Centre for Economic Performance

Discussion Paper

ISSN 2042-2695

No. 1979
February 2024

Home broadband and human capital formation

Rosa Sanchis-Guarner
José Montalbán
Felix Weinhardt

The London School of Economics and Political Science

UKRI Economic and Social Research Council

Abstract
Using administrative data, we estimate the effect of home broadband speed on student-level value-added test scores. Our headline estimate relies on jumps in connection quality between close neighbours that occur across thousands of invisible telephone exchange station catchment-area boundaries. We find that increasing speed by 1 Mbit/s increases test scores by 1.37 percentile ranks, equivalent to 5% of a standard deviation. School-level factors or broadband take-up cannot explain this. Instead, the positive effects are concentrated among high-ability and non-free-school-meal eligible students and result from more education-oriented internet use. Differences in ICT quality can thus lead to increasing education inequalities.

Keywords: broadband, education, spatial regression discontinuity
JEL Codes: J24; I21; I28; D83

This paper was produced as part of the Centre's Education and Skills Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

For comments, we thank Antonio Cabrales, Martin Fernández-Sánchez, Alexandra de Gendre, Clément Malgouyres, Stephan Maurer, Marion Monnet, Diego Puga, and participants at several seminars and conference presentations. We thank Prof Sonia Livingston from LSE for kindly providing access to UKCGO survey data. We are especially grateful to Ben Faber, without whom this project would not have materialized. We also thank anonymous reviewers and editors for comments which lead use to completely re-work an earlier and now resting working paper. We explain the differences in sample choices and methods between these two papers in footnote 3. Felix Weinhardt gratefully acknowledges ESRC seed funding (ES/J003867/1) as well as support from the German Science Foundation through project WE5750/1-1 and the CRC TRR 190 (project 280092119). Rosa Sanchis-Guarner acknowledges funding from LSE/CEP, the British Academy and the Generalitat of Catalunya (Programa Serra Hunter and Programa Beatriu de Pinos).
Rosa Sanchis-Guarner, University of Barcelona, IEB, CESifo and Centre for Economic Performance at London School of Economics. José Montalbán, SOFI at Stockholm University. Felix Weinhardt, European University, DIW Berlin, CESifo and Centre for Economic Performance at London School of Economics.

Published by
Centre for Economic Performance
London School of Economic and Political Science
Houghton Street
London WC2A 2AE

1 Introduction

Technological advances in information and communication technology (ICT) have dramatically changed the education landscape over the past two decades. Currently, the educational technology (*edutech* or *edtech*) industry is experiencing a boom. Most researchers, policymakers, educators, and parents agree on the importance of incorporating these technologies into the learning environment. There is however no consensus on their net impact on educational outcomes (Escueta et al., 2017). While home internet quality can increase learning productivity and widen access to educational opportunity, it may also lead to unproductive distraction, making its net effect on student human capital formation ambiguous (Bulman & Fairlie, 2016). Our paper contributes to the limited existing literature by examining how home broadband quality, rather than school-based usage, impacts learning outcomes.[1]

Our key finding, based on administrative data, is that increases in broadband connection speed improve student test scores aligned with the national curriculum, regardless of any specific online focus. Better test-scores, as a measure of learning, are linked to knowledge accumulation and economic development (Angrist et al., 2021).[2] Our results imply that faster broadband at home is an important factor in teenagers' development of human capital, with important implications for future college attendance (Das et al., 2022). Our heterogeneity analysis reveals that these positive effects are driven by high-ability students from non-free school meal backgrounds. We present empirical evidence supporting a productivity-enhancing mechanism, demonstrating positive effects in utilizing faster broadband for educational purposes at home, without any significant impact on broadband take-up or engaging in distracting activities. Furthermore, our research design allows us to control for school-level heterogeneity, ensuring that our estimates are causally linked solely to the home environment.

There are clear challenges in estimating the impact of home internet quality on test scores. To deal with *active* selection, i.e. household choices of broadband packages, a potential strategy would be to use a measure of local available internet speed,

[1] We discuss the related literature in detail below.

[2] The accumulation of human capital, i.e., the stock of skills, traits and knowledge that an individual possesses (Burgess, 2016), is key for growth, employment and earnings (Becker, 1962; Mincer, 1974; Barro, 2001). Educational outcomes, such as the cognitive skills that people have learned, have been found to be a reliable proxy of human capital (Hanushek & Woessmann, 2012, 2015).

which is determined by the distance between the household and the telephone local exchange (LE) station providing them with telephone and internet services (for DSL connections). This would imply comparing outcomes of students whose residences are located at different distances to LE stations and who hence enjoy different potential home internet speed. However, residential distances to the connected LE stations are not randomly assigned across space because stations are located in places with particular location characteristics, potentially leading to *passive* selection. To overcome selection, we use a well-known feature of digital subscriber line (DSL) broadband technology in the design of our estimation strategy: the length of the copper wire that connects residences to the telephone LE station, which is a key determinant of *available* local internet connection speeds.

We focus on the invisible boundaries generated across LE stations. Each station has a catchment area of residential addresses that it serves in its surroundings. The extent and shape of this catchment area is a byproduct of history: rapid growth in fixed-line telephony during and after World War II, in combination with capacity constraints at the exchange switchboards, led to invisible and essentially randomly placed station-level catchment area boundaries. In our strategy, we focus on households whose residence is located in the vicinity of these boundaries, exploiting variation in distances to the connected station across small catchment area boundary segments, each side connected to a different LE station. Using a spatial regression discontinuity (SRD) design, the causal effect of broadband speed on student test scores is identified by comparing cross-sectional variation across "lucky" households that are supplied with faster broadband access (the side with shorter distances on average) to otherwise similar counterparts that were "unlucky", supplied with slower broadband access (the side with longer distances on average). Due to the irregular geographic shape of the boundaries, some households with short cables (long cables) might live on the slower side (faster side). Hence, our SRD design is *fuzzy*, with the *sharp* SRD design affected by attenuation bias. We explain this empirical setup in detail in section 3.

Multiple sets of administrative microdata allow us to meet the extensive data requirements for implementing our empirical strategy. First, we use administrative standardized and externally marked test score records for the population of 14-year-old English students in national Key Stage 3 (KS3) tests over the period 2005–2008, along with rich information on their background characteristics, including the student pre-internet test score taken at a younger age. Another key feature of our data is that we are able to georeference student residence and school at the most granular spatial scale, i.e., the postcode level, which roughly corresponds to blocks of approximately 15 addresses in England. Second, we use telecommunication network data including the position of the universe of English telephone LE stations (approximately 3,900) and their assignments to each of roughly 1.45 million full postcodes. We complement this with postcode-level internet speed measures. Third, we employ a rich vector of georeferenced control variables that allows us to compute residential proximity to a comprehensive list of local amenities. These data include the universe of property transaction values, from which we can construct local average house prices. Forth, to further investigate the mechanisms behind the effect of broadband speed on educational achievement, we use data from the British Household Panel Survey (BHPS), a large longitudinal survey representative of the UK population, covering the years 2005-2008. We use questions regarding internet and broadband access, information regarding computer use at home, and household background characteristics.

Our main finding from this initial SRD analysis is that available broadband speed has positive effects on externally marked test scores that measure learning progress against the national curriculum. We find that moving 100 meters closer to the LE station increases student test scores by 0.122 percentile ranks. These estimates are intent-to-treat (ITT) impacts of the availability of faster broadband connections. Heterogeneity analysis shows that the results are driven by teenagers from more advantaged socio-economic backgrounds and higher ability. This result implies that pre-existing differences are amplified by faster home broadband. In addition, we find that results persist two years later when students are aged 16.

In Section 4, we present a battery of robustness checks to validate our identification strategy and support our conclusions. In particular, the results are robust to controlling for school-specific broadband features, which suggest that the findings are not driven by school characteristics. Notice that, in the most demanding specifications, we can include school fixed effects that capture any school-level heterogeneity in the most flexible way because students from both sides of the boundaries can attend the same school, giving us enough variation to estimate the coefficients.

We are ultimately interested in the relationship between broadband speed and test scores. In Section 5, we combine an additional linked postcode-to-exchange station telecom network dataset with data on average local internet speeds to estimate the distance-speed decay. Consistent with existing estimates (Ahlfeldt et al., 2017), we find that for each additional 100 meters closer to the connected LE station, the local average speed increases by 0.089 Mbit/s. Combining this finding with our previous results implies that for each increase in the average broadband speed of 1 Mbit/s, which corresponds to an increase of 20% with respect to the baseline average speed, test scores increase on average by 1.37 percentile points. This average effect of one additional Mbit/s is equivalent to approximately 5% of a standard deviation in the national test score distribution. The size of this effect can be interpreted as a cumulative effect of having faster internet speed over a period of six years on average. Assuming that effects are linear over time, the average effect is equivalent to an increase of 0.8% of a standard deviation in the national test score distribution per year of exposure to faster internet speed.

