through a lottery.[18] The last admission criterion is instead inactive: due to

[17] In our data, no pupils are assigned based on the LAC and SEN criteria, and only a handful of pupils in the second cohort are assigned based on the sibling rule. These pupils are excluded from our analysis. Furthermore, in both cohorts, we observe that fewer than the reserved 10% of pupils are assigned on the basis of musical aptitude. Since pupils admitted on the basis of musical aptitude are not assigned through the mechanisms exploited to obtain random variation in free-school attendance, we also exclude them in the analysis.

[18] Pupils who live beyond three miles of the school are never admitted through the distance criterion.

heavy oversubscription, pupils living beyond five miles from the school were never offered a seat in the period of analysis and are dropped from the sample.

To design an identification strategy for this complex setup, we borrow heavily from the RDMD literature. We start with the distance-based RDD: this generates, *ex-post*, a distance cut-off τ that is *ex-ante* unknown to parents. However, pupils who live very close to the school are always admitted and are not 'at risk'. They are therefore dropped from the estimation sample. Similarly, pupils who live too far away from the school are not going to be admitted due to residential proximity – so they are also not at risk of being admitted via the distance criterion. Only applicants who live 'close enough' to the τ distance cut-off are at risk of being admitted through the RDD. The key empirical concern is to determine how close is 'close enough'. We follow Abdulkadiroğlu et al. (2021) and use an optimal bandwidth estimation strategy to determine the level of the bandwidth δ that determines which pupils who are at risk (see Calonico et al., 2014). According to the results in Abdulkadiroğlu et al. (2021), pupils within $[\tau-\delta, \tau+\delta]$ share the same 50/50 risk of being admitted in the RDD draw.[19] We group these pupils together by controlling for a $[\tau-\delta, \tau+\delta]$ dummy (cohort-specific) and a piecewise linear spline in distance from cut-off τ. At this stage, it is worth noticing that, *ex-ante*, pupils within $[\tau-\delta, \tau+\delta]$ are also at risk of being admitted via the lottery among applicants who live within three miles of the school. Indeed, the $[\tau-\delta, \tau+\delta]$ dummy controls for the combination of RDD and lottery risk of being admitted to the school.[20]

Next, we consider applicants who live within three miles of the school, but not close enough to be meaningfully at risk of being admitted via the distance-based RDD assignment. That is, pupils within $(\tau+\delta$, 3 miles] of the school. These pupils are exposed to the same lottery risk of admission, and we consistently group them together by controlling for a $(\tau+\delta$, 3 miles] indicator (cohort-specific). Finally, a similar situation applies among pupils within (3 miles, 5 miles] of the school. These pupils are exposed to the same lottery risk of being admitted, and we group them together by controlling for a (3, 5 miles] (cohort-specific) indicator.[21]

The above approach works for the first cohort of applicants, but not for the second cohort, due to the truncation problem that arises from joining the LA admission mechanism (explained

[19] The thought experiment here is to fix the admission criteria, school capacity, and pupils' priority groups, and resample the distance tie-breaker within this group.

[20] This approach identifies 30 pupils in the $[\tau-\delta, \tau+\delta]$ group in the first cohort and 42 pupils in the second cohort. About a third receives an offer on average – approximately 55% in the first cohort and 20% in the second one – reflecting the 'truncation' issue due to the presence of more highly ranked schools in the LA-wide DA assignment. As explained above, this is only relevant among applicants in the second cohort.

[21] Approximately, 65% and 80% of applicants are in the $[\tau+\delta$, 3 miles] area for cohorts 1 and 2, respectively and 25% and 15% in the (3 miles, 5 miles] experiment. The corresponding figures for the likelihood of an offer are 16% and 4%; and 25% and 12%. Once again, numbers for the second cohort are affected by offer truncation.

above). We again use applicants' lottery numbers as an IV for school offers within the 'lottery' samples. However, as we cannot reconstruct the full outcomes of the LA algorithm, we have to drop pupils in the RDD sample in the second cohort. Luckily, there are only 42 pupils to whom this applies – and the similarity of results including/excluding pupils at risk from the RDD 'experiment' clearly indicates that they do not drive our findings.

4. Data and descriptive facts

We obtained pupil-level data for the first two cohorts of applicants to both free schools. Alongside details on the schools' admission criteria, the data include information on pupils' applications, whether the applicants received an offer from one of the free schools (and under which criterion), and the schools in which they enrolled. Overall, this dataset includes roughly 2,500 applicants.

We matched this information to pupil- and school-level data from the Pupil Level Annual School Census (PLASC) and to pupil-level primary (KS1 and KS2) and secondary (KS4) school exam scores from the National Pupil Database (NPD). These administrative datasets cover the entire pupil population in England's state-funded schools. The datasets give us access to pupil-background variables, including gender; free-school-meal (FSM) eligibility (a proxy for poverty); SEN status; indicators for whether pupils are of White British ethnicity, and whether English is their first language (EFL). These variables are measured at the time of application to secondary school. The data also report the share of days when pupils were absent (justified or unjustified) from school throughout their secondary school years, and whether they were ever suspended ('excluded') for disciplinary reasons.[22]

The school-enrolment data allow us to reconstruct the number of years each pupil spent in either of the two free schools. This is the key independent ('treatment') variable in our analysis. We also gathered information on the schools that unsuccessful applicants attend in the first year of secondary education. This allow us to reconstruct the type of institution they attend (single-sex schools, academies, faith schools, or grammar schools); intake composition (the share of female, FSM-eligible, White British, EFL, and SEN pupils); Ofsted inspection outcomes, overall and by domain (management, behaviour, teaching, and achievement),[23] as well as the pupil-teacher ratios, and teachers' qualifications, gender, and gross salary, and

[22] Data on applicants were matched to pupil-census records by the Department for Education (DfE), using full pupil details, including first name, family name, address, and date of birth, before being anonymised for our purposes. The matching was successful with more than 95% of the pupils linked. However, we lose some more observations because of missing achievement data or missing information on pupil-background characteristics. Our final sample includes approximately 2,200 observations.

[23] We used the data from the most recent inspection before the last cohort in the data finished secondary school.

teacher/headteacher mobility from the School Workforce Census. These data allow us to shed light on some 'standard' characteristics of counterfactual schools attended by applicants who did not enrol in one of the free schools.

Columns (1)–(3) of Table 1 present key descriptive statistics regarding the application data (Panel A) and the composition (Panel B) of both schools. Panel A shows that, on average, 1-in-3 applicants received an offer from the NE school, and 1-in-4 applicants enrolled in the school. However, these figures are different across the two cohorts: while 1-in-5 applicants received an offer and 1-in-5 applicants attended the school in cohort 1, the corresponding figures for cohort 2 are 1-in-2 and 1-in-5 respectively. For the CL school, on average 1-in-10 pupils received an offer and 1-in-11 pupils enrolled in the school. However, there is again cross-cohort heterogeneity: in cohort 1, 1-in-5 pupils received an offer and 1-in-10 enrolled in the school, while the corresponding figures for cohort 2 are 1-in-8 and 1-in-12.

While some of the variation in the share of offers made reflects changes in the number of applicants, cross-cohort differences in offers also reflect the truncation issue discussed above. Since data on preferences for years before 2014 were not collected by the DfE, we cannot characterise the extent of the truncation problem for the cohorts under investigation. However, using data for 2014 we find the following patterns. About 16% and 22% of preferences expressed for the NE and CL schools are first preferences. This evidence suggests that the pool of applicants displays a mixed degree of 'intensity of preference' for the two schools – and that our results are unlikely to only reflect the effect of 'attending your preferred school' (as in Cullen et al., 2006). In terms of preference truncation, among the students who ranked the two schools as their second preference (36% for the NE school and 23% for CL school), 42% and 52% for the NE school and CL school respectively receive an offer from a school they prefer more. Expectedly, the truncation problem is more important for 3rd and 4th choices (e.g., 82.5% of third preferences for the NE school are truncated from above, and 74% for CL school). As discussed, we address this issue by using lottery numbers as instruments for observed offers.

In terms of composition, Panel B of Table 1 shows that the NE school has a lower share of White British and EFL applicants than the CL one, and lower KS1 and KS2 scores in English, Mathematics, and Science.[24] However, the shares of female, FSM-eligible, and SEN applicants are roughly comparable across the two schools. For the CL school, we also know the home-to-school distance among applicants, one of the admission criteria, which is on average equal to roughly 1.9 kilometres.

[24] Pupils can attain KS1 scores between 3 and 21, and KS2 levels between 3, 4, and 5.

Next, Columns (4)–(6) of Table 1 report the national and the catchment-area averages of pre-application pupil characteristics. The catchment areas cover the LA or set of LAs from which each school attracts more than 90% of their applicants. While the catchment area for the NE school overlaps with the LA in which it is located, where nearly all applicants live, the catchment area for the CL school covers five London LAs. For both schools, the share of female applicants is lower than the national and catchment-area averages. The shares of FSM-eligible applicants are in line with the national average, but lower than the catchment-area average. Both schools have lower shares of White British and EFL applicants compared with the national average. However, the CL school has higher shares of White British and EFL applicants compared with the catchment-area average, while the opposite is true for the NE school. Finally, among applicants to both schools, average KS1 and KS2 scores are higher than both the national and catchment-area averages.[25]

Overall, our evidence shows there is some positive self-selection among applicants in terms of prior ability. Conversely, selection along socio-economic status or ethnicity does not appear as prevalent. Although these considerations do not impact the internal validity of our estimates, they are relevant for external validity. As discussed in the Introduction, this is an 'inescapable' feature of all studies that evaluate school effects in the context of school choice in education quasi-markets and using features of the application process, such as lotteries in the case of US charter schools. Parents and pupils are likely to self-select into the pool of applicants for different schools based on expected gains from those schools. Applicants are therefore almost by definition different from non-applicants. Nonetheless, these very same issues mean that the pool of applicants is a key population to focus on in the context of market-based approaches to education reforms – including the free-schools programme that we analyse in this paper.

5. Main Results

5.1. Balancing tests

If our identification strategy works, successful applicants should be comparable to unsuccessful applicants along all observable and unobservable dimensions – conditional on cohort-specific 'experiment' fixed effects and the distance-based running variable that determines assignment for pupils exposed to the RDD admission risk (see Sections 3.2 and 3.3).

[25] Andrews and Johnes (2017) find that free schools tend to open in disadvantaged areas, but that less well-off students from these areas are less likely to attend a free school.

To provide evidence in support of our identification strategy, we test the balancing of treated (successful) and control (unsuccessful) applicants' observable and pre-determined characteristics by estimating the following model with Ordinary Least Squares (OLS):

$$X_i = \gamma \mathrm{D}_i^j + \alpha_{jck} + \phi(d_i) + \varepsilon_i, \qquad if \;\; 0 < P_i^{jck} < 1 \qquad (1)$$

In Equation (1), P_i^{jck} identifies pupil i's probability of being admitted to free school j (NE, *j=1* or CL, *j=2*) in cohort *c* (Cohort 1 or 2) via experiment *k*, representing a stanine in the NE school and the RDD group plus the two lottery bands in the CL school. The restriction $0 < P_i^{jck} < 1$ identifies pupils with a non-degenerate admission risk, thereby excluding from our estimation sample pupils with 'reserved' seats at each school and those not at risk (the 'always seated' and the 'never seated' in the RDMD terminology – see Section 3). X_i is a vector of observable and pre-determined pupil characteristics, including female, White British, EFL, FSM, and SEN dummies, as well as KS1 and KS2 scores in English, Mathematics, and Science; D_i^j is a dummy variable equal to 1 if pupil i received an offer from one of the free schools ($j = 1, 2$) and equal to 0 otherwise; α_{jck} are free school j-by-cohort c-by-experiment k dummies, identifying pupils that face the same risk of being admitted to the schools within each cohort. In addition, $\phi(d_i)$ are cohort-specific piecewise linear running-variable controls for distance from the cut-off for the RDD draws, d_i, parametrised as follows:

$$\phi(d_i) = \sum_{2ck} L_{i,2ck}\left[\phi_{1,2ck} \times d_i + \phi_{2,2ck} \times d_i \times 1(d_i \geq \tau_{2ck})\right] \qquad (2)$$

In Equation (2), $L_i = I(\tau_{2ck} - \delta_{2ck} < d_i < \tau_{2ck} + \delta_{2ck})$. That is, L_i is a dummy variable equal to 1 if pupil i is at risk of being admitted through the RDD mechanism at the CL school ($j = 2$) in Cohort $c \in$ (*1,2)*, and belongs to the [τ-δ, τ+δ] category, and equal to 0 otherwise. In practice, τ_{2ck} is observed in the data as the distance of the last pupil admitted, while bandwidth δ is estimated by using the optimal bandwidth estimation method proposed by Calonico et al. (2014). This specification ignores the truncation problem among the second cohort of applicants to both schools. As discussed, we use applicants' lottery number in an IV strategy to test whether this poses a threat to our identification strategy.

Table 2 reports the results from the balancing tests when pooling both cohorts. Panel A displays the estimates for both free schools, while Panels B and C report separate estimates for the NE and CL schools respectively. The results from a test for joint significance of all estimates, when including all pre-determined variables in the same model, is reported at the bottom of each panel. Irrespective of the chosen sample, the results support the internal validity

of our research design: the treatment and control groups are comparable in terms of applicants' observable characteristics.