In Section 6, we explore mechanisms that may explain our results, with four key findings. First, households on the fast side do not significantly enjoy higher internet usage nor more broadband take-up at home (relative to those on the slow side). This result is important for the interpretation of our findings, i.e. that our estimates relate to changes in the speed of the broadband connection (the intensive margin) and not simply access to high-speed internet (the extensive margin). Second, we find no effect on potentially distracting activities, such as higher use of computers for playing games or hobbies, potentially ruling out the hypothesis of these students decreasing the amount of time spent doing school work due to the more frequent engagement in social media and other distracting activities. Third, we find that teenagers of ages 15-18 are more likely to use the computer for educational work than older members of the household. On top of this, teenagers on the faster side are 27% more likely to use their computer for educational work in comparison with teenagers on the slower side. Our preferred interpretation of the mechanisms driving the positive effect of speed on test scores is a more education-oriented use of the internet combined with limited use of distraction activities.

Our results are relevant in the nowadays digital era in which high-speed broadband internet speed is still a driving force. Currently, 59% of the world population is online (Clement, 2020). While today the average internet speed is faster, the coverage and speed available to English households between 2005 and 2008 are comparable to the infrastructure currently available in large parts of the world. For comparison, in 2019, only approximately 16% of the world population had broadband access (Roser et al., 2020), with average connection speeds of 7.2 Mbit/s (McKeay, 2017). Notably, many families live without broadband connection in developed countries, with only 56% and 51% of low-income families with broadband subscriptions in the US and UK, respectively (Zuo, 2021; Commission et al., 2021). We, therefore, believe that our estimates based on English student population data in the mid-to late-2000s have high external validity for other countries today and can inform current policy.

Our paper is closely related a small but growing literature linking home broadband technology and student test scores.[3] [4] Malamud et al. (2019) find no significant effects of home internet access on student achievement. This result is based on a credible randomized controlled trial giving out laptops and free internet to students in several low-achieving primary schools in Peru in grades 3-5. Our paper expands on Malamud et al. (2019) by identifying effects based on a broader and older student population that covers all socioeconomic levels of school-age teenagers, in a developed country context. Moreover, our estimates are driven by changes in broadband speed, rather than access.

[3] Notice that this paper differs from earlier and resting related working paper by Faber et al. (2015) in several dimensions. Directly following referee suggestions, we now consider only years when the broadband market was already mature in England and when no other confounding technology with broadband was widely available (i.e., 2005 to 2008). This resulted in redefining our treatment as continuous changes in available broadband speed and not on zero-one variation, i.e. faster versus slower. This change in periods and treatment definition enables us to control for pupil-specific (pre-broadband area) KS1 results, allowing to estimate value-added regressions which account for a large fraction of individual heterogeneity across the boundaries and largely improves the precision of the estimates. Besides, we take into account that due to necessary local-level data aggregation the spatial RD that we estimate is fuzzy in nature, which further increases precision (see p. 18, footnote 16 and Table A.2 for discussion of sharp estimates). Taken together, this changes a less precise zero finding as documented in Faber et al. (2015) into the precisely estimated (non-zero) effects that we show in this paper. Additionally, differently to the earlier version, using survey data and our new methodology, we now present direct evidence on mechanisms that drive the reduced-form estimates on test scores. Finally, we provide an estimate of the speed-student performance relationship by means of using Wald-estimate.

[4] A growing literature on the impact of broadband on (other) socioeconomic outcomes includes papers on its positive effects on labor productivity and wages (Akerman et al., 2015), economic growth (Czernich et al., 2011), capitalization of the property market (Ahlfeldt et al., 2017), health choices (C-sections) (Amaral-Garcia et al., 2019), and marriage rates (Bellou, 2015), and negative effects on political participation (Falck et al., 2014; Campante et al., 2018; Gavazza et al., 2019), sex crime (Bhuller et al., 2013) and social capital (Geraci et al., 2022)

In another key study, Dettling et al. (2018) show that students with broadband access in their postal codes perform better on the SAT and apply to a larger set of colleges in the US. This finding is mostly driven by access, rather than quality of the connection and the effect on college applications is interpreted as coming from better information acquisition and a reduction of direct effort in applications. Similar to our findings, positive effects are concentrated in high-SES students. We complement and extend these findings by studying effects of home broadband on existing administrative achievement tests in the school environment, which are specifically designed to test cognitive ability as part of the regular curriculum. Importantly, Dettling et al. (2018) rule out that their findings are driven by investment of the students in their broader academic records, such as improvements in their GPA. In contrast, we find that our positive effect stems from an improvement in academic performance measured by test scores. A main advantage of our outcome of interest, the Key Stage 3 test, is that it is a low-stakes exam from the students' perspective and designed to test student progress in the English education curriculum with no explicit online-training resources. This means that our estimates are informative about the impact of the home environment on the learning and knowledge accumulation that determines human capital formation in a general sense, in contrast to specific ICT skills or targeted preparation for specific test-taking.

A related literature focuses on the relationship between home computer access and education outcomes using quasi-experimental methods as well as randomized interventions to identify the causal effect (Malamud & Pop-Eleches, 2011; Fairlie & Robinson, 2013; Vigdor et al., 2014; Beuermann et al., 2015; Cristia et al., 2017). These articles often report positive effects on outcomes directly related to computer access but no impact – or only a modest or even negative one – on student academic outcomes. Interestingly, Malamud & Pop-Eleches (2011) find that parental rules for computer ownership and homework can mitigate negative effects of computer access on human capital among low-SES children in Romania. In contrast to this literature, we do not estimate effects of (extensive-margin) changes in access to computers but instead of changes in (intensive-margin) broadband speed.

Finally, other papers analyse the effects of ICT and broadband at the school level (Angrist & Lavy, 2002; Rouse & Krueger, 2004; Goolsbee & Guryan, 2006; Machin et al., 2007; Belo et al., 2014; Falck et al., 2018; Lakdawala et al., 2023). Connecting the two, Belo et al. (2016) show that broadband at school can increase broadband usage at home. In this paper, we study effects of home broadband on learning. We view this as a related but different question compared to effects at the school, where for example teachers are present and need to think about how to incorporate the technology into the learning process. This is different to effect of having better broadband at home, which we show in a developed-country context matters for learning outcomes and at the same time amplifies existing inequalities in ability and by socio-economic background.

2 Background, Data and Descriptive Statistics

2.1 Broadband Expansion in England

The empirical setting of this study is England over the time period 2005–2008, where the broadband market was already developed. The rollout of DSL broadband technology in England, as part of the United Kingdom (UK), started in the major urban centers at the beginning of the 2000s and proceeded rapidly. This process involved technological upgrades of the infrastructure of telephone LE stations – the same ones that provide telephony services to a number of connected premises around them – to allow them to offer broadband internet services through copper cable. By the end of 2004, 80% of the LE stations had been equipped to provide broadband services, covering 97% of local residences, which could subscribe to receive broadband services at home. That year, 54% of households had an internet connection, of which 6.2 million (approximately 25%) were broadband. By the start of our estimation period in 2005, 99% of English addresses were connected to broadband-enabled telephone LE stations.

Even if most of the technological upgrades took place between 2000 and 2005, penetration rates were low in the first years (approximately 10% in 2003) and only started growing in 2004. By then, infrastructure was completely rolled out across space, and the take-up rate increased steadily. The broadband internet take-up rate rose from approximately 30% in 2005 to over 60% by 2008 (Eurostat). This increase in take-up was related to decreases in prices and changes in attitudes and internet content. In 2008, 84% of students used the internet, and among those, 90% used it for their homework (Livingstone & Bober, 2005). Due to this, and similarly to existing work in the UK context (Nardotto et al., 2015), we focus on the post-2005 period in our analysis. In this context, we can focus on the impact of broadband speed for a given state of technology and exploit very local variations in its quality.

The relatively fast broadband development in the UK also makes this period an interesting one to study from today's perspective. In 2007, more than half of UK homes had broadband access, with an average connection speed of 4.6 Mbit/s (OfCom, 2009b). While today the average internet speed is faster, the coverage and speed available to English households between 2005 and 2008 are comparable to the infrastructure currently available in large parts of the world. For comparison, in 2019, only approximately 16% of the world population had broadband access (Roser et al., 2020), with average connection speeds of 7.2 Mbit/s (McKeay, 2017). And even today many families live without any broadband connection in developed countries, with only 56% and 51% of low-income families with broadband subscriptions in the US and UK, respectively (Zuo, 2021; Commission et al., 2021).

Note that our period of analysis ends in 2008 for two reasons: (i) mobile broadband and cable/fibre internet technologies became more widespread in the UK after 2009, reducing the efficacy of our empirical approach, and (ii) the standardized exam that we use to measure the educational achievement of teenagers in this paper (the KS3 for students at age 14) was discontinued. The combination of the testing regime and the state of development of the broadband infrastructure in England in the period 2005-2008 offers a unique opportunity to study the effects of home broadband speed on student performance.

2.2 Data

2.2.1 Administrative Student Records

In the English educational system, student academic performance is assessed in national exams that are administered through externally marked tests. The English education curriculum is organized into four key stages (KSs). Compulsory education starts at age 6 and ends at age 16 with the fourth and final KS4 (the General Certificate of Secondary Education [GCSE] examinations). There are several reasons why the KS3 exam at age 14 is the most suitable for our analysis. First, the KS3 test is externally marked and thus comparable across students and schools. Second, the test is low stakes, so there are no incentives for teachers or students that would drive a wedge between test scores and real achievement. Third, the test is finely graded (mostly zero to 100); therefore, in combination with our sample size, it is possible to detect even small effect sizes. Finally, all students are tested in the three main compulsory subjects: English, mathematics and science. Students have very limited options in choosing subjects or specializing according to interest or ability before the KS3, in stark contrast to the educational period before the KS4 test two years later. This feature of the KS3 exam makes it particularly suitable to test for heterogeneity across groups that might later on (endogenously) specialize in different fields.

We employ administrative data containing information on the universe of students enrolled in English state schools (approximately 95% of pupils) who took the KS3 test from 2005–2008 in England. These data are supplied by the Department for Education (DfE). To match the student information with the telecom network data that we describe below, we first use the restricted-access version of the National Pupil Database (NPD), from which we extract the full residential postcode for each registered student in a given year. British postcodes are associated with a small number of addresses (15 on average) and in denser areas usually correspond to housing blocks. In the second step, we use the unique student identifiers to link their residential information to individual test score results, which are also provided as part of the NPD.[5]

Following the education literature, we transform these scores into percentile ranks for each test and cohort, i.e., separately by year-subject. These subject percentiles are then added into a total score, which we percentilize to obtain an average total score ranging from 1 to 100. We conduct this transformation to make our results comparable to other countries' national exams as well as across cohorts/subjects. Transforming raw scores into percentile ranks has the goal of keeping the ordinal information in the outcome variable and removing the cardinal differences between units of interest, which might be driven by the setup of any particular exam paper, for instance.

We use additional data from the DfE NPD for each pupil in our KS3 2005–2008 sample and collect information on their KS1 test scores (taken at age 7). For this sample, this corresponds to tests taken during 1998–2001, when most of the rollout of broadband

[5] The DfE formerly distinguished between the NPD and the Pupil Level Annual School Census (PLASC), which is now treated as part of the NPD. Note that no information is available on private schools, which enroll approximately 6–7% of the English student population (Ryan & Sibieta, 2010).

internet had not yet taken place and the level of broadband take-up was essentially zero. In contrast with the KS3, this test is marked by the schools, is only available for the subjects of mathematics and English, and is graded on a coarse scale. However, adding this information to our empirical models allows us to estimate individual-level value-added results, controlling for pupil-specific time-invariant ability and background characteristics, which in turn improve the precision of the estimates and the explanatory power of the models. We also obtain information on the location, size and type of the school that the pupils attend, which we use to construct school-level controls and, for some specifications, school fixed effects.

In addition to test scores, the administrative data give us access to a series of observable student characteristics, such as gender, ethnicity, and student eligibility for a free school meal (FSME), which is a common proxy for family income. We exploit these data at two scales: to construct individual-level controls and to calculate postcode-year-specific demographics based on the population of pupils of all ages, which we also use as local area control variables in the regressions.

2.2.2 Average House Prices and Area Socioeconomic Characteristics

We use a number of additional datasets to improve precision and to validate our approach. First and foremost, we use transaction-level data on property sales in England over the estimation period. The data are administrative records from the England and Wales Land Registry, covering all property transactions over this period. We use the reported property address information to link these property transaction values to individual residential postcodes. The postcode-year averages are based on several million individual property transactions that occurred in England over the period 2005–2008. Local house prices capitalize many desirable (and undesirable) local attributes and are likely to capture a large number of unobserved spatial characteristics of the areas.

Even though our empirical analysis is based on a spatial discontinuity design that compares only very proximate households, it could still be the case that catchment area boundaries coincide with physical barriers such as roads or rivers and that either the slower or the faster side of the boundaries has a higher likelihood of hosting a given type of local amenity, the combination of which could lead to bias in the boundary effect. Using a GIS with detailed attribute data from the UK Ordnance Survey, the com-

Table 1: Summary Statistics.

	All Sample	Baseline Sample within 300 Meters
	(1)	(2)
A. Outcome Variables		
Average Percentile Rank Score (Mean)	50.21	49.76
	(28.67)	(28.66)
Average Percentile Rank Score in English	50.26	50.21
	(28.59)	(28.51)
Average Percentile Rank Score in Maths	51.60	51.09
	(28.03)	(28.05)
Average Percentile Rank Score in Science	51.50	50.84
	(27.97)	(27.98)
B. Discontinuity Variables		
Distance to the Segment (Meters)	679.13	156.29
	(547.1)	(79.1)
Distance to the LE Station (Meters)	1,511.96	1,866.97
	(862.7)	(878.3)
Share on the "Fast" Side	0.54	0.51
	(0.50)	(0.50)
Average "jump" (Meters)	763.36	930.29
	(659.7)	(598.5)
C. Pupils & School Characteristics		
Distance to School (Meters)	2,602.5	2,413.35
	(3,433.14)	(3,008.3)
White	0.838	0.785
	(0.37)	(0.41)
Male	0.499	0.498
	(0.50)	(0.50)
Free School Meal	0.141	0.155
	(0.35)	(0.36)
Pre-KS3 Score	44.45	43.98
	(24.61)	(24.65)
Number of Schools	2,864	2,610
D. Area Socioeconomic Characteristics		
Share of White Pupils	0.825	0.769
	(0.29)	(0.32)
Share of Free School Meal Pupils	0.151	0.165
	(0.24)	(0.24)
Share of Community Schools	0.638	0.620
	(0.48)	(0.48)
Average House Prices (Pounds)	193,092.6	190,596.8
	(118,208.4)	(110,831.9)
Observations	1,115,594	183,892

Notes: This table shows descriptive statistics for the outcome variables (Panel A), treatment variables (Panel B), pupils and school characteristics (Panel C), and density and area socioeconomics (Panel D). The first column reports statistics for the whole sample of pupils and postcodes. The second column shows statistics for our baseline sample, which are pupils and postcodes located within 300 meters of the LE station boundary segment. Standard deviations are reported in parentheses.

mercial real estate consultancy CBRE and the DfE, we compute euclidean distances between each English postcode and the following features: nearest school (primary or secondary), nearest road (class A, class B and motorways), nearest rail station (which captures centrality), nearest water body (river, stream, marsh or lake) and nearest supermarket. One of the major concerns is that *passive* endogeneity arises because local geography correlates with the location of the LE stations and of households. Taking these variables into account allows us to properly test whether observable geographic features are an endogeneity concern in our setting and ultimately control for these variables in our empirical specification to increase statistical precision.

Finally, we combine data from the Office for National Statistics Postcode Directory (ONSPD) and the DfE to control for local density by calculating the number of premises in each postcode (which is fairly stable over time) and the number of students (of all ages) per premise.

2.2.3 Telecom Infrastructure and Postcode Broadband Speed Data

In order to construct measures of available speed we use telecommunication network information, including the location of the universe of English LE stations (approximately 3,900) and their assignments to each of roughly 1.45 million full postcodes, which we obtained from OfCom, the UK telecom regulator. This data allows us to link each residential address to its connected telephone LE and to allocate each household to a unique LE catchment area. We exploit the richness of this dataset to devise the empirical strategy set out in section 3.

In Section 5, we use additional data on postcode-level realized internet speed from Ofcom, the British telecom regulator in the UK, to estimate the distance-to-LE speed relationship.[6] The major fixed-line broadband suppliers (ISPs) provide data on individual speed tests to Ofcom, which aggregates the information by area in different years. The data from these suppliers cover over 80% of the market.[7] Data at the finest geographical level, the postcode, have been available since 2012 and are published yearly.

Average postcode-level speed is calculated from information on millions of active broadband connections provided to the regulator and are based on *modem sync speed*,

[6] These data are available from the Ofcom Infrastructure reports – now called Connected Nations – accessible via the Ofcom webpage and the National Archives webpages.

[7] The suppliers include BT, Virgin Media, Everything Everywhere, O2, KCom, TalkTalk and Sky.

which captures the highest possible speed at which data can be transferred across the line with the use of a particular DSL technology (OfCom, 2012). The indicator captures the speed at which the modem in a customer's home connects to the equipment in the telephone exchange, and it is directly related to the subscription package headline speed. This way of measuring the line speed contrasts with speed tests obtained using modems at home and performed by users, who usually report slower speeds, which are affected by the time of the day at which the data transfer is done, the number of devices connected simultaneously and the quality of home software and internet equipment. Postcode average speeds are a reflection of the subscriptions to different broadband packages, which results in a mix of technologies: primarily ADSL but also cable or fibre internet.

A potential concern is the deployment of non-distance-sensitive technologies in Britain from 2008 (cable) and 2010 (fibre). Ofcom information for small geographies were only made accessible from 2012, when superfast technologies were already available in some areas. First, note that even if in 2012 68% of England already had access to superfast broadband (yielding speeds over 30 Mbit/s), approximately 75% of the subscribed broadband connections were still using ADSL technology. In this sense, local averages for 2012 are the result for a majority of ADSL connections which are sensitive to distance to the LE. Furthermore, since most packages offered 8–10 Mbit/s headline speeds (OfCom, 2009b) in the period 2005–2008, we focus on postcodes with average (download) speeds that are realistic for this sample period.