Estimation results obtained for each cohort separately are reported in Tables A1 and A2 in the Online Appendix and are in line with those in Table 4. While a few variables are unbalanced, the extensive set of characteristics is jointly balanced. In unreported models, we also found that the balancing estimates were very similar when excluding all pupils in the RDD sample, thus only comparing pupils who are admitted through one of the lotteries.

Finally, Table A3 in the Online Appendix reports the results of balancing tests for the lottery-number instrument used to deal with the truncation problem for the second cohort of applicants. We report school-specific results for the second cohort of applicants, and exclude pupils admitted through the distance-based RDD assignment mechanism used by the CL school (as we do not have lottery draws for pupils admitted via the RDD mechanism). Overall, the evidence suggests that our identification strategy works: the lottery number is uncorrelated with most pre-determined pupil characteristics. The p-values for the tests of joint significance of all estimates are 0.818 and 0.153 for the NE and CL schools respectively.

5.2. Average free-school effects

We estimate the effects of years of attendance at each free school using the following model:

$$Y_i = \beta YearsOfExposure_i^j + \phi_{jck} + \psi(d_i) + \eta_{jc}X_i + \epsilon_i, \;\; if \;\; 0 < P_i^{jck} < 1 \quad (3)$$

Equation (3) is analogous to Equation (1), but the dependent variable Y_i is pupil *i*'s KS4 average test scores across all subjects taken by students, while the treatment variable of interest is $YearsOfExposure_i^j$, measuring the number of years spent by pupil *i* in either of the two free schools. To increase the precision of our estimates, we include the controls in vector X_i and allow their coefficients to differ by each school-by-cohort group. Given that actual free school attendance is endogenous, we estimate Equation (3) with Two-Stage Least Squares (TSLS), using the free-school offer dummy D_i^j as an instrument for $YearsOfExposure_i^j$. The relevant first-stage equation is:

$$YearsOfExposure_i^j = \gamma D_i^j + \psi_{jck} + \mu(d_i) + \kappa_{jc}X_i + v_i, \;\; if \; 0 < P_i^{jck} < 1 \quad (4)$$

Again, this specification ignores the truncation problem for cohort 2. We therefore also use the IV strategy that exploits applicants' lottery numbers as instrument (described in Section 3) to assess the importance of this issue and validate our main estimates.

Results from models pooling both cohorts are presented in Table 3. Columns (1)–(3) display results when analysing both schools together, Columns (4)–(6) display estimates for the NE school only, Columns (7)–(9) display estimates for the CL school only, and Columns (10)–(12) display results when excluding the 72 pupils who were exposed to the RDD risk of admission at the CL school across the two cohorts. This is the relevant benchmark for the estimates that address preference truncation (reported in Table 4 and described below).

Columns (1), (4), (7), and (10) report reduced-form effects of a free-school offer on KS4 tests, showing that successful applicants have KS4 scores that are roughly 0.16 standard deviations higher than unsuccessful applicants. This effect is statistically significant in all samples and comparable in magnitude across the two schools. Next, columns (2), (4), (8), and (11) report the first-stage effects of a free-school offer on years of exposure. We see that successful applicants spend on average two more years than unsuccessful applicants in the two free schools. The effect is again comparable across the two schools and strongly significant in all samples, as confirmed by the first-stage F statistic, which always displays values considerably larger than 100. Finally, columns (3), (6), (9), and (12) report the TSLS effects of years of free-school exposure on KS4 scores. The results show that spending an additional year in a free school boosts achievement by 0.076 standard deviations. We obtain a marginally higher effect size for the NE school (0.087 standard deviations) than for the CL school (0.059 and 0.062 standard deviations), but the differences are not large. Importantly, we find that excluding pupils at risk of being admitted through the RDD mechanism makes no relevant difference for the estimation of the CL-school effect.[26]

Separate results for each cohort are reported in Tables A4 and A5 in the Online Appendix. These results are broadly similar to our main findings. When pooling both schools, we find an impact of 0.082 standard deviations in the first cohort and an effect of 0.072 standard deviations in the second cohort. Cohort-by-school specific results are also in line with the findings in Table 3. The NE school has an impact of 0.089 and 0.084 standard deviations for cohorts 1 and 2, respectively, while the CL school has effects between 0.055 and 0.075 standard deviations, depending on the cohort and on whether we include pupils at risk of being admitted through

[26] Excluding students 'at risk' of RDD admission is conceptually very different from disregarding altogether pupils who might be selected by distance, and focussing on the two lottery groups at the CL school. This is because *ex-ante* it is impossible to know who is only at risk of admission from a lottery draw among pupils living within 3 miles of the school. This set can only be identified by using the insights of the RDMD literature and by categorising pupils as 'always seated', 'never seated' and 'conditionally seated' for the RDD experiment.

the RDD mechanism. However, these findings are noisier – which is expected given the smaller sample sizes – and the estimates for the CL school are not significant.[27]

Next, we investigate the potential problem of offer truncation among applicants in the second cohort. Our findings are presented in Table 4, in which we compare the effects for each school when using the offer dummy (Panel A) or the lottery number (Panel B) as instruments for years of free-school exposure, among pupils in the second cohort only. As discussed, in the analysis of the CL school, we drop pupils at risk of being admitted through the RDD mechanism as we do not have a credible way to address the issue of truncation. This is not a significant exclusion – approximately 40 pupils – and does not affect our main estimates.

The instrumental variable estimates that we obtain using the offer IV are 0.072 (s.e. 0.026) when pooling the two schools; 0.084 (s.e. 0.037) for the NE school; and 0.055 (s.e. 0.038) for the CL school. The corresponding estimates we obtain using the lottery number IVs are 0.085 (s.e. 0.026), 0.101 (s.e. 0.048), and 0.060 (s.e. 0.102). We therefore conclude that the two approaches lead to economically similar and statistically indistinguishable effects.

While the potential problem of truncation only affects the second cohort of applicants, an IV strategy that instruments admission offers with lottery number can in principle be replicated for both cohorts to investigate the issue further. Unfortunately, for the CL school, we do not have access to applicants' lottery numbers among applicants in the first cohort. However, we can pool both cohorts and compare the findings across the two cohorts for the NE school. The evidence in this respect is reassuring. When pooling the two cohorts for the NE school and instrumenting years of attendance with lottery numbers we find an estimate at 0.088 (s.e. 0.026), virtually identical to the one obtained when instrumenting years of attendance with the offer dummy. The same applies when analysing the two cohorts separately for this school.

Overall, these findings suggest that offer truncation is not a significant threat to causal identification in our context. Perhaps, this is not surprising: offers were made using random lottery draws – which are by construction orthogonal to students' preferences (not revealed to the schools), the extent of truncation, and its possible impact on our estimated effects.

The effects reported so far concern average pupil achievement across all subjects. In Table A1 in the Appendix, we focus on the three core subjects that all students take at the end of

[27] We also estimated models where we look at free-school attendance at KS4, as opposed to years of exposure. Our findings are reported in Table A6 in the Online Appendix and confirm that both schools have a similarly positive, large and significant effect on KS4 test scores. Furthermore, we carried out a battery of check to validate the robustness of our results, including: *i)* dropping the pupil-level controls; *ii)* controlling for KS2-level dummies instead of the linear KS2 score; and *iii)* checking sensitivity to the choice of polynomial and bandwidth utilised in the RDD sample for the CL school. Our estimates are reported in Table A7 in the Appendix, which shows that the main results are robust to these changes.

secondary education – namely, English, Maths, and Science. We also investigate the impact of attending a free school on the number of exams taken at KS4. We find positive effects of attending the two free schools on scores in each of the three core subjects. Due to data limitations, we only observe science scores for one cohort, and as a result the positive effects for science are imprecisely estimated for the CL school. Importantly, the two free schools not only improve average grades and increase the share of pupils attaining a pass grade; they also increase the incidence of pupils who achieve top-level grades. Conversely, we find no evidence that the higher scores obtained in the core subjects are achieved at the expense of students entering a significantly lower number of exams. Although the coefficient for the number of exams entered for the CL school pupils has a negative sign (consistent with its educational philosophy described in Section 2.4), it is small and not statistically significant.

To sum up, our main findings are as follows. We find that the average positive impact of the two free schools on KS4 achievement corresponds to 0.076 standard deviations per year of exposure. This impact is slightly larger for the NE school, at 0.087 standard deviations, than for the CL school, at 0.062 standard deviations, but the difference is far from statistically significant. We also find that dealing with offer truncation issues that affect the second cohort of students does not affect our findings. Overall, our effects are comparable to the impact of some of the most successful US charter schools (see Cohodes and Parham, 2021), but smaller than the effects of some of the Boston charters studied by Abdulkadiroğlu et al. (2011).

5.3. Heterogeneous effects by pupil background

Given that the two free schools under investigation adopt different pedagogical approaches, we may expect them to benefit different subgroups of the pupil population. To investigate these issues, in Table 5 we estimate the impact of free-school attendance on KS4 test scores across different subgroups of pupils, pooling the two cohorts of applicants. Panel A reports the effects of the NE school, and Panel B reports the effects of the CL school.[28] Throughout our discussion, we mostly emphasise the patterns that emerge when comparing point estimates – instead of the statistical significance of any such differences – as we believe sample sizes are too small to precisely pin down heterogeneous effects.[29]

[28] Note that we use offer as an instrument for years of attendance rather than lottery numbers – as the latter would entail losing one cohort of students at the CL school. For consistency with the previous tables, we also exclude pupils admitted on the basis of proximity to residence.

[29] We do not consider issues due to multiple-hypothesis testing. However, a simple Bonferroni correction would reveal none of the differences we identify is significant.

The estimates show that the NE school is especially effective for males: the point estimate is a strongly significant 0.117 among boys and only a non-significant 0.038 among girls. While the two estimates are not statistically different (the p-value on the significance of the difference is 0.162), they display an economically meaningful heterogeneity. We also find that non-FSM-eligible pupils appear to benefit slightly more from attending the NE schools than FSM-eligible pupils: the point estimate is a significant 0.081 among the former and a non-significant 0.058 among the latter. We also find a precisely estimated and sizeable effect among non-White British pupils, and a slightly larger but statistically insignificant one among White British pupils. However, since the school mostly enrols non-White British pupils, it is difficult to draw strong conclusions in this respect. Finally, we find that the NE school is slightly more effective for initially low-achieving pupils, based on their KS2 performance: the point estimate is a statistically significant 0.084 among low-achieving students, compared to a statistically insignificant 0.066 among high-achieving pupils.

We can contrast these findings with the insights from our visit. When asked which groups of pupils benefited more from attending the school, the headteacher suggested 'disadvantaged students'. While this lines up with the breakdown in terms of low/high prior achievements, it does not square well with the FSME dichotomy. Furthermore, when asked whether the school educational approach particularly favours boys, the trust CEO suggested that this is possible as 'girls manage to stay afloat in the other schools in the area, whereas boys get lost and disengage from education'. Although this *ex-post* rationalisation is in line with our findings, the gender divide was not offered as one of the key patterns of heterogeneity perceived on the ground.

On the other hand, the CL school delivers considerably larger effects among White British and non-FSM-eligible pupils than among non-White British and FSM-eligible pupils: the free-school impact is a strongly significant 0.124/0.075 among White British/non-FSM-eligible pupils, and an insignificant 0.010/0.020 among non-White British/FSM-eligible pupils. The White/non-White difference is borderline statistically significant with a p-value of 0.075. We find no clear differential effects depending on gender – as estimates for boys and girls are similar in magnitude – but some evidence that the school is more effective for high-achieving pupils than low-achieving ones: the point estimate for high-achieving students is a borderline significant 0.084, compared to an insignificant 0.051 among low-achieving ones.

Does this evidence line up with the information we gathered from our interviews? To a large extent, it does. Indeed, the headteacher lamented that the school is more effective and more attractive for pupils from more affluent families, probably because of its CL pedagogical philosophy. This was identified as one of the main 'failures' of the school, as it was set up to

serve a relatively disadvantaged community and provide it with a rigorous but inspiring schooling alternative.

Irrespective of the perceptions of stakeholders we spoke to, these heterogeneous patterns appear in line with the educational approaches championed by the two schools. Previous research suggests the 'no excuse' approach might work especially well for boys, who may need a more disciplinarian approach, and low-achieving pupils, who need an approach focussed on changing expectations and aspirations. On the other hand, the 'classical liberal' approach is markedly more effective among White British/non-FSM-eligible pupils, and to some extent among high-achieving ones. This may be because the 'classical liberal' approach – based on a knowledge-rich curriculum and a 'specify and excel' attitude – works better among pupils from a more advantaged background than for less well-off pupils.

5.4. The free-school effects on behavioural outcomes

Given growing evidence on the importance of school (and teacher) effects on non-cognitive outcomes (e.g., Jackson, 2018), we investigate the impact of the free schools on behavioural outcomes. Specifically, we estimate effects on: *i)* the probability that pupils sit their KS4 exams at the school where they started Year 7; *ii)* unauthorised absences; and *iii)* school exclusions for disciplinary reasons.[30] Our results are presented in Table 6.

We find that pupils enrolling in one of the free schools in Year 7 are substantially more likely to stay in the same school until their KS4 exams compared to unsuccessful applicants who enrolled elsewhere (approximately a 25 percentage-point increase over the average among counterfactual students at 69% – amounting to a 36% increase in retention). Pupils spending longer time in one of the two free schools also miss fewer school sessions due to unauthorised absences, irrespective of whether we use the share of sessions missed (impact sizes of around 9% and 18% for the NE and CL schools, respectively) or a dummy for whether they have missed any session (impact sizes of 3.6% and 5.3% for the NE and CL schools). However, we do not find significant differences in terms of disciplinary exclusions, although there is a marginally significant effect on the probability of any exclusion for the NE school. Still, such exclusions are rare – 14 percent of pupils have been excluded for at least a session, but the average number of sessions missed is less than half a day per year. We may therefore lack power to uncover small effects. This null effect on disciplinary exclusions may be the result of

[30] Grade repetition is very rare in England so we cannot look at this outcome. Note that we use a binary treatment variable indicating whether pupils enrolled in the free schools in Year 7 when analysing effects on the probability of remaining at the same school until the KS4 exams. However, we revert to using years of exposure as treatment variable when studying the two school effects on absences and exclusions.

two competing forces: better behaviour among pupils on the one hand, and stricter discipline on the other. Overall, these results suggest that the positive effects of the two free schools on pupil achievement may be partly explained by their negative impact on pupil mobility and positive effect on pupil behaviour.

We also analyse potential heterogeneous effects of the free schools on these behavioural outcomes, but fail to detect insightful patterns. Our results are reported in Tables A8–A10 in the Online Appendix. We find that that the NE school reduces school mobility to a larger extent among low-achieving pupils than among high-achieving ones – but its impact is also larger for girls than for boys, and for FSM-eligible pupils compared to non-FSM-eligible ones. Meanwhile, the CL school has a larger negative impact on pupil mobility among White British pupils compared with non-White British ones – but also a larger impact on FSM-eligible pupils than among non-FSM-eligible ones. These patterns are difficult to square with the heterogeneous effects on pupil achievement. Similarly, when analysing the impact of the free schools on unauthorised absences, the findings are hard to reconcile with our evidence on their heterogeneous effects on KS4 test scores. For the NE school, we find a larger impact among non-White British pupils and among low-achieving pupils than among White British pupils and high-achieving ones, which is in line with the heterogenous effects on achievement. However, we also find a larger effect for girls than boys, which is not in line with the evidence on achievement. Meanwhile, for the CL school, we find larger effects among White British pupils than among non-White British ones, which is consistent with the effects on achievement, but also a larger impact among low-achieving pupils, which is not in line with the impact we document on achievement. Finally, we find little heterogeneity in terms of the free schools' effects on disciplinary exclusions.

In conclusion, our evidence indicates that behavioural mechanisms that operate by simply increasing educational continuity (i.e., fewer absences and school changes) are unlikely to fully account for our results on KS4 test scores.

6. Understanding mechanisms

6.1. School attributes

What mechanisms could explain the effects of the free schools? We start by considering a wide range of 'standard' school characteristics and analyse whether the two free schools differ along these dimensions compared to counterfactual schools attended by rejected applicants.[31]

[31] Our results are unlikely to depend on the fact that the schools are new – with students and teachers being 'on board' and galvanised – which would suggest that their effectiveness could be diluted over time when such

We carry out this analysis at the pupil level and study whether students starting secondary education (in Year 7) at one of the two free schools experience different school characteristics compared to pupils who applied to the free schools but ended up attending other schools. As before, we overcome the endogeneity of the decision to enrol in a free school by using the offer dummies as instruments. We look at four categories of school characteristics: *i)* pupil demographics; *ii)* characteristics of the teaching body; *iii)* institutional features; and *iv)* Ofsted reports.[32] Our results are reported in Table 7. Column (1) reports the results for the NE school compared with the counterfactual schools, while column (2) reports the results for the lottery sample at the CL school (to limit problems related to the truncation issue discussed above).[33]

In terms of demographics, we find that successful applicants enrolling in the NE school experience lower shares of FSM-eligible pupils (-4 percentage points), White British pupils (-20 percentage points), and female pupils (-9 percentage points) than unsuccessful applicants who enrol in counterfactual schools. Meanwhile, successful applicants attending the CL school are exposed to a higher share of White British pupils (+16 percentage points) than those in counterfactual schools. In other words, there is evidence that the pupil composition differs somewhat between the two free schools and the counterfactual ones, although there is little consistency between the two free schools under study in this respect. This suggests peer effects are unlikely to be an important mechanism.

There are also differences in the characteristics of the teaching body between the free schools and the counterfactual ones. Pupils in the NE school enjoy a lower pupil-teacher ratio than those in counterfactual schools (-2.8 pupils per teacher, with the average close to 14.5) and a higher share of qualified teachers (+5 percentage points, with an average of 95%), while those in the CL school face a substantially higher share of male teachers (+12 percentage points, with the average close to 39%), and a lower share of qualified teachers (-19 percentage points, with the average being 94%). Furthermore, teachers in the NE school are paid less than those in the counterfactual schools (-£3,000 per year, with the average salary being £37,500),

'novelty effects' wear out. The fact that the schools are still effective five and six years after opening (at the end of the KS4 phase for the first and second cohort, respectively) speaks against such possibility.

[32] These regressions can only be estimated on the sub-sample of pupils who had a valid school identifier for Year 7. Missing school identifiers could be due to short-term episodes of mobility or transitions to/from independent schools. They affect roughly 12% of the sample. Tables A11 and A12 in the Online Appendix replicate the estimates for the balancing and school effectiveness presented above within this sample – and provide results that are virtually identical to our main findings.

[33] Note that, again, we are instrumenting attendance using offer dummies rather than lottery numbers. Using lottery numbers would imply losing the second cohort of students for the CL school for which lottery numbers are not available. This would significantly limit the precision of our estimate. However, as we showed before, dealing with truncation issues by using lottery instruments does not meaningfully change our findings.

while those in the CL school are paid more (£9,000 per year, with the average salary being £43,500). We also find that between the first and the second year of opening, the NE school had higher teacher mobility than its counterfactual schools – while the opposite is true for the CL school. When focussing on headteachers, however, we find opposite patterns.[34] Overall, these differences are unlikely to consistently explain the large positive effects on KS4 outcomes, with similar effect sizes across the two schools.

The institutional setup of counterfactual schools attended by applicants who do not enrol in one of the free schools is also on average different. Compared to free-school attendees, pupils in counterfactual schools are more likely to enrol in single sex schools (both free schools are mixed gender) and faith schools (both free schools are not religious). They are less likely to attend an academy or a free school. However, we see no statistically significant difference in the probability of attending a selective school. Once again, it is unlikely that these differences drive the effects on pupil achievement. Bertoni et al. (2022) study the relative effectiveness of different secondary school types in the England and find that only selective grammar schools deliver higher academic returns on average. Neither of the free schools we analyse is selective, nor are the counterfactual schools.

Finally, applicants who do not enrol in the NE school end up in schools with worse Ofsted ratings in all aspects of the evaluation (by 1.2–1.4 points, on a scale of 1 to 4, with 1 being 'outstanding' and 4 being 'inadequate'), while applicants who do not enrol in the CL school end up in schools with better inspection ratings in most respects (by 0.14–0.31 points), with the exception of the behavioural domain: this is the area in which the CL school excels compared to the counterfactuals (0.64 points higher).

This pattern confirms the insights we gathered in Section 5.4: improved behaviour might be a potential mechanism explaining the positive effects of the two free schools. But, as already highlighted, it is unlikely to fully account for the effects of the free schools on achievement and their heterogeneity by pupil background.[35]

6.2. School pedagogical approaches

One channel we are yet to quantify relates to the free schools' distinctive pedagogical approaches: their 'no excuse' and 'classic liberal' philosophies. To investigate this issue, we

[34] Due to data limitation, we can only measure teacher/headteacher mobility across two adjacent years.

[35] In unreported regression, we investigated whether the characteristics of the counterfactual schools discussed above help explain the heterogeneous effects we find on test scores. Once again, we did not observe heterogeneous patters along these dimensions that matched the patterns we found when analysing KS4 achievements.

carry out a quantitative text analysis on web-scraped VE statements for the two free schools and the counterfactuals in our sample.

To do so, we start by creating a taxonomy of 'concepts' that characterise the pedagogical approaches of the free schools. These are pinned down by key words – including core values and drivers – and key expressions in the VE statements that resonate with the impressions we gathered from our school visits and semi-structured interviews. We then create an extended 'dictionary' that associates such concepts to a wide range of synonyms and akin expressions that can be 'searched' in the statements of counterfactual schools. Finally, we create two synthetic measures that capture whether a key concept – representing an aspect of the free schools' pedagogical approach – is 'over-represented' at a given schools compared to what would be expected by looking at the full sample of VE statements (correcting or not for the complexity of the words used). More details are provided in the Appendix.

Descriptive statistics for our indexes are reported in Appendix Table 2. The top panel focusses on the NE school, while the bottom panel focusses on the CL school. Column (1) reports the value of the (standardised) unadjusted index for each of the key concepts for the two free schools. Column (2) and (3) report the mean and median of the index among counterfactual schools. Column (4) reports the share of counterfactual schools that have words or expressions related to that concept. Finally, Columns (5) and (6) report the free-school rank within the relevant sample (i.e., NE plus 52 counterfactuals; or CL plus 96 counterfactuals), using either the simple or the complexity-adjusted index.

Starting with the top panel, we see that some of what we thought were characterising features of the NE school are clearly present among counterfactuals. For example, 85% and 90% of the counterfactual schools also focus on purpose and autonomy, and the NE school only ranks 24th and 42nd (using the complexity-adjusted index) in these domains. On the other hand, the no excuse and strict routines features of the free school appear more unique: only 7.7% and 9.6% of counterfactual schools relate to these concepts – and the NE school ranks at the top (1st or 2nd/3rd depending on the index considered). Interestingly, while many other schools also refer to hard work and a culture of success – concepts that are present in 92% and 98% of the VE statements – the NE school clearly puts much more emphasis on this idea, ranking between 3rd or 11th depending on the index. This evidence suggests that our index works well: not only does it capture the recurrence of a concept within a given statement, but also the intensity with which it is emphasised by that school.

Moving to the bottom panel, we find once again that some of the practices adopted by the CL school are quite common. Two of its core values – hard work and high standards – appear

in 82% and 95% of the VE statements of the counterfactuals, and the school only ranks 45th and 47th (using the complexity adjusted index). On the other hand, the classical liberal and knowledge-rich concepts are quite unique: they only appear in 2% and 10% of the statements – and the school is very highly ranked (1st and 4th looking at Column 6). More strikingly, we find that teacher-led learning does not appear in any of the counterfactuals. Finally, while politeness is a concept that emerges from 86% of the VE statements, the CL school emphasises it to a larger extent, ranking 4th and 15th in Columns (5) and (6).

As for the 'standard' school attributes considered above, we merge this information to our main dataset at the pupil level and use our regression framework to study whether individuals who do not attend one of the free schools join counterfactuals with similar or different pedagogical approaches. Again, we control for experiment-by-cohort effects (and pupil characteristics) and instrument attendance at a free school at the start of secondary education with offer receipt. Our findings are presented in Table 8, where we concentrate on the standardised complexity-adjusted index (results when using the simple index are similar). The top panel focusses on the NE school, while the bottom panel reports results for the CL one (lottery sample only).

We find that pupils attending the NE school are significantly more likely than those in counterfactual schools to be exposed to a 'no excuse' approach to education, characterised by a strong emphasis on strict routines, culture of success, and hard work, as well as to the values of trust, fairness, and mastery. These are characterised by the school as follows: 'We are loyal (…) and always do what we say we will do' (trust); 'We are open-minded (…) and play by the rules' (fairness); and 'The urge to get better and better (…)' (mastery). Once again, this seems to emphasise a 'no excuse' approach to education.[36]

For pupils attending the CL school, we find that the classical-liberal philosophy – alongside the concepts of knowledge-rich, competitive atmosphere, and debates – are significantly more prevalent than among pupils at counterfactual schools. Expectedly, we find that a teacher-led approach to learning is a key feature of the CL school, while high aspirations seem to be emphasised less than at the counterfactuals. This is an interesting juxtaposition given that Dobbie and Fryer (2013) find that both features are amongst the strongest correlates with school effectiveness in their study of US charters. We also find that the concepts of

[36] Other elements of the 'no excuse' approach identified among US charters – such as extended instructional time and high dosage tutoring (see Angrist et al., 2013 and Dobbie and Fryer, 2013) – are not common in England because of institutional constraints (mostly operating via the teaching workforce). Cirin's (2014) survey of academies' and free schools' use of their freedom found that only 8% and 4% had changed the length of the school day and teaching terms, respectively.

politeness and kindness are over-represented among pupils in the CL school. While the first relates to the disciplined, non-permissive approach to pupils' behaviour at school, the second is more closely related to ideas such as altruism, compassion, and consideration of others (and is one of the school's core values). Finally, hard work and high standards are not substantially more salient elements of the environment faced by pupils at the CL school compared to the counterfactuals, despite both being among its core values.