2.2.4 Survey Microdata on Broadband Take-up and Computer Use at Home

To further investigate the mechanisms behind the effect of broadband speed on educational achievement, we use data from the British Household Panel Survey (BHPS) in 2007-2008. The BHPS is a large longitudinal survey representative of the UK population, with each wave covering about 15,000 geo-referenced households. We used the survey version with home location small geography identifiers, Lower Layer Super Output Area (LSOA), which we used to link home residence's location with respect to the telecom infrastructure. We combine the individual and household respondents' questionnaires and focus on wave Q, which is the only one that includes questions related to the use of computers at home and the availability and use of the internet. First, we explore household-level questions regarding the availability of computers, internet connections and broadband subscriptions. Second, we examine information regarding computer use at home for individuals 15 to 55 years of age. Individuals have to answer whether they use the computer at home for different tasks such as connecting to the internet, playing games or doing educational work. In addition, we include household background characteristics such as socio-economic status group, gender, household composition, housing tenure and age in the empirical estimations.

2.3 Summary Statistics

Table 1 provides the descriptive statistics of key variables in the whole and our baseline estimation sample, with mean values and standard deviations. In the full sample, our data cover slightly more than 1.1 million students living in over 400,000 postcodes and attending more than 2,860 schools in England over the period 2005–2008.[8] As we discuss below, our estimation sample is constructed by focusing on households within 300 meters of an LE catchment area boundary segment.

Panel A provides the descriptive statistics of our outcome variables for the subject-specific tests and the mean of the three. Panel B of Table 1 provides summary statistics on these two distances for both the full and estimation samples. We explain below how these measures are constructed. Panel C reports the pupil-level characteristics and shows that the vast majority of the pupils are white, approximately 14% are entitled to free school meals, and students live on average 2.5 kilometers from their schools. Panel D displays the postcode-level characteristics, which show similar values in the proportion of white and FSME students at the local level, a majority of community-type schools and an average house price of approximately £190,000. The table shows that the composition of pupils and area characteristics for the whole sample are highly similar to those of the estimation sample.

[8]The raw data include approximately 500,000 pupil observations per year. To prepare the sample for our empirical strategy, we exclude observations that (i) have implausible or inconsistent values, (ii) are assigned to segments that only have observations on one side, thus making it impossible to perform within-segment-year comparisons, (iii) cannot be linked to school or local area characteristics, or (iv) are located in postcodes that had broadband services enabled for less than six months.

3 Empirical Strategy

3.1 Sorting Issues and Identifying Variation

To estimate the effects of home broadband on education outcomes, a major identification challenge has to be addressed: household-level observed broadband speed – i.e., package choices – is likely related to learning outcomes through confounding factors that are difficult to directly control for. We refer to this as *active* sorting; e.g., better-off households invest in better connections to boost outcomes. As a result of this type of sorting, using data on observed broadband speeds is problematic. The approach used by the existing literature is therefore to focus on variation in *available* broadband speeds, which depends on location choices but not on broadband subscription choices.

In this context, we exploit a feature of the DSL-broadband technology: a salient feature of the technology is that once a home is connected to a broadband-enabled LE station, the *available* connection speed depends on the length of the copper wire connection between the residence and the LE station.[9] We use this characteristic by exploiting "jumps" in distance to the LE across catchment area boundaries.

In the UK, the location of LE stations was determined during the deployment of the English landline telephony network, which mainly occurred before and during World War II. Importantly, distance to the LE station did not affect the quality of traditional telephone services, so the setup of the network was designed to maximize the number of connections from the minimum number of stations (with the goal of achieving cost-effectiveness). However, while we believe that it is unlikely that households in the past or present actively sort on the basis of distance to the connected exchange station, there are several reasons to believe that LE station location is far from random and is potentially correlated with other local neighborhood characteristics that do matter for household sorting. One can argue that most households are not likely to be aware that the speed of their broadband connection is related to their location choices, and even if they are, they might not precisely know where LE stations are located. House-

[9]Distance to the LE is not the only driver of the variation across DSL subscriptions, as other factors can also affect observed speed, such as the quality of the hardware and software used, the number of simultaneous users in the household, the day of the week and time of day, the size and upkeep of the LE station and other technical factors such as varying quality of in-house wiring, unconnected microfilters, or varying performance by the ISP. See OfCom (2009a) and OfCom (2009b) for more details. See Ahlfeldt et al. (2017) for estimates of the determinants of household observed broadband speed.

holds might not be located at different distances from the LE because they know and care about the speed-distance relationship (abstracting from *active* sorting). However, they might sort with respect to other geographical features also correlated with station location. For example, LE stations appear to have been placed at central locations (local town centers) that were also close to major road junctions for hosting the exchange switchboard infrastructure. We refer to this as *passive* sorting. As local geography correlates with both the location of LE stations and the location of households across space, comparing households located within an LE would lead to biased estimates.

To overcome identification issues related to *passive* sorting, instead of comparing different locations within LEs (e.g., households connected to the same LE but at different distances from the station), we use a strategy that compares variation in DSL-cable length across neighboring locations. In spirit, this strategy is similar to Black (1999) who compares house prices across school catchment area boundaries. As we show later, house prices are fully balanced in our setting because these boundaries were unknown to the general population over the study period. To ensure that other local characteristics do not affect our estimates, we compare households located very close to each other and thus with similar geographical features but with connections to different LEs and hence with different broadband *availabilities*. These boundaries give rise to substantial cross-sectional variation in the quality of the available DSL-broadband speed due to discontinuous jumps in the length of the copper wire that connects residences on either side of the boundary to their assigned LE stations. This holds across any given boundary segment. In a nutshell, we compare the distance to the connected exchange of postcodes on both sides. We call the resulting discontinuous changes in average distances to the LE on each side of small boundary segments "jumps".

Note that in this paper, we do not compare outcomes between DSL-connected and unconnected places but exploit differences in broadband speed quality across locations.[10] The period for which we could compare locations with internet dial-up connections to those already connected to a broadband-enabled LE is that encompassing the rollout of the DSL infrastructure, e.g., mostly 2000–2004. The technological upgrade between these two technologies would have allowed us to exploit a 10x increase in ex-

[10] Additional technological and regulatory changes also took place during our period of analysis; however, by comparing within-year cross-boundary segments, we focus solely on cross-sectional variation to obtain our estimates.

pected internet speeds; however, the take-up rate of broadband services before 2005 was negligible, reducing the probabilities of finding an identifiable treatment. Instead, we use data for the years 2005–2008, when broadband infrastructure was almost universally enabled and when take-up rates were already substantial and growing, and focus on differences on available internet speed across neighbouring locations.

We argue and provide evidence below that using the variation in available broadband speed across local boundaries addresses the discussed endogeneity concerns. This variation has to be exploited at very small scales to avoid spatial confounders correlated with location and outcomes. To leverage the richness of the data, it is essential to use very disaggregated information, in terms of both geographical scale and sample size; thus, the available geolocated administrative data are key for the application of a robust empirical strategy. Next, we explain how we take this setting to our data and our estimation approach.

3.2 Construction of the Discontinuity and Treatment Variables

The core of our empirical strategy is the construction of the boundaries of LE catchment areas. The first thing to note is that these boundaries do not coincide with any other administrative boundaries, in particular school district boundaries. For this, we use information on the precise geolocation of the universe of English LE stations (approximately 3,900). In particular, we use the assignment of each of the English postcodes to the LE that provides telephone and internet services to its premises. There are approximately 1.45 million full postcodes in England. Each postcode contains approximately 15 households on average, and the postcode areas are often as small as a single building, especially in denser areas. Panel A of Figure 1 illustrates the location of an exchange station A and the associated catchment area. Figure A.1 depicts the LE catchment areas for all of England.

We match individual postcodes to their closest exchange station boundary segment. We first extract from the boundary for each catchment area the segments that are unique borders (lines) between two different LE stations. This can be done by cutting the polygon boundaries at intersections. Second, the boundaries are divided into smaller segments (henceforth called boundary segments), which are on average 3.2 kilometers long (S.D. of 1 kilometer). In the third step, we assign all postcodes in England to the particular boundary segment that it is closest to, conditional on which LE they are connected to.[11] This determines which boundary segment each postcodes belong to. Panel B of Figure 1 shows such particular boundary segment in red and close postcodes matched to this it as little dots. Note that to compare households with similar geographical surroundings, we only use postcodes within 300 meters of the boundary segment in our preferred estimation sample. Robustness checks on the sensitivity of our findings to the baseline sample selection are discussed extensively in Section 4.4.

For the boundary-postcodes, e.g those assigned to a particular side of a boundary-segment, we can now compute two distance measures that are required for the construction of the treatment and SRD variables. First, we calculate the Euclidean distance from the postcode centroid to the connected LE station; this approximates the connecting copper cable length, which is a measure of internet speed quality.[12] Second, we calculate the distance between postcode centroids and its boundary segment. This will allow us to compare postcodes close to each other.

It is important to reiterate that there are two different types of distances. The first is the distance from the postcodes to the connected LE station, which is an important determinant of the available broadband speed. This distance increases as we approach the boundary segment and changes discontinuously when we cross an LE boundary segment. This is the distance measure that gives rise to the variation in broadband speed across boundary segments. The second measure is the distance to the LE boundary segment. This distance is used to identify close neighbors, i.e., to select which locations we use as comparisons within a short segment, and it is our SRD running variable.