We also investigated whether the prevalence of these differing elements of the free schools' pedagogical approaches, compared to counterfactual schools, vary by pupil background. We found that this is not the case (results available from the authors on request). This finding suggests that the differential impact of the two free schools on achievement among students with different characteristics is not driven by their differential exposure to the NE and CL teaching philosophies. Instead, it suggests that these different educational approaches are likely to be differentially effective depending on pupil background, in ways that are reminiscent of what has been documented in the US charter literature.[37]

7. Conclusions

Do 'bottom-up' educational initiatives that expand the supply side of the education market work? To shed light on this question, this paper has assessed the causal effects of two start-up schools that were established as part of the early roll out of the free-school programme in England. Using detailed admission data for the first two cohorts of applicants to these schools and an identification strategy exploiting lotteries and a distance-based RDD to draw causal conclusions, we have shown that these two schools substantially increase pupil achievement.[38] In addition, we have highlighted that the different pedagogical approaches adopted by the schools appear to benefit different types of pupils. While the school with a 'no excuse' approach is especially effective for boys, and somewhat more effective for low-achieving pupils than high-achieving ones, the 'classical liberal' school delivers considerably better

[37] Certainly, we cannot observe schools' *actual* adherence to their aspirational VEs. Since English education policy has increasingly come to emphasise the importance of the general aspects of the free schools' pedagogical approaches (e.g., aspirations and high standards), one might expect other schools to pay lip service to aspects of these concepts in their VEs. If that were the case, we may underestimate the *actual* differences between the two free schools and the other schools in terms of educational philosophies.

[38] Dobbie and Fryer (2020) show that charter schools have the potential to impact university and labour market outcomes, and our future research will consider these issues. Currently, however, it is not meaningful to analyse longer-term outcomes: students in first cohort would have finished university at the earliest in 2020, while those in the last cohort would have completed university in 2022. We thus believe it is better to wait a few years before carrying out an analysis of the long-term effects of attending the free schools.

results for White British and non-poor pupils than for non-White British and poor pupils, and marginally better results among high-achieving pupils compared to low-achieving ones.

Since we only evaluate the effects of two schools, it is hard to generalise our findings to the free-school sector as a whole. This is especially true since free schools are far from homogenous in their practices, as they subscribe to different educational philosophies. How different are our two free schools from the rest of the sector? Using our quantitative text analysis, we can shed some light on this issue. Our insights are summarised in Table 3 in the Appendix, which compares our two free schools' pedagogical approaches to those of other free schools that existed at the time. We find that quite a few of the other free schools have adopted many of the educational concepts embedded in the NE school, including 12% that explicitly use the expression 'no excuses' (or similar) in their VE statements. Conversely, we find that fewer free schools resemble the CL one, especially when it comes to its classics-infused pedagogy (no other free schools' VE statement mention this concept) and the teacher-led approach to learning (only one other school explicitly mentions this).[39]

What, then, to make of our results? First, our causal evidence strongly suggests that bottom-up initiatives focussed on education provided by a set of heterogeneous and independently managed schools – with the state confined to funding, governing, and monitoring such independent players – can be successful at improving pupils' learning. Second, our analysis of heterogeneous effects and our quantitative text analysis of the schools' pedagogical philosophies suggest that these improvements arise from match-specific gains occurring because of the complementarities between schools' pedagogical offer and students' specific background.

Taken together, these results provide some valuable insights regarding the functioning of quasi-markets in education with differentiated and autonomous school providers. As maintained by supporters of these approaches to education, there are potential gains from autonomy and differentiation as these features can improve the match between pupils' needs and schools' offer, thus improving standards. However, this conclusion hinges on the idea that all families actively exercise choice and are in a position to identify the schools that best cater for their children's needs. In turn, this consideration suggests that – for free schools to be 'tide that lifts all boats' – school choice and admission systems should be designed in ways that make the matching between pupils and schools as efficient as possible.

[39] This is another high-profile free school: Michaela Community School (in Wembley, North-West London). The school is well-known for its regimented teaching philosophy and combines a 'no excuse' approach with concepts that are similar to those of the CL school in our taxonomy (such as knowledge rich and teacher-led learning).

More narrowly, our paper also make some important contributions to the debate on the effects of autonomous schooling, especially in relation to the English academies' programme. Free schools are radically different from early sponsored academies (studied by Eyles and Machin, 2019), as they are not affiliated with a sponsor and therefore do not benefit from the additional resources or expertise brought by such associations. Instead, free schools share many features with converter academies, and available evidence shows that outstanding converter academies can enhance pupil achievement (see Bertoni et al., 2020). However, methodological considerations caution against a straight-forward comparison of these findings. Indeed, in this study, we have exploited lotteries and a distance-based RDD to study the causal effects of free schools among a set of applicants. Conversely, research on converter (and sponsored) academies uses a 'legacy' difference-in-difference approach to by-pass selection issues. These estimates pin down the impact of autonomy on pupils who chose these schools prior to their conversion to academy status – and are therefore unlikely to encompass strong match-specific effects that would be estimated among a pool of applicants. Although both sets of estimates are internally valid, they refer to sub-populations with different unobservable propensities to apply to, and attend, an autonomous school – and are therefore also likely to reflect different expected returns from autonomous schools.

References

Abdulkadiroğlu, A., J. D. Angrist, S. M. Dynarski, T. J. Kane and P. A. Pathak (2011). Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots. *Quarterly Journal of Economics*, *126*(2): 669-748.

Abdulkadiroğlu, A., J. D. Angrist, Y. Narita and P. A. Pathak (2017). Research design meets market design: Using centralized assignment for impact evaluation. *Econometrica*, *85*(5): 1373-1432.

Abdulkadiroğlu, A., J. D. Angrist, Y. Narita and P. A. Pathak (2021). Breaking Ties: Regression Discontinuity Design Meets Market Design. Forthcoming, *Econometrica*.

Allen, R. and R. Higham (2018). Quasi-markets, school diversity and social selection: Analysing the case of free schools in England, five years on. *London Review of Education*, 16 (2): 191–213.

Andrews, J. and R. Johnes (2017). *Free Schools in England.* Education Policy Institute report.

Angrist, J. D., S. M. Dynarski, T. J. Kane, P. A. Pathak and C. R. Walters (2010). Inputs and Impacts in Charter Schools: KIPP Lynn. *American Economic Review*, vol. 100(2): 239-243.

Angrist, J. D, P. A. Pathak and C. R. Walters (2013). Explaining Charter School Effectiveness. *American Economic Journal: Applied Economics*, vol. 5(4), 1-27.

Arteaga, F., A. J. Kapor, C. A. Neilson and S. D. Zimmerman (2022). Smart Matching Platforms and Heterogeneous Beliefs in Centralized School Choice. *Quarterly Journal of Economics*, vol. 137(3), 1791-1848.

Bertoni, M., S. Gibbons, and O. Silva (2020). School choice during a period of radical school reform. Evidence from academy conversion in England. *Economic Policy*, *35*(104): 739-795.

Bertoni, M., F. Gonschorek, T. Klein and O. Silva (2022). School types: Evidence on the relative effectiveness of different school arrangements in England. Mimeo, London School of Economics.

Biasi, B. and S. Ma (2023). The Education-Innovation Gap. Mimeo Yale School of Management.

Böhlmark, A. and M. Lindahl (2015). Independent Schools and Long-run Educational Outcomes: Evidence from Sweden's Large-scale Voucher Reform. *Economica*, 82(327): 508-551.

Calonico, S., M. D. Cattaneo and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6): 2295-2326.

Cirin, R. (2014). Do Academies Make Use of Their Autonomy? Department for Education, Research Report 366, London.

Cohodes, S. and K. Parham (2021). Charter Schools' Effectiveness, Mechanisms, and Competitive Influence. Forthcoming, *Oxford Research Encyclopedia of Economics and Finance.*

Cullen, J. B., B. A. Jacob and S. Levitt (2006). The Impact of School Choice on Participants: Evidence From Randomized Lotteries. *Econometrica*, vol. 75(5), 1191-1230.

Dobbie, W. and R.G. Fryer (2013). Getting Beneath the Veil of Effective Schools: Evidence from New York City. *American Economic Journal: Applied Economics*, vol 5(4), 28-60.

Dobbie, W. and R. G. Fryer (2020). Charter Schools and Labor Market Outcomes. *Journal of Labor Economics*, vol. 38(4), 915-957.

Dynarski, S., D. Hubbard, B. Jacob and S. Robles (2018). Estimating the Effects of a Large For-Profit Charter School Operator. NBER Working Paper No. 24428.

Eyles, A., S. Machin, and O. Silva (2017). Academies 2: The new batch. *Fiscal Studies*, *39*(1): 121-158.

Eyles, A. and S. Machin (2019). The Introduction of Academy Schools to England's Education. *Journal of the European Economic Association*, 17(4): 1107-1146.

Fryer, R. G. (2014). Injecting Charter Schools Best Practices into Traditional Public Schools: Evidence from Field Experiments. *Quarterly Journal of Economics*, vol 129(3), 1355-1407.

Gibbons, S, S. Machin and O. Silva (2008). Competition, Choice and Pupil Achievement. *Journal of the European Economic Association*, vol. 6(4), 912-947.

Green, F., R. Allen, and A. Jenkins (2015). Are English free schools socially selective? A quantitative analysis. *British Educational Research Journal*, 41 (6): 907–24.

Hoxby, C. (2000). Does Competition Among Public Schools Benefit Students and Taxpayers? *American Economic Review*, 90(5), 1209-38.

Jackson, C. K. (2018). What do test scores miss? The importance of teacher effects on non–test score outcomes. *Journal of Political Economy*, 126(5), 2072-2107.

Terrier, C., S. Machin, S. McNally, and G. Ventura. (2020). *Closing the Gap Between Vocational and General Education? Evidence from University Technical Colleges in England.* IZA Discussion Paper 13837.

Tables

Table 1. Descriptive statistics - pupil-level data.

	(1)	(2)	(3)	(4)	(5)	(6)
Variable	Both schools	NE school	CL school	National sample	NE School 'Catchment Area'	CL School 'Catchment Area'
Panel A. Application patterns						
Offer FS	0.203	0.348	0.098	-	-	-
Start FS	0.152	0.237	0.091	-	-	-
Panel B. School composition						
KS4 score (standardized)	0	-0.192	0.142	-	-	-
Female	0.437	0.427	0.445	0.489	0.487	0.487
FSME	0.185	0.180	0.189	0.183	0.226	0.253
White British	0.399	0.133	0.601	0.797	0.499	0.435
English FL	0.501	0.278	0.671	0.854	0.565	0.522
SEN	0.184	0.188	0.181	0.211	0.250	0.222
KS2 - English Level	4.431	4.281	4.552	4.212	4.147	4.262
KS2 - Maths Level	4.436	4.316	4.530	4.231	4.145	4.296
KS2 - Science Level	4.437	4.316	4.534	4.246	4.150	4.276
KS1 - English Level	15.521	14.969	15.927	14.933	14.158	14.718
KS1 - Maths Level	16.223	15.562	16.763	15.763	14.913	15.667
KS1 - Science Level	15.706	15.027	16.262	15.566	14.764	15.372
Distance	-	-	1.921	-	-	-
Observations	2179	914	1265	1,543,129	12,583	17,177

Notes: 'Catchment Area' refers to the Local Authority (LA) or set of LAs from which each of the two schools attract more than 90% of their applicants.

Table 2. Balancing tests.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Dependent variable:	White	Male	EFL	FSM	SEN	KS2 English	KS2 Maths	KS2 Science	KS1 English	KS1 Maths	KS1 Science
Panel A. Pooling both schools. Observations: 2,112											
Free School Offer	-0.014	-0.040	0.007	0.004	-0.018	0.031	-0.002	0.023	0.187	0.059	0.176
	(0.024)	(0.030)	(0.027)	(0.023)	(0.021)	(0.034)	(0.034)	(0.033)	(0.197)	(0.193)	(0.184)
Joint significance	*0.798*										
Panel B. NE school. Observations: 914											
Free School Offer	-0.009	-0.030	-0.014	0.008	-0.012	0.022	0.006	0.028	0.096	-0.063	0.085
	(0.024)	(0.036)	(0.032)	(0.028)	(0.027)	(0.040)	(0.040)	(0.041)	(0.227)	(0.225)	(0.209)
Joint significance	*0.981*										
Panel C. CL School. Observations: 1,198											
Free School Offer	-0.024	-0.061	0.049	-0.006	-0.030	0.052	-0.020	0.013	0.388	0.326	0.377
	(0.052)	(0.052)	(0.049)	(0.039)	(0.035)	(0.062)	(0.064)	(0.057)	(0.384)	(0.369)	(0.370)
Joint significance	*0.733*										
Cohort-by-experiment dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort-specific running variable controls	CL only	CL only	CL only	CL only	CL only	CL only	CL only	CL only	CL only	CL only	CL only

Notes: each coefficient comes from a different OLS regression. The dependent variable is reported in each column's heading. The specification adopted is the one in Equation (1). Joint significance tests are obtained after jointly estimating the models for all outcomes. Robust standard errors in parentheses. ***: p<0.01, **: p<0.05, *: p<0.1.

Table 3. Free school effectiveness – the impact on KS4 test scores.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Sample:	*Both schools*			*NE school*			*CL school – full sample*			*CL school – lottery only*		
Equation:	Reduced form	1st stage	Structural form	Reduced form	1st stage	Structural form	Reduced form	1st stage	Structural form	Reduced form	1st stage	Structural form
Estimator:	OLS	OLS	TSLS	OLS	OLS	TSLS	OLS	OLS	TSLS	OLS	OLS	TSLS
Dependent variable:	KS4 Score Std.	Years of Exposure	KS4 Score Std.	KS4 Score Std.	Years of Exposure	KS4 Score Std.	KS4 Score Std.	Years of Exposure	KS4 Score Std.	KS4 Score Std.	Years of Exposure	KS4 Score Std.
Free School Offer	0.167***	2.189***		0.171***	1.969***		0.159**	2.675***		0.172**	2.759***	
	(0.044)	(0.135)		(0.055)	(0.163)		(0.075)	(0.234)		(0.081)	(0.250)	
Years of exposure			0.076***			0.087***			0.059**			0.062**
			(0.020)			(0.026)			(0.029)			(0.030)
Kleibergen-Paap F statistic		*264.6*			*146.6*			*131*			*121.9*	
Observations	1,914	1,914	1,914	852	852	852	1,062	1,062	1,062	1,005	1,005	1,005
Cohort-by-experiment dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort-specific running variable controls	Yes	Yes	Yes	No	No	No	No	Yes	Yes	No	No	No
Pupil-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: each coefficient comes from a different OLS or TSLS regression. The dependent variable is reported in each column's heading. The specification adopted is the one in Equation (3). Pupil-level controls include female, White British, EFL, FSM, SEN dummies, as well as the level of end-of-primary scores in English, Maths and Science. The number of observations differs from Table 4 due to the presence of missing values in some pupil-level controls. The Kleibergen-Paap robust first-stage F statistic for the significance of the instrument is reported at the bottom of the table. Robust standard errors in parentheses. ***: p<0.01, **: p<0.05, *: p<0.1.

Table 4. Addressing truncation - Cohort 2.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Sample:		*Both schools*			*NE school*			*CL school – lottery only*	
Equation:	Reduced form	1st stage	Structural form	Reduced form	1st stage	Structural form	Reduced form	1st stage	Structural form
Estimator:	OLS	OLS	TSLS	OLS	OLS	TSLS	OLS	OLS	TSLS
Dependent variable:	KS4 Score Std.	Years of Exposure	KS4 Score Std.	KS4 Score Std.	Years of Exposure	KS4 Score Std.	KS4 Score Std.	Years of Exposure	KS4 Score Std.
Panel A. Offer dummy as IV									
Free School Offer	0.167***	2.334***		0.165**	1.955***		0.175	3.213***	
	(0.063)	(0.178)		(0.078)	(0.205)		(0.120)	(0.352)	
Years of exposure			0.072**			0.084**			0.055
			(0.026)			(0.037)			(0.038)
Kleibergen-Paap F statistic		*172.3*			*83.40*			*83.40*	
Panel B. Lottery number as IV									
Lottery number	-1.766	-0.150		-2.736	-0.278		-1.149	-0.069	
	(0.202)***	(0.082)*		(0.359)***	(0.140)**		(0.240)***	(0.069)	
Years of exposure			0.085			0.101			0.060
			(0.046)*			(0.048)**			(0.102)
Kleibergen-Paap F statistic		*76.67*			*58.12*			*22.91*	
Observations	1,138	1,138	1,138	364	364	364	774	774	774
Cohort-by-experiment dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pupil-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: each coefficient comes from a different OLS or TSLS regression. The dependent variable is reported in each column's heading. The specification adopted is the one in Equation (3) for Panel A, while Panel B uses the lottery number (percentile rank varying between 0 and 1) instead of the offer dummy as instrument for years of exposure. Pupil-level controls include female, White British, EFL, FSM, SEN dummies, as well as the level of end-of-primary scores in English, Maths and Science. For the CL school, the sample excludes students exposed to RD admission risk. The number of observations differ from Table 4 due to the presence of missing values in some pupil-level controls. The Kleibergen-Paap robust first-stage F statistic for the significance of the instrument is reported at the bottom of each panel. Robust standard errors in parentheses. ***: p<0.01, **: p<0.05, *: p<0.1.

Table 5. Heterogeneous effects on KS4 test scores - by pupil characteristics.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample:	White British	Non-White British	Male	Female	FSM	Non-FSM	High KS2	Low KS2
Equation:	Structural form	Structural form	Structural form	Structural form	Structural form	Structural form	Structural form	Structural form
Estimator:	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS
Panel A: NE school								
Years of exposure	0.102	0.081***	0.117***	0.038	0.058	0.081***	0.066	0.084***
	(0.096)	(0.027)	(0.032)	(0.046)	(0.074)	(0.028)	(0.065)	(0.027)
Observations	112	740	492	360	144	708	347	505
K-P F statistic	*6.010*	*132.2*	*82.31*	*52.65*	*22.38*	*116.1*	*26.55*	*126.5*
P-value of difference	*0.832*		*0.162*		*0.767*		*0.798*	
Panel B: CL school – lottery only								
Years of exposure	0.124***	0.010	0.070	0.055	0.020	0.075**	0.084*	0.051
	(0.045)	(0.046)	(0.043)	(0.045)	(0.100)	(0.032)	(0.049)	(0.036)
Observations	597	408	561	444	193	812	439	566
K-P F statistic	*62.53*	*52.55*	*68.09*	*51.30*	*15.31*	*105.2*	*35.25*	*103.6*
P-value of difference	*0.075*		*0.817*		*0.596*		*0.585*	
Cohort-by-experiment dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pupil-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: each coefficient comes from a different TSLS regression. The dependent variable is the standardised KS4 score, and the sample considered is listed in the column heading. The specification adopted is the one in Equation- (3). Pupil-level controls include female, White British, EFL, FSM, SEN dummies, as well as the level of end-of-primary scores in English, Maths and Science. For the CL school, the sample excludes students exposed to RD admission risk. The Kleibergen-Paap (K-P) robust first-stage F statistic for the significance of the instrument is reported at the bottom of each panel. Robust standard errors in parentheses. ***: p<0.01, **: p<0.05, *: p<0.1

Table 6. Mechanisms: effects on behavioural outcomes.

	(1)		(2)	
Sample:	*NE school*		*CL school – lottery sample*	
Equation:	Structural form		Structural form	
Estimator:	TSLS		TSLS	
Outcome variable:	Coefficient	Std.Err.	Coefficient	Std.Err.
Year 7 and KS4 in the same school	0.273***	0.076	0.241***	0.079
Share of classes missed for unauthorised absences	-0.001*	<0.001	-0.002*	0.001
Dummy - positive share of classes missed for unauthorised absences	-0.027*	0.016	-0.033*	0.018
Average yearly # of sessions lost due to disciplinary exclusions	0.014	0.095	0.261	0.173
Dummy - positive # of sessions lost due to disciplinary exclusions	0.024*	0.013	0.019	0.014
Observations	883		1,055	
Cohort-by-experiment dummies	Yes		Yes	
Cohort-specific running variable controls	No		No	
Pupil-level controls	Yes		Yes	

Notes: each coefficient comes from a different TSLS regression. The dependent variable is listed in each row, and the sample considered is listed in the column heading. The endogenous variable is "starting Year 7 in the school" for the "same school" regression, and "years of exposure to free schools" for the other variables. The specification adopted is the one in Equation (3). Pupil-level controls include female, White British, EFL, FSM, SEN dummies, as well as the level of end-of-primary scores in English, Maths and Science. For the CL school, the sample excludes students exposed to RD admission risk. Robust standard errors in parentheses. ***: p<0.01, **: p<0.05, *: p<0.1.

Table 7. Mechanisms: the characteristics of the counterfactual schools.

	(1)		(2)	
Sample:	*NE school*		*CL school – lottery sample*	
Equation:	Structural form		Structural form	
Estimator:	TSLS		TSLS	
Outcome variable:	Coefficient	Std.Err.	Coefficient	Std.Err.
Panel A. Demographics				
% Female pupils	-8.70***	2.40	-6.40	3.90
% FSM pupils	-4.165***	1.369	-0.235	1.647
% White British pupils	-20.919***	4.097	16.127***	1.857
Panel B. Features of the teaching body				
Pupil/teacher ratio	-2.795***	0.218	0.524*	0.273
% male teachers	0.217	1.122	12.013***	1.384
% qualified teachers	5.074***	1.027	-18.573***	0.536
Gross salary (£)	-2,995***	306	9,227***	254
Teacher turnover	0.039***	0.009	-0.063***	0.012
Headteacher turnover	-0.033	0.024	0.843***	0.035
Panel C. School institutional features				
Mixed gender school	0.171***	0.043	0.310***	0.057
Academy/free school	0.413***	0.067	0.581***	0.058
Faith school	-0.309***	0.053	-0.474***	0.055
Selective (grammar) school	0.009	0.026	-0.023	0.021
Panel D. Ofsted reports				
Overall grade	-1.408***	0.166	0.280***	0.085
Management	-1.225***	0.154	0.311***	0.077
Behaviours	-1.206***	0.130	-0.642***	0.079
Teaching	-1.333***	0.157	0.140*	0.084
Achievement	-1.408***	0.166	0.262***	0.085
Observations	837		859	
Cohort-by-experiment dummies	Yes		Yes	
Cohort-specific running variable controls	No		No	
Pupil-level controls	Yes		Yes	

Notes: each coefficient comes from a different TSLS regression. The dependent variable is listed in each row, and the sample considered is listed in the column heading. The regressions exclude pupils for whom the Year 7 school identifier is missing. The specification adopted is the one in Equation (3). Pupil-level controls include female, White British, EFL, FSM, SEN dummies, as well as the level of end-of-primary scores in English, Maths and Science. For the CL school, the sample excludes students exposed to RD admission risk. Robust standard errors in parentheses. ***: p<0.01, **: p<0.05, *: p<0.1.

Table 8. Text analysis on schools' 'Vision and Ethos Statements' - pupil-level counterfactual school regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A. NE school.	No excuses	Strict routines	Culture of success	Good manners	Hard work	Trust	Fairness	Autonomy	Purpose	Mastery
Start at FS	0.995***	1.526***	0.482***	-0.092	1.439***	2.131***	1.119***	-0.801***	-0.286	1.311***
	(0.157)	(0.130)	(0.158)	(0.163)	(0.141)	(0.038)	(0.146)	(0.136)	(0.181)	(0.149)
Panel B. CL School – lottery only	Classic liberal	Knowledge-rich	Compet. Atmosphere	Teacher-led learning	Politeness	Debate	High aspirations	Kindness	Hard work	High standards
Start at FS	2.952***	1.293***	2.890***	3.032***	0.856***	2.714***	-0.244*	0.623***	-0.006	-0.280***
	(0.023)	(0.074)	(0.028)	(0.001)	(0.093)	(0.055)	(0.132)	(0.106)	(0.094)	(0.091)
Cohort-by-experiment dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pupil-level control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: each coefficient comes from a different TSLS regression focussing on whether pupils start at one of the two free schools instruments with offer from that school. The dependent variable is an index measuring excess occurrence of a key work (and its synonyms) in the free school statement relative to expected occurrence in body of statements and adjusting for the 'complexity' (rare occurrence) of the words used in the statements. Index values have been standardized. Different key words (and their synonyms) are analysed as reported in each column's heading. The specification adopted is the one in Equation (3). For the CL school, the sample excludes students exposed to RD admission risk. Pupil-level controls include female, White British, EFL, FSM, SEN dummies, as well as the level of end-of-primary scores in English, Maths and Science. Robust standard errors in parentheses. ***: p<0.01, **: p<0.05, *: p<0.1.

APPENDIX – DATA DESCRIPTION

A1. Characterising schools' pedagogical approaches

The goal of our quantitative text analysis exercise is to characterise the pedagogical approaches used by the two free school considered in our analysis and compare them to the educational approaches used by: *i*- the set of counterfactual schools attended by students who do not enrol in one of the two free schools; and *ii*- the set of other secondary, non-religious free schools that existed during the period of our investigation.

To do so, we web-scraped the 'vision and ethos' (VE) statements for these schools as well as 'letters of welcome' from their headteachers (most of the time the former were hyper-linked in the latter). We carried out the scraping for the counterfactuals in Summer/Autumn 2022, and in Winter 2023 for the other free schools. All the schools we considered had a website and had posted VE statements and letters of welcome. In total, our samples include: the NE school and its 52 counterfactuals; the CL school and its 96 counterfactuals; and a set of 102 other free schools that we compare to either the NE or the CL school.

As a starting point for the investigation, we created a set of 'concepts' that characterise the pedagogical approaches of the two free schools. These are identified using key words – including core values and drivers – and key expressions in the VE statements that are recurrent and chime with the insights we obtained from our school visits and semi-structured interviews. For the NE school, these concepts are: 'no excuses', 'strict routines', 'culture of success', 'good manners', 'hard work', 'trust', 'fairness' (the latter three are core values), 'autonomy', 'purpose' and 'mastery' (the last three are drivers). For the CL school, the concepts are: 'classic liberal', 'knowledge rich', 'competitive atmosphere', 'teacher-led learning', 'politeness', 'debate', 'high aspirations', 'kindness' and 'hard work' (the last three are core values).