In each side of a boundary-segment, postcodes differ individually in their copper cable lengths, but on average the group in each side is relatively closer or further away from its connected LE station. To make this geographical setting operational, we calculate the *average* distance to the connected LE of the postcodes on each side of a boundary-segment. The side that has shorter average distances is labelled as the

[11] We know the precise geolocation of the postcode centroids using the British National Grid Eastings and Northings to the 1 meter precision from the National Statistics Postcode Directory.

[12] A shortcoming of our approach (common to other, similar papers) is that we can only calculate crow-fly distances between the centroids of the postcodes of the location of pupils' homes and the LE station to which they are connected; in urban areas, there can be a substantial gap between this and the actual length of the connection cable (OfCom, 2009b). Nevertheless, given the very small scale of our geographical units of observation, we can approximate this in a more precise way than other studies that use data for larger geographical units.

Fast side, and the other side is defined as the *Slow* side. The difference in the average distances between the two sides captures the "jump" in average cable length when we cross the boundary segment; from it we identify the impact of changes in speed broadband.[13] This is illustrated in Panel B of Figure 1. Here, the average exchange-distance for the boundary postcodes in catchment area A is larger compared to catchment area D. As a result, the postcodes on the boundary segment connected to D are labelled the *Fast* side, whereas the other side is the *Slow* side.

In the previous example, all the postcodes in the *Fast* side are on average but also individually closer to their connected LE station than all the postcodes in the *Slow* side (sharp RDD). However, it could be the case that two addresses located on different sides of the segment are not necessarily different in the way that we would expect. Due to the irregular geographic shape of the boundaries, some households with shorter cables (longer cables) might live on the on-average slower side (faster side). If the jump in average distances between sides is attenuated because some postcodes are "assigned to the wrong side", this could result in attenuation bias. We illustrate such this situation in Panel C of Figure 1. [14] To take into account these cases, we implement a fuzzy spacial RDD using an IV strategy, which we explain in detail in the next subsection.[15]

3.3 Specification and IV Strategy

Our goal is to estimate the causal effect of broadband internet speed on test scores for 14-year-old students. The basic framework of our analysis capturing the relationship of interest is the following:

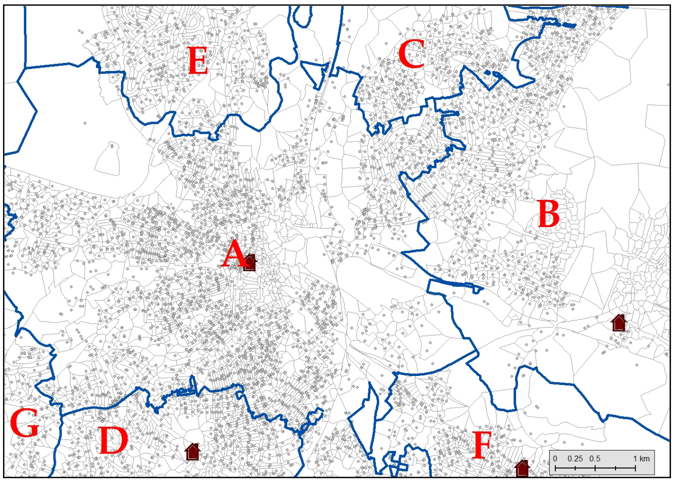
$$TestScore_{ipnlst} = \beta DistLE_{pl} + g(D_{pn}) + X_i'\Lambda + Z_{is}'\Theta + A_{pt}'\Phi + L_p'\Psi + \delta_{nt} + \epsilon_{ipnlst} \quad (1)$$

[13] Sometimes the jump between the two sides is small, if the variation in average distances is relatively low. For this reason, we exclude segments in which the jump is below 100 meters, and for our main results, we focus on segments with jumps of at least 300 meters. We extensively discuss the robustness of this choice in Section 4.4.

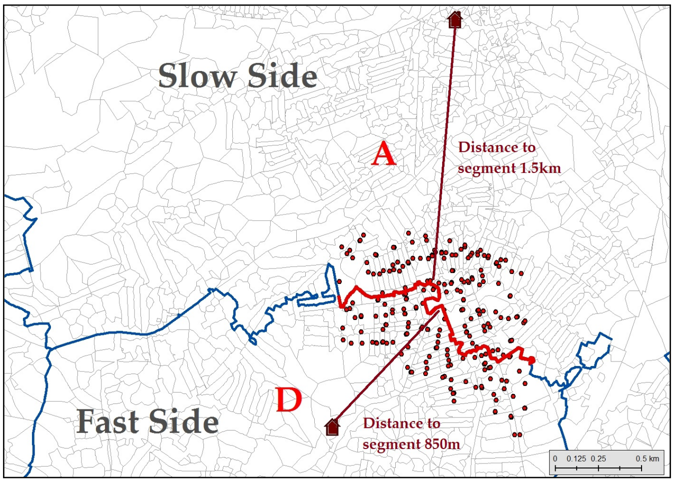
[14] For segment AB, the average distance to the connected LE is quite similar on both sides, approximately 1.6–1.7 kilometers. The shape of the segment is irregular and slightly diagonal, tilted to the west. Two sets of postcodes are selected to explain this situation. The triangular-shaped ones are both around 3 kilometers away from the LE station, but the shorter of the two (3,043 meters) is located on the *Slow* side, and the longer segment in the pair (3,049 meters) on the *Fast* side. The pentagon-shaped postcodes display the inverse situation: the one with longer distance, 2,467 meters, is located on the *Fast* side, while that with the shorter distance (2,402 meters) is on the side with longer distances on average.

[15] For completeness, we also present (attenuated) sharp SRD estimates, see Table A.2.

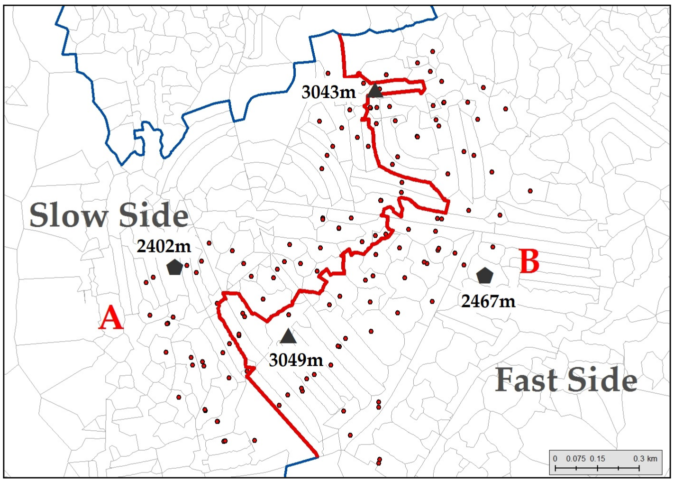
Figure 1: Graphical Illustration of Empirical Strategy.



(a) Overview of LE catchment areas and boundary segments



(b) Postcode sample boundary A–D segment



(c) Postcode sample boundary A–B segment – Fuzzy

Notes: The light gray dots represent the precise location of the postcode centroids for a selection of LE catchment areas at the 1 meter-resolution precision. The blue lines represent the LE station boundaries. The building symbols display the exact locations of LE stations. Subfigure (a) shows an overview of LE station areas with underlying postcode area polygons and centroids. One letter is allocated to each LE. Note the irregular shapes and the fact that LE stations are not always located in the center of the LE area. Subfigure (b) and (c) zooms in on the boundary segment of a particular LE catchment area, AD (b) and AB (c). The dots mark the postcodes located within 300 meters of the red boundary segments.

where $TestScore_{ipnlst}$ is the percentile rank in the KS3 test of pupil i living in postcode p in boundary-segment neighborhood n associated with LE station l and attending school s at time t; $DistLE_{pl}$ is the distance to connected LE station l from home postcode p; X_i' is a vector of student background characteristics, such as preinternet student performance on the KS1 test, gender and free school-meal eligibility status; Z_{is}' is a vector of characteristics of the school attended by pupil i, such as school type and distance between home and school; A_{pt}' is a vector of postcode-year-specific characteristics, such as local average housing prices, share of students eligible for free school meals, and white (population) pupils; L_p' is a vector of time-invariant postcode attributes such as density (e.g., number of delivery points) and distance to different amenities (e.g., nearest rail station or road); δ_{nt} is boundary segment-by-year fixed effects, which guarantee that we are comparing students within the same segment-year of the LE boundary; and ε_{ipnlst} is the error term. To ease interpretation of the estimates, we measure $DistLE_{pl}$ as "negative distance", e.g., proximity to the LE station. Thus, *beta* captures the changes in test scores when we come closer to the LE by one meter and available broadband speed improves. We cluster the standard errors at the segment-by-year level.[16]

To apply the empirical model to the data, we need to specify two additional pieces of information. First, the definition of $g(D_{pn})$ captures the relationship of the postcode distance to the boundary segment. This deterministic function has the spirit of the running variable in a nonspatial regression discontinuity design. In our setting, we are interested in controlling for distance to the boundary segment because of *passive* sorting, as explained in Section 3.1 above. As we have a large sample of pupils, even when we restrict observations to postcodes within 300 meters of the border, we can carefully control for distance to the boundary segment, where each side is connected to a different LE station. This means that we can effectively compare the test scores of students with differences in home broadband access living very close to each other.