We then created an extended 'dictionary' that associates such concepts to a wide range of synonyms and akin expressions that can be 'looked for' in the statements of the counterfactuals and other free schools. The extent of synonyms/akin expressions varies considerably between concepts. For example, for the NE school, we identify 47 alternative expressions for 'culture of success', but only 8 for 'trust'. For the CL school, we find 55 akin expressions for 'high standards', but 6 for 'competitive atmosphere'. The full dictionaries can be accessed using the links provided in the Online Appendix. To clarify our approach, here we provide some examples. Starting with NE school, the 'no excuse' concept is associated with expressions that relate to 'signing a contract' with parents or students; not taking shortcuts; not making exceptions; 'doing whatever it takes'; and doing 'what we say we will do'. 'Strict routines' is

instead associated to basic/clear/simple/strict rules and protocols; expressions that emphasise the importance of rules/routines/protocols; and expressions that recall the idea of 'doing simple things well every day'. Finally, 'culture of success' is a very large field that includes '100% everyday'; '100% excellence'; '100% success' (and similar); 'achieving the best'; 'be the best'; 'cool to be smart'; 'reaching for the stars'; and several variations of expressions including the words high/excellent and standards/achievements (or similar). For the CL school, the 'classic liberal' concept includes expressions such as 'Latin'; ancient/past/great associated with scholars/thinkers/philosophers; 'the classics'; and various expressions using the word philosophy. 'Knowledge rich' instead considers as akin expressions (amongst others) 'deep knowledge'; 'knowledge based'; deep or rich understanding (and similar); 'specify and excel'; 'power of reason'; and 'critical thinking'. Finally, 'teacher-led learning' is linked with the following ideas: assess, test, and evaluate, associated with feedback or report; pupil-teacher discussions, dialogues or interactions; and teacher-focussed learning or lessons.

Next, we used these dictionaries to search for the set of synonyms/akin expressions in the VE statements of the counterfactual schools and the other free schools of the time – and any time a synonym/akin expression was identified we replaced it with the relevant key concept identifying the NE and CL pedagogical approaches listed above. For example, we substituted 'signing a contract' with 'no excuse'; or 'critical thinking' with 'knowledge rich'.

Prior to doing this, all VE statements (and dictionaries) were lemmatised, harmonised (e.g., we deleted punctuation and transform all fonts into lower-case) and stop-words were removed (we took care of not removing stop-words when these would change the meaning of our key concepts – for example the word 'no' in the expression 'no excuses'). Importantly, to only identify legitimate changes, we carried out the following two checks. First, we made sure that all bigrams containing some of our single-word concepts that obviously refer to 'spurious' ideas were not replaced by our dictionary search. For example, the word 'trust' when associated with 'multi-academy' or a trust's specific name (e.g., Harris or Dixons) is by-passed by our re-classification of synonyms/akin expressions. Second, we manually checked all the expressions that were replaced by our algorithm against the original VE statements – and manually recoded changes that were not 'legitimate', so that they are not included in our reclassification. Across the four possible comparisons carried out in our text analysis (NE vs. counterfactuals; CL vs. counterfactuals; NE vs. other free schools; and CL vs. other free schools), we manually recoded only 4.5-5% of such inappropriate changes out of 1,200-2,200 total changes, depending on the sample considered. These figures are reassuring and suggest that our dictionaries are wide

enough to identify several synonyms/akin expressions but not so broad that we pick up many unrelated concepts.[1]

Finally, we created two indices that, for each school, indicate whether a key concept is 'overrepresented' in the VE statement of that school with respect to its prevalence across the set of VE statements being considered. More formally (and focussing, for example, on the comparison between the NE school and its counterfactuals), we define the following quantities: the document *d* is the set of statements *s* included in the analysis of the NE school and its counterfactuals; n_d^s is the number of statements *s* in the document *d* (*d* = 53 for the NE school and its counterfactuals); n_s^i is the set of words/expressions associated to concept *i* recorded in statement *s*; n_d^i is the number of those words/expressions in document *d*; N_s is the total number of words in statement *s*; N_d is the total number of words in the document *d*; and finally f_{is} and f_{id} are the counts of words/expressions for concept *i* in statement *s* and in the whole document *d*, respectively.

We then define $X_{is} = \frac{f_{is}}{Ns}$ to measure the frequency of concept *i* in statement *s*; and $\tilde{X}_i = \frac{\sum_1^{n_d^s} X_{is}}{n_d^s}$ to measure the (unweighted) average frequency of concept *i* across the statements in the document *d* – that is, the expected frequency of that concept in the document. Our first index is then defined as $EX_{is} = \frac{X_{is}}{\tilde{X}_i}$ and measures the 'excess frequency' of concept *i* in statement *s* (i.e., the ratio of the actual frequency to expected frequency).

In order to construct our second index, we further define the statement-level average excess frequency across concepts as $EX_s = \frac{\sum_1^{n_s^i} EX_{is}}{n_s^i}$ – i.e., a measure of the excess frequency across all concepts *i* within statement *s*. This measure captures the 'complexity' of the words used in statement *i*. Our second complexity-adjusted index is then defined as $\widetilde{EX}_s^i = \frac{EX_{is}}{EX_s}$, which adjusts the excess frequency of concept *i* in statement *s* for the average statement complexity. The intuition for this adjustment is that the expression 'no excuse' (for example) should carry more weight if the statement only uses other words that are common (e.g., pupils, teachers, school) than if the statement only comprises salient (and thus infrequent) concepts (e.g., strict routines, autonomy, purpose, and mastery).

[1] Note also that we carried out this process iteratively – that is, we went back and forth between scrutinising the VE statements of counterfactuals/other free schools and adding synonyms/akin expressions to the dictionaries based on new expressions/words we identified when carrying out our checks. Though this approach was time consuming, it resulted in a classification that we can trust as it has been completely 'ground-truth-ed'.

We construct such indexes for our four comparisons – i.e., NE school vs. counterfactuals; CL school vs. counterfactuals; NE school vs. other free schools during the observation window; and CL schools vs. other free schools. All these indexes vary at the school level. Their descriptive statistics are presented in Appendix Tables 2 and 3 and discussed in the main text (Section 6.2 and Conclusions).

APPENDIX – TABLES AND EXHIBITS

Appendix Table 1. Free school effectiveness – Core subjects breakdown.

	(1)		(2)	
Sample:	*NE school*		*CL school – lottery sample*	
Equation:	Structural form		Structural form	
Estimator:	TSLS		TSLS	
Outcome variable:	Mean of outcome	Free school effects	Mean of outcome	Free school effects
Panel A. English				
Average grade	5.14	0.168***	5.67	0.114**
Pass	0.80	0.023*	0.89	0.013*
Achieve top level	0.23	0.039***	0.32	0.028*
Panel B. Mathematics				
Average grade	5.12	0.111**	5.53	0.139***
Pass	0.80	0.005	0.86	0.009
Achieve top level	0.27	0.027**	0.34	0.042***
Panel C. Science				
Average grade	5.12	0.027**	5.20	0.077
Pass	0.74	0.036**	0.71	0.022
Achieve top level	0.11	0.043**	0.17	0.011
Panel D. Exams				
Number of exams entered	9.07	0.040	8.71	-0.121
Cohort-by-experiment dummies	Yes		Yes	
Cohort-specific running variable controls	No		No	
Pupil-level controls	Yes		Yes	

Notes: each coefficient comes from a different TSLS regression. The dependent variable is listed in each row, and the sample considered is listed in the column heading. Number of observations varies by outcomes. For English and Mathematics: 845 and 829 pupils for NE and CL, respectively. For Science, 322 and 573 for NE and CL, respectively (due to data issues only one cohort can be considered). For number of exams: 892 and 1136 for NE and CL, respectively. The specification adopted is the one in Equation (3). Pupil-level controls include female, White British, EFL, FSM, SEN dummies, as well as the level of end-of-primary scores in English, Maths and Science. For the CL school, the sample excludes students exposed to RD admission risk. The first column presents mean outcomes for counterfactual students with no offer from one of the two free schools. The second column reports coefficients and significance levels from robust standard errors. ***: p<0.01, **: p<0.05, *: p<0.1.

Appendix Table 2. Descriptive statistics – Text analysis on schools' 'Vision and Ethos Statements'.

Variable	(1) Index free school	(2) Mean index counter.	(3) Median index counter.	(4) Share of non-zero index	(5) Free school rank	(6) Alternative index rank
Panel A: No Excuse (NE) school						
No excuses	4.192	-0.081	-0.281	7.69	1	3
Strict routines	5.165	-0.09	-0.285	9.62	1	2
Culture of success	1.837	-0.035	-0.208	98.08	3	11
Good manners	0.783	-0.011	-0.108	86.54	10	10
Hard work	3.717	-0.071	-0.296	92.31	1	5
Trust	4.829	-0.093	-0.469	36.54	1	2
Fairness	2.956	-0.057	-0.304	55.77	2	5
Autonomy	-0.843	0.016	-0.201	90.38	44	42
Purpose (high expect.)	-0.213	0.004	-0.118	84.62	28	24
Mastery	4.138	-0.080	-0.423	28.85	2	3
Panel B: Classic Liberal (CL) school						
Classic liberal	8.856	-0.093	-0.141	2.11	1	1
Knowledge-rich	4.866	-0.051	-0.291	10.53	2	4
Compet. atmosphere	6.561	-0.069	-0.263	7.37	1	2
Teacher-led learning	9.695	-0.102	-0.102	0.00	1	1
Politeness	2.082	-0.022	-0.255	86.32	4	15
Debate	6.288	-0.066	-0.340	12.63	1	2
High aspirations	-0.180	0.002	-0.095	71.58	49	55
Kindness	1.808	-0.019	-0.222	64.21	8	20
Hard work	0.120	-0.001	-0.293	82.11	33	45
High standards	0.470	-0.005	-0.268	94.74	24	47

Notes: Index refers to excess occurrence of a key work (and its synonyms) in the free school statement relative to expected occurrence in body of statements. Alternative index corrects for the 'complexity' (rare occurrence) of the words used in the statements. Index values have been standardized. Observations. Panel A: 53 schools for the NE set (1 free school and 52 counterfactuals). Panel B: 97 schools for the CL set (1 free school and 96 counterfactuals).

Appendix Table 3. Representativeness of NE and CL pedagogical practices among free schools

Variable	Index free school (FS)	Mean index other FS	Median index other FS	Share of non-zero index	Free school rank	Alternative index rank
Panel A: No Excuse (NE) school						
No excuses	6.653	-0.066	-0.245	12	1	2
Strict routines	7.196	-0.072	-0.177	3	1	2
Culture of success	1.747	-0.017	-0.360	97	5	13
Good manners	0.530	-0.005	-0.299	90	21	20
Hard work	5.829	-0.058	-0.283	93	1	2
Trust	4.312	-0.043	-0.506	42	2	2
Fairness	3.347	-0.033	-0.316	54	1	6
Autonomy	-0.710	0.007	-0.146	91	75	70
Purpose (high expect.)	-0.135	0.001	-0.281	91	44	41
Mastery	6.332	-0.063	-0.359	19	1	1
Panel B: Classic Liberal (CL) school						
Classic liberal	10.00	-0.099	-0.099	0	1	1
Knowledge-rich	4.610	-0.046	-0.329	14	2	3
Compet. atmosphere	3.927	-0.039	-0.280	9	4	5
Teacher-led learning	9.614	-0.095	-0.123	1	1	1
Politeness	1.525	-0.015	-0.239	84	7	24
Debate	7.781	-0.077	-0.259	10	1	1
High aspirations	-0.592	0.006	-0.274	93	69	74
Kindness	1.229	-0.012	-0.332	61	12	22
Hard work	-0.240	0.002	-0.251	93	50	71
High standards	-0.012	0.000	-0.314	97	40	57

Notes: The descriptive statistics are based on quantitative text analysis of the 'Ethos and Vision' statements of the two schools under analysis and the other secondary, non-special, non-religious free schools that existed up to the year 2016. Index refers to excess occurrence of a key work (and its synonyms) in the free school statement relative to expected occurrence in body of statements. Alternative index corrects for the 'complexity' (rare occurrence) of the words used in the statements. Index values have been standardized. Observations. 102 schools in both panels.

Appendix Exhibit 1

Questionnaire for semi-structured interviews conducted at the schools in person.

Interviews conducted by O. Silva and G. Heller-Sahlgren on 30th June 2022 (NE school) and 5th July 2022 (CL school).

Narrative: Thanks for sharing your application data – we have now analysed the evidence and we will now briefly present our approach and findings.

Note: Discuss the results without giving too much detail on heterogeneous effects.

Narrative: We would like now to discuss with you what difference you think your school makes for your students, how do you think these differences come about and who benefits the most from attending your school.

Narrative: What we say will be used in guiding our further analysis and we might quote – with your permission – some of what we say. We can of course send you any quote before we use it so that you can approve it and correct it.

Before we start, I would like to ask permission to record our conversation.

Q1: If you were to think about the characteristics of the school that bring about these differences, what would you say are the most important features?

> *Hint for the interviewer:* follow up, with specific questions about the role of class size/peers/teachers (salary and qualifications)/institutional differences (such as autonomy) – depending on what is mentioned.

Q2: Would you say that these differences really make your school stand-out from the alternative schools children could choose? Which one is a defining feature?

> *Hint for the interviewer:* we are trying to get at the counterfactual schools.

Narrative: When presenting our results, we have discussed the ethos of your school as clearly NE/CL [choose] on the basis of the vision statements that we found on your school website.

Q3: Would you say that this is indeed an appropriate way to characterise your school's teaching philosophy?

Q4: If you were to explain what NE/CL is, how would you describe this teaching approach?

> *Hint for the interviewer:* this question is active only if they do not describe their philosophy when answering the previous question.

Q5: If you were to use three to five key words to summarise this approach, what would they be?

Q6: Why was this approach chosen?

Q7: Do you think this approach was designed because it benefits students at large or in order to cater for the specific needs of some subset of pupils?

Narrative: We are also curious to know whether you think your school – and its pedagogical approach – has a different impact on students with different backgrounds.

Q8: Who do you think is more positively affected by attending your school?

> *Hint for the interviewer:* we should not disclose our groups. But if they do not go in that direction, we should specifically ask White/non-White; boys/girls; FSME or not; high/low achievers.

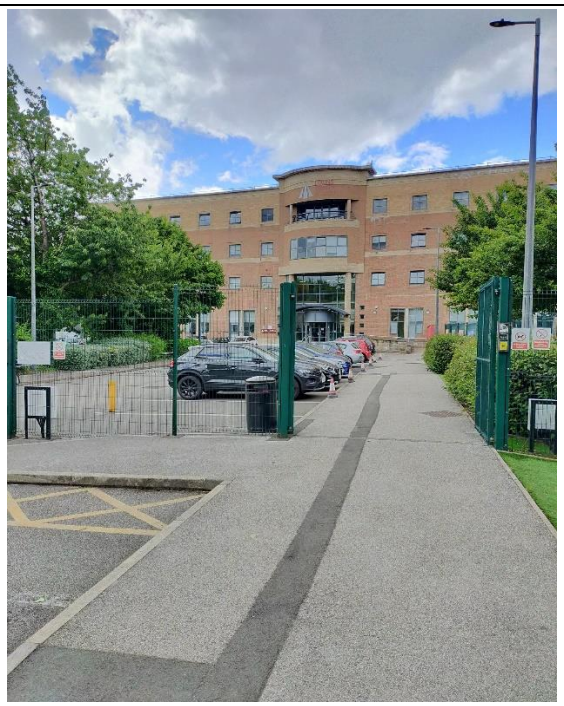
According to our analysis, we find that… [discuss our evidence]

Q9: Are you surprised by these patterns?

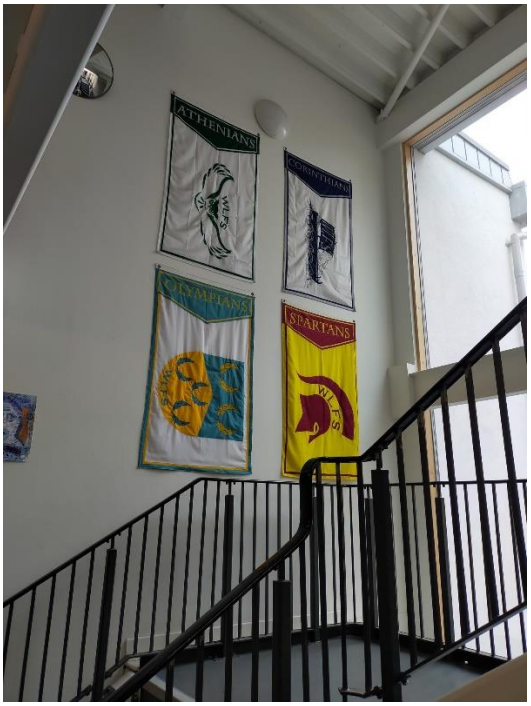
Q10: Would you think that these heterogenous effects arise from the specific pedagogical approach of your school?

Final question: is there anything else you would like to add that might improve our analysis?

Appendix Exhibit 2 – Panel A – Pictures from the NE school visit



Appendix Exhibit 2 – Panel B – Pictures from the CL school visit



ONLINE APPENDIX - TABLES

Online Table A1. Balancing tests - Cohort 1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Dependent variable:	White	Male	EFL	FSME	SEN	KS2 English	KS2 Maths	KS2 Science	KS1 English	KS1 Maths	KS1 Science
Panel A. Pooling both schools. Observations: 798											
Free School Offer	0.025	-0.073	-0.005	-0.007	-0.036	0.051	-0.000	0.097**	0.240	0.082	0.228
	(0.035)	(0.044)	(0.040)	(0.031)	(0.031)	(0.050)	(0.050)	(0.049)	(0.288)	(0.284)	(0.273)
Joint significance	*0.321*										
Panel B. NE school. Observations: 511											
Free School Offer	0.008	-0.040	-0.043	0.028	-0.043	0.035	0.018	0.109*	0.003	-0.021	0.056
	(0.039)	(0.054)	(0.048)	(0.042)	(0.038)	(0.057)	(0.055)	(0.060)	(0.333)	(0.336)	(0.302)
Joint significance	*0.802*										
Panel C. CL School. Observations: 287											
Free School Offer	0.059	-0.138*	0.072	-0.079**	-0.022	0.090	-0.043	0.068	0.782	0.318	0.621
	(0.072)	(0.076)	(0.071)	(0.040)	(0.050)	(0.098)	(0.102)	(0.086)	(0.561)	(0.533)	(0.576)
Joint significance	*0.246*										
Cohort-by-experiment dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort-specific running variable controls	CL only	CL only	CL only	CL only	CL only	CL only	CL only	CL only	CL only	CL only	CL only

Notes: each coefficient comes from a different OLS regression. The dependent variable is reported in each column's heading. The specification adopted is the one in Equation (1). Joint significance tests are obtained after jointly estimating the models for all outcomes. Robust standard errors in parentheses. ***: p<0.01, **: p<0.05, *: p<0.1.

Online Table A2. Balancing tests - Cohort 2.

Dependent variable:	(1) White	(2) Male	(3) EFL	(4) FSME	(5) SEN	(6) KS2 English	(7) KS2 Maths	(8) KS2 Science	(9) KS1 English	(10) KS1 Maths	(11) KS1 Science
Panel A. Pooling both schools. Observations: 1,315											
Free School Offer	-0.049	-0.011	0.017	0.013	-0.002	0.013	-0.004	-0.041	0.142	0.039	0.134
	(0.031)	(0.040)	(0.036)	(0.033)	(0.030)	(0.046)	(0.046)	(0.045)	(0.270)	(0.263)	(0.249)
Joint significance	*0.834*										
Panel B. NE school. Observations: 403											
Free School Offer	-0.025	-0.022	0.011	-0.008	0.015	0.010	-0.006	-0.045	0.173	-0.097	0.108
	(0.031)	(0.049)	(0.042)	(0.039)	(0.038)	(0.056)	(0.056)	(0.055)	(0.312)	(0.304)	(0.290)
Joint significance	*0.882*										
Panel C. CL School. Observations: 912											
Free School Offer	-0.098	0.010	0.030	0.058	-0.036	0.018	-0.001	-0.034	0.077	0.331	0.188
	(0.072)	(0.071)	(0.068)	(0.064)	(0.049)	(0.080)	(0.081)	(0.076)	(0.526)	(0.511)	(0.482)
Joint significance	*0.909*										
Cohort-by-experiment dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort-specific running variable controls	CL only	CL only	CL only	CL only	CL only	CL only	CL only	CL only	CL only	CL only	CL only

Notes: each coefficient comes from a different OLS regression. The dependent variable is reported in each column's heading. The specification adopted is the one in Equation (1). Joint significance tests are obtained after jointly estimating the models for all outcomes. Robust standard errors in parentheses. ***: p<0.01, **: p<0.05, *: p<0.1.

Online Table A3. Balancing tests - Lottery number instruments (Cohort 2 only).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Dependent variable:	White	Male	EFL	FSM	SEN	KS2 English	KS2 Maths	KS2 Science	KS1 English	KS1 Maths	KS1 Science
Panel A. NE school. Observations: 402											
Lottery number	-0.008	0.044	-0.058	0.045	-0.013	-0.056	0.050	0.136	-0.323	0.365	-0.267
	(0.049)	(0.087)	(0.074)	(0.071)	(0.067)	(0.098)	(0.097)	(0.101)	(0.567)	(0.545)	(0.545)
Joint significance	*0.818*										
Panel B. CL School. Lottery-risk students only. Observations: 871											
Lottery number	0.106	-0.222***	-0.128**	-0.044	-0.028	-0.018	-0.038	-0.059	-0682	-0.497	-0.741
	(0.059)	(0.069)	(0.064)	(0.059)	(0.057)	(0.085)	(0.087)	(0.082)	(0.542)	(0.522)	(0.496)
Joint significance	*0.153*										
Cohort-by-experiment dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: each coefficient comes from a different OLS regression. The dependent variable is reported in each column's heading. The specification adopted is the one in Equation (1). Joint significance tests are obtained after jointly estimating the models for all outcomes. Robust standard errors in parentheses. ***: p<0.01, **: p<0.05, *: p<0.1.

Online Table A4. Free school effectiveness - Cohort 1 – Impact on KS4 test scores.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Sample:	*Both schools*			*NE school*			*CL school – full sample*			*CL school – lottery only*		
Equation:	Reduced form	1st stage	Structural form	Reduced form	1st stage	Structural form	Reduced form	1st stage	Structural form	Reduced form	1st stage	Structural form
Estimator:	OLS	OLS	TSLS	OLS	OLS	TSLS	OLS	OLS	TSLS	OLS	OLS	TSLS
Dependent variable:	KS4 Score Std.	Years of Exposure	KS4 Score Std.	KS4 Score Std.	Years of Exposure	KS4 Score Std.	KS4 Score Std.	Years of Exposure	KS4 Score Std.	KS4 Score Std.	Years of Exposure	KS4 Score Std.
Free School Offer	0.167***	2.028***		0.177**	1.985***		0.144	2.129***		0.168	2.239***	
	(0.062)	(0.203)		(0.077)	(0.256)		(0.101)	(0.317)		(0.107)	(0.338)	
Years of exposure			0.082***			0.089**			0.068			0.075
			(0.030)			(0.037)			(0.047)			(0.048)
Observations	743	743	743	488	488	488	255	255	255	231	231	231
Kleibergen-Paap F statistic		*99.9*			*59.9*			*45.2*			*43.97*	
Cohort-by-experiment dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort-specific running variable controls	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	No	No	No
Pupil-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: each coefficient comes from a different OLS or TSLS regression. The dependent variable is reported in each column's heading. The specification adopted is the one in Equation (3). Pupil-level controls include female, White British, EFL, FSM, SEN dummies, as well as the level of end-of-primary scores in English, Maths and Science. The number of observations is different from Table A1 due to the presence of missing values in some pupil-level controls. The Kleibergen-Paap robust first-stage F statistic for the significance of the instrument is reported at the bottom of the table. Robust standard errors in parentheses. ***: p<0.01, **: p<0.05, *: p<0.1.

Online Table A5. Free school effectiveness - Cohort 2 – Impact on KS4 test scores.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Sample:	*Both schools*			*NE school*			*CL school – full sample*			*CL school – lottery only*		
Equation:	Reduced form	1st stage	Structural form	Reduced form	1st stage	Structural form	Reduced form	1st stage	Structural form	Reduced form	1st stage	Structural form
Estimator:	OLS	OLS	TSLS	OLS	OLS	TSLS	OLS	OLS	TSLS	OLS	OLS	TSLS
Dependent variable:	KS4 Score Std.	Years of Exposure	KS4 Score Std.	KS4 Score Std.	Years of Exposure	KS4 Score Std.	KS4 Score Std.	Years of Exposure	KS4 Score Std.	KS4 Score Std.	Years of Exposure	KS4 Score Std.
Free School Offer	0.167***	2.334***		0.165**	1.955***		0.171	3.140***		0.175	3.213***	
	(0.063)	(0.178)		(0.078)	(0.205)		(0.110)	(0.326)		(0.120)	(0.352)	
Years of exposure			0.072***			0.084**			0.055			0.055
			(0.026)			(0.037)			(0.036)			(0.038)
Observations	1,171	1,171	1,171	364	364	364	807	807	807	774	774	774
Kleibergen-Paap F statistic		*172.3*			*91.3*			*92.8*			*83.4*	
Cohort-by-experiment dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort-specific running variable controls	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	No	No	No
Pupil-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: each coefficient comes from a different OLS or TSLS regression. The dependent variable is reported in each column's heading. The specification adopted is the one in Equation (3). Pupil-level controls include female, White British, EFL, FSM, SEN dummies, as well as the level of end-of-primary scores in English, Maths and Science. The number of observations is different from Table A2 due to the presence of missing values in some pupil-level controls. The Kleibergen-Paap robust first-stage F statistic for the significance of the instrument is reported at the bottom of the table. Robust standard errors in parentheses. ***: p<0.01, **: p<0.05, *: p<0.1.