Our preferred control function for the relation with the running variable is the fol-

[16]Table A.5 shows that different clustering choices for our preferred specification do not change our conclusions.

lowing:

$$
\begin{aligned}
g(D_{pn}) &= \sum_{b=1}^{B_s} \left(\gamma_b^{Slow}\, Slow_{pn} * I\left(DistSegment_{pn} = b\right) * DistSegment \right) \\
&+ \sum_{b=1}^{B_f} \left(\gamma_b^{Fast}\, Fast_{pn} * I\left(DistSegment_{pn} = b\right) * DistSegment_{pn} \right) \qquad (2)
\end{aligned}
$$

where we control for the distance to the LE invisible boundary segment by using distance bin dummies of 100 meters (b) interacted with the distance to the boundary segment on both sides of the cutoff. In terms of the notation, $Slow_{pn} * I\left(DistSegment_{pn} = b\right)$ is an indicator variable for postcode p on the *Slow* side of boundary segment n that equals one if the distance of postcode p to segment n is within bin b. $Fast_{pn} * I\left(DistSegment_{p} = b\right)$ is defined analogously for all postcodes on the *Fast* side of the boundary segment n. This semiparametric approach allows for more flexibility in controlling for distance to the segment. Instead of imposing a certain functional form on the polynomial, we estimate the coefficients γ_b^{Slow} and γ_b^{Fast} for each small distance bin for the segments (i.e., 0–100, 100–200, and 200–300), thereby capturing the shape of the polynomial in a flexible way on each side of the boundary segment. This approach is flexible and avoids either oversimplifying the underlying relationship (as would be the case if, for example, we used a linear polynomial) or overfitting by using high-order polynomials (Gelman & Imbens, 2019). The results are robust to alternative definitions of distance bins, to the use of distance as a continuous variable, and for a more flexible approach of higher-order polynomials.

One final step is required for estimation. In order to overcome the attenuation bias induced by the assignment of postcodes into *Fast* (closer) and *Slow* (further away) sides, we use an instrumental variables strategy and a fuzzy SRD specification. Using the information on which side of each boundary segment has a lower average distance to the LE, we construct segment-side-specific *Fast* dummies that we use as instruments for the actual postcode-level distance to the LE. The first-stage equation is:

$$
DistLE_{pl} = \pi Fast_{pn} + g(D_{pn}) + X_i'\Lambda + Z_{is}'\Theta + A_{pt}'\Phi + L_p'\Psi + \delta_{nt} + \varepsilon_{ipnlst} \qquad (3)
$$

where $Fast_{pn}$ is a dummy variable equal to one if student i living in postcode p is located on the *Fast* side of LE station boundary segment n and zero otherwise. As above, $DistLE_{pl}$ is measured as proximity to the connected LE station. The coefficient π captures how much closer postcodes are to their connected LE stations on the *Fast* side relative to the average distance on the *Slow* side.

Since the telephone LE station boundaries are historically given and under the assumption that households do not sort on each side of the boundary within the boundary segment, we can focus on households whose residences are located in the vicinity of the LE boundary segment, considering those on the *Fast* side very similar to those on the *Slow* side. The broadband speed assigned to these households can be considered "locally" randomly assigned within segment-by-year. Assuming that in the absence of the treatment, the outcome variable is a smooth function of distance to the LE boundary, the causal effect of broadband internet speed is identified by comparing outcomes for pupils who live close on the *Fast* side of an LE station boundary (treatment group) with those who are near but live on the *Slow* side (control group). This effect is captured by the IV estimate of β in specification 1. This strategy estimates a local average treatment effect (LATE) of broadband internet speed on student performance by comparing "lucky" households that are supplied with faster broadband access to otherwise similar counterparts that were "unlucky" in terms of being supplied with slower broadband access. A specific feature of this strategy is that it generates variation over multiple thousand telephone LE station boundary segments, which vary in distance to the LE station and in the jump in distance across the boundaries.

This identification strategy is similar in spirit to those of Falck et al. (2014), Ahlfeldt et al. (2017) and Amaral-Garcia et al. (2019). Falck et al. (2014) estimate the effects of information disseminated over the internet on voting behavior in Germany between 2005-2008, where entire (small) locations happened to be located too far away from an exchange station to access broadband. In particular, towns farther than five kilometers from an exchange could not obtain any broadband internet without costly further technological upgrades. This characteristic allows the authors to exploit differences in outcomes between places that were connected to the DSL network and those that were not. However, this approach is not directly applicable to the British context because of a much denser network of LE stations, related to the smaller size of the country, with all places connected to the network, and the relatively quick rollout of the broadband infrastructure. Similarly to us, Ahlfeldt et al. (2017) and Amaral-Garcia et al. (2019) exploit discontinuities in the proximity to connected stations across LE catchment areas boundaries. They exploit changes in local speeds induced by technological upgrades over time. In contrast, we exploit spatial cross-section variation in distance to the connected LE station within year for a stable technology (ADSL).

3.4 Estimating the Impact of Broadband Speed on Test Scores

In a simple framework, faster broadband internet could affect teenagers' school performance in different ways. High internet speed allows students to access more online content per unit of time. If test scores are determined in a learning production function, we can think of speed impacting learning productivity. For each hour of study, students can access more information, shifting the learning production function upwards if, for example, they can access more learning resources such as Wikipedia or online interactive materials. However, broadband could affect learning investment by reducing study hours if students divert time to non-learning online activities, such as gaming or using social media. This second channel could be more or less relevant depending on whether online distractions replace offline distractions. The first channel would have a positive effect on human capital formation, while the second would have a negative effect. Coefficient β captures the net impact of these channels, and our aim is to identify an unbiased estimate of this coefficient.[17]

Henceforth, the causal estimation of coefficient β in equation 1 informs us about the impact of *available* broadband speed on test scores. It captures how much test scores change as we approach an LE station via the relationship between copper cable length and potential speeds. We estimate a LATE impact by using a proxy of broadband speed that changes over space. Our treatment measures local availability of faster broadband rather than student-specific faster connections and provides a measure of the intent-to-treat (ITT) impact.[18] While the estimation of this parameter is desirable from an identification point of view, it is also a relevant policy variable. By investing in network improvements and expansions, public policy can influence the availability of fast broadband in different locations, while afterwards, particular households might

[17] This model can become more complex if, for example, we take into account the changing nature of web content over time or the interaction of school and home ICT use, but discussing this is beyond the scope of this paper.

[18] Related papers also provide ITT estimates (Ahlfeldt et al., 2017; Dettling et al., 2018; Amaral-Garcia et al., 2019; Geraci et al., 2022).

sort into different broadband packages and use internet in different ways for reasons that might correlate with the outcomes of interest.

We are also interested in the impact that broadband speed changes have on student performance. For the reasons explained above and due to the lack of data appropriate to the time period and geographical scale of our analysis, we cannot directly estimate this impact in the data. Our approach is to employ additional data to compute the distance-speed relation present during our study period, combining linked postcode-to-exchange station telecom network data with data on the local average internet speed experienced by households. This approach is not dissimilar to other papers such as Ahlfeldt et al. (2017) and Amaral-Garcia et al. (2019).[19] We compute the Wald estimate dividing our LATE coefficient (effect of distance to LE station on test scores) by this parameter (distance to LE station on average speed) to obtain a measure of the impact of changes in speed on test scores and calculate the standard error using the Delta method.

Similarly to Dettling et al. (2018) and Zuo (2021), we re-scale our ITT estimates to provide a measure of treatment-on-treated (TOT) effect. We adjust our Wald estimate by the relevant population of interest. We do this using additional survey data to obtain information on the appropriate broadband take-up rates during the period of analysis (those of families with teenagers during 2005–2008 in England) and on the proportion of pupils using the internet for school work. This allows us to discuss the economic relevance of our results and to contextualize the size of our estimates. We extensively discuss our quantification estimates in Section 5.

3.5 Estimation of the mechanisms

This subsection uses the telecom network data matched to the BHPS to analyze the channels behind the effect of broadband speed on educational achievement. We adapt

[19] In Ahlfeldt et al. (2017), the authors estimate a hedonic price model of the effect of broadband speed on individual house prices using a prediction of local yearly broadband speed constructed from a separate dataset. To overcome the lack of availability of postcode-year speed data, the authors predict postcode-level speeds served by different LE with different technologies for years 1995-2010 using a large individual broadband speed test dataset in years 2009-2010, which includes rich information on internet packages, type of technology and distance to LE. Amaral-Garcia et al. (2019) rely on their estimates to construct speed predictions by ADSL technology and geographical aggregation (LSOA). They use the predicted speeds to assess the impact of broadband internet speed on C-sections. These strategies are close to a two-step procedure, similar to the one we implement by means of a Wald estimate (with adjusted S.E.).

the empirical strategy to the smaller survey dataset, and compare individuals on the fast and slow sides of LE catchment area boundaries (reduced-form). To estimate the effect of broadband internet speed on broadband take-up and computer presence at home, we ran the following specification:

$$Y_{ihmlt} = \beta Fast_{hm} + g(D_{hm}) + X_i'\Omega + Z_h'\Lambda + \delta_{mt} + \epsilon_{ihmlt} \tag{4}$$

where Y_{ihmlt} is the broadband take-up or computer use at the household of individual i living in LSOA h in boundary neighbourhood m associated with LE station l at time t; $Fast_{hm}$ is a dummy variable equal to one if student i living in LSOA h is located on the *Fast* side of LE station boundary m and zero otherwise; $X_i'\Omega$ includes individual and household characteristics including age, gender and socio-economic status; $Z_h'\Lambda$ is a vector of LSOA socio-economics characteristics, δ_{mt} are boundary-by-year fixed effects, which guarantee that we are comparing students within the same LE boundary-year; and ε_{ihmlt} is the error term. We cluster the standard errors at the boundary-by-year level. The relationship of the LSOA distance to the LE boundary is captured by the control function $g(D_{hm})$, which includes distance bin dummies of 100 meters interacted with distance to the boundary, differently estimated in each side.

To go deeper into the mechanisms driving our findings, we then focus on a set of computer uses by individuals of different ages. We estimate the following specification:

$$\begin{aligned} Y_{ihmlt} = {} & \beta(Fast_{hm} * Aged15-18_{ihmlt}) + \gamma Aged15-18_{ihmlt} + \rho Fast_{hm} + g(D_{hm}) \\ & + X_i'\Omega + Z_h'\Lambda + \delta_{mt} + \epsilon_{ihmlt} \end{aligned} \tag{5}$$

where $Aged15-18_{ihm}$ identifies individuals i aged between 15 and 18 years old in LSOA h in boundary neighbourhood m associated with LE station l at time t.

This specification is akin to a differences in discontinuities design. We compare two age groups (teenagers and adults) across boundaries (fast and slow sides). This setup allows us to explore whether teenagers use computers differently to adults in the same household, and whether teenagers on the fast side behave differently to those on the slow side. Hence, we can focus on the population of interest, i.e. teenagers. The interaction term, β, captures the differential computer use of teenagers on the fast side with respect to those on the slow side.

For consistency with section 3.2, we focus on boundaries with jumps of at least 300 meters. Our baseline regression focuses on individuals living within one kilometer of the LE catchment area boundary to improve statistical power. Additionally, we restrict our sample to individuals younger than 55 years old to focus on the population whose children are in school years.

4 Results

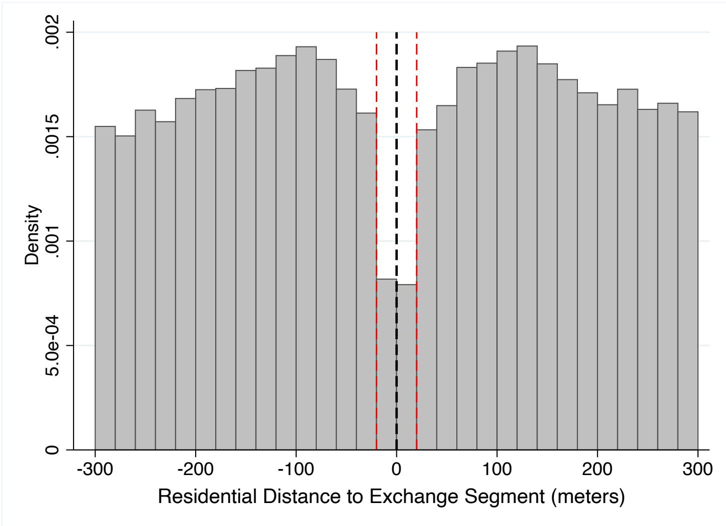
4.1 Internal Validity of the Empirical Strategy

The internal validity of the fuzzy SRD requires that there is no endogenous sorting on either side of the LE boundary segment. In the case of this study, there are two key features that make manipulation at the LE station boundary segment highly unlikely. First, LE catchment areas are either invisible or unlikely to be known by households. Second, LE boundaries do not coincide with administrative boundaries of any kind, such as school district boundaries. The fact that both our identification strategy is not bound to other administrative boundaries where variables affecting our outcome of interest might change discontinuously at the cutoffs, and that we have granular data to use a data-driven approach to generate boundary segments, provide a unique advantage to estimate the causal effects of home broadband speed.[20]

Figure 2 presents a histogram representation of pupil density at the LE station boundary segment for our preferred estimation sample. It allows for visual inspection of whether bunching takes place. It reveals no evidence of systematic manipulation of residential distance to the LE boundary segment around the thresholds. However, we observe that density drops within 20 meters of the LE boundary segment (indicated by the red vertical dashed lines). This drop is a byproduct of the data construction: postcode distances to the boundary segment are measured from the centroids, and there are very few oddly shaped postcodes where the centroid falls within 20 meters of the boundary segment. The fact that the drop is of the same magnitude on both sides of the discontinuity is a positive indicator that it is unlikely that this feature has a differential effect on one particular side of the LE boundary segment.

[20]In the data construction, we removed boundary segments that intersect with natural boundaries (rivers). While it is now possible to determine (rough) boundary locations online, this was not possible

Figure 2: Distribution of Pupil Distance from Residential Distance to Exchange Segment.



Notes: The figure shows the density of pupils in our preferred estimation sample at the LE boundary segment. The black dashed vertical line presents the LE station boundary segment. The red dashed vertical lines show the boundaries of the donut strategy. Each bar contains bins of 20 meters' distance.

We also perform more in-depth analyses and formal tests of bunching, as reported in Section A.2. We formally test for bunching following McCrary (2008) and Cattaneo et al. (2018). The McCrary (2008) test fails to reject the null of no significant jump at the LE boundaries.[21] Cattaneo et al. (2018) propose a test that is robust to bandwidth selection issues. We fail to reject the null hypothesis of no discontinuous jump at the LE boundary segment.

Since the density drop within 20 meters is a byproduct of the geographical resolution of the data, we use the so-called donut strategy, which excludes observations within 20 meters of the discontinuity in our main specifications, following Angrist et al. (2019) and Leuven & Løkken (2020), among others. Discarding these observations improves the precision of our estimates since we eliminate potentially rare postcodes and spillovers across LE boundaries.[22]

Another direct check of a violation of instrument exogeneity is to test whether pupil, school and area/postcode baseline characteristics are "locally" balanced on either side of the LE boundary segment. If these variables are unbalanced on either side

during the period studied.

[21] We also test for various alternative residential distances to the LE boundaries. The results are robust to different specifications. See Table A.1.

[22] At the boundary segment, very local spillovers might exist where households on the *Slow* side could connect to the WiFi routers of neighbors on the *Fast* side. The fraction of pupils who are discarded represents less than 1% of our estimation sample.

of the boundary segment, it would indicate selection problems around the discontinuities. Table 2 tests the balance of a battery of background characteristics. The first column of Table 2 regresses the main outcome on the predetermined background characteristics. The regressions further control for distance bins and include segment-by-year fixed effects. The results of this column show that the background characteristics are economically and statistically important in explaining the variation in the outcome variable. We strongly reject the null hypothesis that these variables are jointly equal to zero. The third column of Table 2 displays the results of local linear regressions for each of the predetermined pupil, school, density, area socioeconomic, housing price and amenity characteristics. We show that most of the regressors are not significantly different from zero, and the coefficients are very small. In particular, important determinants of our outcome variable, such as individual preinternet scores, free school meal eligibility and housing prices, are statistically indistinguishable from zero.[23] We find that postcodes on the *Fast* side of the LE boundary segment are more likely to be closer to a school and a rail station and less likely to be white. However, the magnitude of the differences in these characteristics with those on the *Slow* side of the LE boundary segment is very small (0.6%, 0.3% and 1.8% with respect to the baseline mean, respectively). Given the number of variables tested, it is not surprising to find some small imbalances in individual variables. In line with this, we fail to reject the joint test in which all the coefficients are equal to zero at conventional levels of statistical significance. This means that we do not find evidence that third factors, some of which are extremely important in explaining later test scores, change discontinuously across our boundary segment.

[23] Ahlfeldt et al. (2017) find statistically significant impacts of faster broadband internet on house prices. Note that such result does not raise concerns about the potential bias of our baseline estimates. As we explain in Section 3, we implement a different empirical strategy. While they exploit changes in local speeds induced by technological upgrades over time, we utilize spatial cross-section variation in distance to the connected LE station within a year. Thus, the fact that we find house prices are balanced on either side of the discontinuity (see Panel D of Table 2) does not contradict the main findings of Ahlfeldt et al. (2017).

Table 2: Balance of Baseline Student, School and Area/Postcode Characteristics.