Online Table A6. Free school effectiveness – Alternative treatment - KS4 at free school (instead of years of exposure).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Sample:	*Both schools*			*NE school*			*CL school – full sample*			*CL school – lottery only*		
Equation:	Reduced form	1st stage	Structural form	Reduced form	1st stage	Structural form	Reduced form	1st stage	Structural form	Reduced form	1st stage	Structural form
Estimator:	OLS	OLS	TSLS	OLS	OLS	TSLS	OLS	OLS	TSLS	OLS	OLS	TSLS
Dependent variable:	KS4 Score Std.	KS4 at the Free School	KS4 Score Std.	KS4 Score Std.	KS4 at the Free School	KS4 Score Std.	KS4 Score Std.	KS4 at the Free School	KS4 Score Std.	KS4 Score Std.	KS4 at the Free School	KS4 Score Std.
Free School Offer	0.167***	0.408***		0.171***	0.381***		0.159**	0.466***		0.172**	0.493***	
	(0.044)	(0.028)		(0.055)	(0.033)		(0.075)	(0.051)		(0.081)	(0.053)	
KS4 at the Free School			0.409***			0.447***			0.341**			0.348**
			(0.105)			(0.135)			(0.161)			(0.164)
Kleibergen-Paap F statistic		*216.6*			*133.8*			*84.6*			*85.1*	
Observations	1,914	1,914	1,914	852	852	852	1,062	1,062	1,062	1,005	1,005	1,005
Cohort-by-experiment dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort-specific running variable controls	Yes	Yes	Yes	No	No	No	No	Yes	Yes	No	No	No
Pupil-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: each coefficient comes from a different OLS or TSLS regression. The dependent variable is reported in each column's heading. The specification adopted is the one in Equation (3). Pupil-level controls include female, White British, EFL, FSM, SEN dummies, as well as the level of end-of-primary scores in English, Maths and Science. The number of observations differs from Table 4 due to the presence of missing values in some pupil-level controls. The Kleibergen-Paap robust first-stage F statistic for the significance of the instrument is reported at the bottom of the table. Robust standard errors in parentheses. ***: p<0.01, **: p<0.05, *: p<0.1.

Online Table A7. Free school effectiveness - Robustness test on the reduced form effects.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Sample:	*Both schools and cohorts*		*NE school – both cohorts*		*CL school – both cohorts*		*CL school – both cohorts*		
Robustness:	Drop all pupil level controls	KS2 level dummies instead of linear	Drop all pupil level controls	KS2 level dummies instead of linear	Drop all pupil level controls	KS2 level dummies instead of linear	Linear running variable, 1.5*Optimal bandwidth	Quadratic running variable	Cubic running variable
Estimator:	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Dependent variable:	KS4 Score Std.	KS4 Score Std.	KS4 Score Std.	KS4 Score Std.	KS4 Score Std.	KS4 Score Std.	KS4 Score Std.	KS4 Score Std.	KS4 Score Std.
Free School Offer	0.184***	0.178***	0.198***	0.179***	0.157*	0.175***	0.158**	0.174**	0.191**
	(0.052)	(0.044)	(0.062)	(0.055)	(0.096)	(0.075)	(0.078)	(0.076)	(0.077)
Observations	2,078	1,914	882	852	1,196	1,062	1,089	1,074	1,087
Cohort-by-experiment dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort-specific running variable controls	Yes	Yes	No	No	No	No	Yes	Yes	Yes
Pupil-level controls	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes

Notes: each coefficient comes from a different OLS regression. The dependent variable is reported in each column's heading. The specification adopted is the one in Equation (1). Pupil-level controls include female, White British, EFL, FSM, SEN dummies, as well as the level of end-of-primary scores in English, Maths and Science. The number of observations varies by column because of the presence of missing values in the covariates (Columns 1-6) and because of different optimal bandwidth choices (Columns 7-9). Robust standard errors in parentheses. ***: p<0.01, **: p<0.05, *: p<0.1.

Online Table A8. Heterogeneous effects on persistence in the school - by pupil characteristics.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample:	White	Non-White	Male	Female	FSM	Non-FSM	High KS2	Low KS2
Equation:	Structural form	Structural form	Structural form	Structural form	Structural form	Structural form	Structural form	Structural form
Estimator:	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS
Panel A: NE school								
Years of exposure	0.225	0.229***	0.191**	0.360***	0.370**	0.228***	0.115	0.359***
	(0.255)	(0.078)	(0.093)	(0.133)	(0.152)	(0.087)	(0.173)	(0.082)
Observations	121	762	508	375	155	728	355	528
K-P F statistic	*10.82*	*133.4*	*89.10*	*56.04*	*30.59*	*118*	*32.21*	*126.7*
Panel B: CL school – lottery only								
Years of exposure	0.294**	0.172	0.205*	0.239**	0.583***	0.172*	0.253*	0.246**
	(0.119)	(0.112)	(0.105)	(0.119)	(0.196)	(0.090)	(0.138)	(0.096)
Observations	628	427	583	472	206	849	457	598
K-P F statistic	*74.94*	*70.38*	*77.31*	*71.35*	*23.46*	*123.5*	*37.85*	*150.5*
Cohort-by-experiment dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pupil-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: each coefficient comes from a different TSLS regression. The dependent variable is the standardised KS4 score, and the sample considered is listed in the column heading. The specification adopted is the one in Equation (3). Pupil-level controls include female, White British, EFL, FSM, SEN dummies, as well as the level of end-of-primary scores in English, Maths and Science. For the CL school, the sample excludes students exposed to RD admission risk. The Kleibergen-Paap (K-P) robust first-stage F statistic for the significance of the instrument is reported at the bottom of each panel. Robust standard errors in parentheses. ***: p<0.01, **: p<0.05, *: p<0.1.

Online Table A9. Heterogeneous effects on unauthorised absences - by characteristics.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample:	White	Non-White	Male	Female	FSM	Non-FSM	High KS2	Low KS2
Equation:	Structural form	Structural form	Structural form	Structural form	Structural form	Structural form	Structural form	Structural form
Estimator:	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS
Panel A: NE school								
Years of exposure	0.000	-0.001*	-0.000	-0.003*	0.002	-0.001	0.001	-0.002**
	(0.003)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)
Observations	120	738	490	368	152	706	340	518
K-P F statistic	*10.92*	*139.5*	*94.62*	*52.40*	*24.22*	*123.5*	*30.42*	*131.9*
Panel B: CL school – lottery only								
Years of exposure	-0.004**	-0.000	-0.001	-0.003**	-0.006	-0.001*	-0.001	-0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)
Observations	493	383	487	389	197	679	326	550
K-P F statistic	*82.57*	*80.83*	*85.81*	*80.59*	*29.04*	*144.1*	*51.78*	*122.3*
Cohort-by-experiment dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pupil-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: each coefficient comes from a different TSLS regression. The dependent variable is the standardised KS4 score, and the sample considered is listed in the column heading. The specification adopted is the one in Equation (3). Pupil-level controls include female, White British, EFL, FSM, SEN dummies, as well as the level of end-of-primary scores in English, Maths and Science. For the CL school, the sample excludes students exposed to RD admission risk. The Kleibergen-Paap (K-P) robust first-stage F statistic for the significance of the instrument is reported at the bottom of each panel. Robust standard errors in parentheses. ***: p<0.01, **: p<0.05, *: p<0.1.

Online Table A10. Heterogeneous effects on average exclusions - by pupil characteristics.

Sample:	(1) White	(2) Non-White	(3) Male	(4) Female	(5) FSM	(6) Non-FSM	(7) High KS2	(8) Low KS2
Equation:	Structural form	Structural form	Structural form	Structural form	Structural form	Structural form	Structural form	Structural form
Estimator:	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS
Panel A: NE school								
Years of exposure	-0.347	0.070	-0.070	0.233**	0.169	0.000	0.187	-0.023
	(0.297)	(0.099)	(0.152)	(0.112)	(0.282)	(0.095)	(0.169)	(0.114)
Observations	121	762	508	375	155	728	355	528
K-P F statistic	11.14	129.5	85.39	53.88	23.79	117.2	28.45	128.6
Panel B: CL school – lottery only								
Years of exposure	-0.026	0.480	0.305	0.094	-0.395	0.379*	0.255	0.255
	(0.185)	(0.306)	(0.262)	(0.208)	(0.299)	(0.199)	(0.262)	(0.224)
Observations	628	427	583	472	206	849	457	598
K-P F statistic	65.35	56.88	67.24	59.55	18.57	106.9	35.43	112
Cohort-by-experiment dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pupil-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: each coefficient comes from a different TSLS regression. The dependent variable is the standardised KS4 score, and the sample considered is listed in the column heading. The specification adopted is the one in Equation (3). Pupil-level controls include female, White British, EFL, FSM, SEN dummies, as well as the level of end-of-primary scores in English, Maths and Science. For the CL school, the sample excludes students exposed to RD admission risk. The Kleibergen-Paap (K-P) robust first-stage F statistic for the significance of the instrument is reported at the bottom of each panel. Robust standard errors in parentheses. ***: p<0.01, **: p<0.05, *: p<0.1.

Online Table A11. Balancing tests - Excluding students without valid Year 7 school identifiers.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Dependent variable:	White	Male	EFL	FSME	SEN	KS2 English	KS2 Maths	KS2 Science	KS1 English	KS1 Maths	KS1 Science
Panel A. Pooling both schools. Observations: 1,809											
Free School Offer	-0.004	-0.055*	-0.000	-0.006	-0.023	0.029	0.000	0.025	0.135	-0.032	0.167
	(0.025)	(0.032)	(0.028)	(0.025)	(0.023)	(0.035)	(0.035)	(0.035)	(0.208)	(0.203)	(0.192)
Joint significance	*0.793*										
Panel B. NE school. Observations: 865											
Free School Offer	-0.007	-0.044	-0.016	0.001	-0.018	0.019	0.013	0.022	0.027	-0.145	0.032
	(0.025)	(0.038)	(0.033)	(0.029)	(0.028)	(0.041)	(0.041)	(0.042)	(0.236)	(0.232)	(0.216)
Joint significance	*0.957*										
Panel C. CL School. Observations: 944											
Free School Offer	0.004	-0.082	0.038	-0.023	-0.034	0.055	-0.031	0.034	0.408	0.253	0.508
	(0.058)	(0.058)	(0.057)	(0.045)	(0.040)	(0.069)	(0.070)	(0.063)	(0.429)	(0.407)	(0.401)
Joint significance	*0.686*										
Cohort-by-experiment dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort-specific running variable controls	CL only	CL only	CL only	CL only	CL only	CL only	CL only	CL only	CL only	CL only	CL only

Notes: each coefficient comes from a different OLS regression. The dependent variable is reported in each column's heading. The specification adopted is the one in Equation (1). Joint significance tests are obtained after jointly estimating the models for all outcomes. The number of observations differs from Table 2 due to the presence of missing values in pupils' Year 7 school identifiers. Robust standard errors in parentheses. ***: p<0.01, **: p<0.05, *: p<0.1.

Online Table A12. Free school effectiveness on KS4 test scores - Excluding students without valid Year 7 school identifiers.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(7)	(8)	(9)
Sample:	*Both schools*			*NE school*			*CL school – full sample*			*CL school – lottery only*		
Equation:	Reduced form	1st stage	Structural form	Reduced form	1st stage	Structural form	Reduced form	1st stage	Structural form	Reduced form	1st stage	Structural form
Estimator:	OLS	OLS	TSLS	OLS	OLS	TSLS	OLS	OLS	TSLS	OLS	OLS	TSLS
Dependent variable:	KS4 Score Std.	Years of Exposure	KS4 Score Std.	KS4 Score Std.	Years of Exposure	KS4 Score Std.	KS4 Score Std.	Years of Exposure	KS4 Score Std.	KS4 Score Std.	Years of Exposure	KS4 Score Std.
Free School Offer	0.162***	2.360***		0.166***	2.079***		0.154*	3.059***		0.179**	3.131***	
	(0.046)	(0.138)		(0.056)	(0.166)		(0.079)	(0.231)		(0.085)	(0.247)	
Years of exposure			0.069***			0.080***			0.050**			0.057**
			(0.019)			(0.025)			(0.026)			(0.027)
Kleibergen-Paap F statistic		*291.1*			*156.1*			*176.1*			*160.8*	
Observations	1,687	1,687	1,687	811	811	811	876	876	876	828	828	828
Cohort-by-experiment dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort-specific running variable controls	Yes	Yes	Yes	No	No	No	No	Yes	Yes	No	No	No
Pupil-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: each coefficient comes from a different OLS or TSLS regression. The dependent variable is reported in each column's heading. The specification adopted is the one in Equation (3). Pupil-level controls include female, White British, EFL, FSM, SEN dummies, as well as the level of end-of-primary scores in English, Maths and Science. The number of observations differs from Table A7 due to the presence of missing values in pupils' Year 7 school identifiers. The Kleibergen-Paap robust first-stage F statistic for the significance of the instrument is reported at the bottom of the table. Robust standard errors in parentheses. ***: p<0.01, **: p<0.05, *: p<0.1.

ONLINE APPENDIX - DICTIONARIES

Link to the ‘no excuse’ pedagogical approach dictionary: click here.

Link to the ‘classic liberal’ pedagogical approach dictionary: click here.