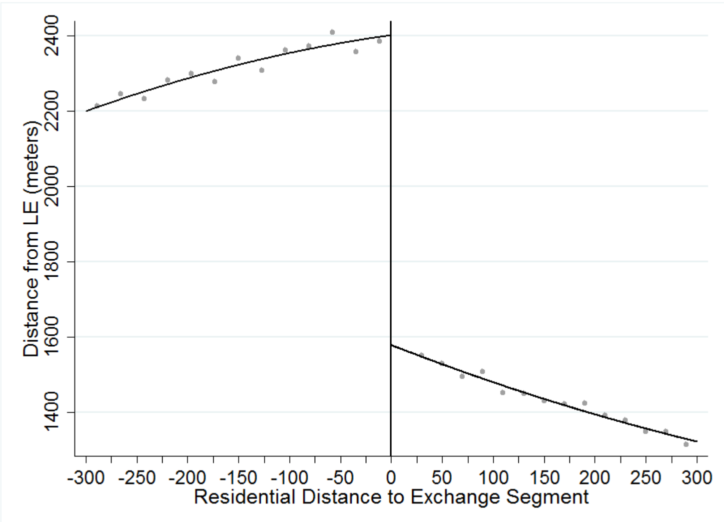
	Outcome	Instrument	
	Average Percentile Rank Score	*Slow* Side Baseline Mean	*Fast* Side Nonparametric Estimate
	(1)	(2)	(3)
A. Student Characteristics			
White	-0.592**	0.78	-0.014*
	(0.244)		(0.007)
Male	-2.430***	0.5	0.005
	(0.133)		(0.010)
Free School Meal	-13.749***	0.16	0.003
	(0.206)		(0.008)
Pre-KS3 Score	0.819***	43.89	-0.153
	(0.002)		(0.508)
B. School Characteristics			
Log Distance Home to School	3.699***	7.39	0.012
	(0.112)		(0.017)
Log School Size	1.305***	6.88	0.007
	(0.140)		(0.013)
Community School	-9.951***	0.62	0.003
	(0.214)		(0.009)
C. Density & Area Socioeconomics			
Log Number of Delivery Points	-0.944***	3.34	0.013
	(0.132)		(0.015)
Log Number of Pupils per Premise	-1.754***	-1.76	0.016
	(0.121)		(0.016)
Share of White Pupils	1.416***	0.77	-0.008
	(0.389)		(0.005)
Share of Free School Meal Pupils	-22.724***	0.17	-0.001
	(0.354)		(0.006)
D. Housing Prices			
Log Average Housing Price	13.310***	12.04	0.014
	(0.270)		(0.009)
E. Amenities			
Log Distance to Closest Road	0.045	5.44	-0.013
	(0.103)		(0.022)
Log Distance to Closest School	1.416***	5.83	0.033**
	(0.134)		(0.016)
Log Distance to Closest Supermarket	1.156***	6.56	0.013
	(0.191)		(0.012)
Log Distance to Closest Rail Station	0.118	7.22	0.020**
	(0.259)		(0.009)
Log Distance to Closest Water Body	-0.144	6.5	-0.002
	(0.136)		(0.017)
Joint F-test	10,381		1.23
Joint P-value	0		0.23
No Segm-years	7,096		7,096
No Observations	183,892		183,892

Notes: Column (1) regresses the main outcome on the predetermined background characteristics. Column (2) shows the baseline mean on the *Slow* side for the predetermined background characteristics. Column (3) performs local linear regressions for each of the background characteristics. Each regression controls for distance bins and segment-by-year fixed effects. F-tests (and corresponding p-values) are for the joint significance of the variables reported in each column. Standard errors clustered at the segment-by-year level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.2 Discontinuities in Distance to the LE Station

We now investigate the discontinuities across the boundary segment in the distances to the connected LE stations. Figure 3 plots the distance to the connected LE station as a function of household distance to the LE boundary segment. The left side shows the average "stacked" distances to connected exchange stations of all boundary segments on the *Slow* side. As we move closer to the boundary segment, the distance to the exchange increases. Then, there is a clear discontinuity. On the *Fast* side, the average distance to the (different) connected exchange in turn decreases as we move farther away from the boundary segment and closer to the LE station.

Figure 3: Distance from the Connected LE Station Jumps across the Boundary Segment.



Notes: The dots represent the average distance to the LE station per 25-meter interval of residential distance to the LE boundary-segment boundaries. The solid lines are fitted values from a third-order polynomial approximation, which is estimated separately on both sides of the cutoffs. "Residential Distance to Exchange Segment (Meters)" refers to the residential distance to the LE boundary-segment boundaries. The black vertical line is the stacked LE station boundaries.

In Panel A of Table 3, we estimate the magnitude of this jump, which is the first stage in our fuzzy SRD setup. The results confirm that being on the *Fast* side of the threshold is a strong and statistically highly significant predictor of the distance to the connected LE station. Our preferred specification coefficient, shown in the fifth column of Table 3, estimates that households just on the *Fast* side are 816.6 meters closer to their LE stations (relative to those on the *Slow* side). This coefficient and the corresponding estimates from the various other specifications, which are discussed in more detail below, are statistically significant at the 1% level.

4.3 Impact on Student Test Scores

Panel B of Table 3 shows the fuzzy SRD nonparametric estimates on the average student percentile rank score. Each column presents estimates from increasingly saturated specifications. Column 1 includes student test scores from a prebroadband period in England (KS1 test scores from age 6/7) as well as the distance bins and segment-by-year fixed effects. Adding this predetermined control variable effectively changes the empirical strategy to a fuzzy SRD *value-added* design. Using this specification, moving one meter closer to the LE station increases the national KS3 exam performance at age 14 by 0.00123 percentile ranks. The results are statistically significant at the 5% level. In column 2, we add further individual controls (e.g., ethnicity, free school meal eligibility, and gender), which on their own have explanatory power, as documented in Table 2. To increase precision, we include interactions between the different combinations of student individual controls. The main effect increases slightly further to 0.00131 per meter, which is statistically significant at the 1% level. Columns 3 and 4 introduce the additional time-variant and time-invariant area and amenity controls from Table 2. This hardly affects the estimates. Finally, the last column includes additional school-level controls. In column 5, we control for school type, size and distance from home. This does little to the estimates.

Because of the stability of the estimates across columns, we choose the estimate from column 5 as our preferred specification, for which we discuss more results below.[24] The baseline estimate is statistically significant at the 5% level (p-value=0.01). As explained above, due to the irregular geographic shape of several boundaries, it is critical to use a fuzzy SRD design since sharp RD estimates would suffer from attenuation bias. Table A.2 illustrates the attenuation bias, showing its substantial magnitude. Although many of the estimates are statistically significant at conventional levels, the sharp SRD estimates are approximately five times lower than the preferred fuzzy SRD estimates.

At face value, this estimated effect of about 0.0012 percentile ranks per meter seems very small and in fact is only detectable due to the combination of our empirical approach with formidable student census data that includes national test scores (and

[24] The results are robust to the use of a triangular kernel. The point estimates are larger in magnitude (0.00147), but we cannot reject the null hypothesis that the estimates are different from the baseline estimates. These results are available upon request.

previous test scores) at a very high geographical resolution. In Section 5, we provide direct estimates that allow us to quantify this positive effect in terms of broadband speed and usage to show that the effect size is economically meaningful.

Table 3: Fuzzy SRD Estimates: Impact of Exchange Distance on Student Performance.

	(1)	(2)	(3)	(4)	(5)
A. Avg. jump in distance to connected exchange when crossing boundary to the *Fast* side (in m)					
Nonparametric	-814.3***	-813.8***	-813.3***	-816.4*** *	-816.6***
Estimates	(16.00)	(15.99)	(15.99)	(15.93)	(15.93)
B. Effect of exchange distance on national KS3 exam scores at age 14 (in percentile ranks)					
Nonparametric	0.00123**	0.00131***	0.00120**	0.00120**	0.00122**
Estimates	(0.00049)	(0.00048)	(0.00048)	(0.00048)	(0.00047)
Student Preinternet Score KS1	Yes	Yes	Yes	Yes	Yes
Student Controls	No	Yes	Yes	Yes	Yes
Time-Variant Area & Amenities	No	No	Yes	Yes	Yes
Time-Invariant Area & Amenities	No	No	No	Yes	Yes
School Controls	No	No	No	No	Yes
Observations	183,892	183,892	183,892	183,892	183,892

Notes: The table shows the fuzzy SRD nonparametric estimates for the average distance to the LE station (Panel A) and test scores on the KS3 (Panel B). Columns report different specifications. Each coefficient comes from a separate regression, where the running variable is the residential distance to the LE station boundary. The window size for the residential distance to the LE station boundary is $\pm$ 300 meters. We only include boundaries where the distance jumps by at least 300 meters. Standard errors clustered at the segment-by-year level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.4 Robustness Checks

4.4.1 Sensitivity to Sampling Choices

All results discussed so far are estimated using the sample of postcodes that fall within 300 meters of the boundary segment and where the distance jump to the connected exchange across the boundary segment is at least 300 meters. To clarify that these choices do not affect our conclusions, in Table A.3, we estimate our preferred specification with different samples, using only observations falling within 100 meters of the boundary segment up to those falling within 1 kilometer on either side. The effect sizes are slightly larger in samples closer to the boundary segment (0.00157 in the 100 meter sample) but not significantly different from each other at conventional levels.[25]

In Table A.4, we use our main sample of postcodes within 300 meters of the boundary segment but restrict our attention to boundaries based on the magnitude of the

[25]Note that the results are unchanged when we remove the area and amenities controls in comparing observations very close to the boundary segment, such as those at 100 or 150 meters.

distance jump in crossing to the *Fast* side. In the sample that includes all boundaries except those where the distance jumps by less than 100 meters, the average change in distance from crossing to the *Fast* side is 664 meters (see column 1, Panel A). In contrast, when we exclude all boundaries with jumps of less than 500 meters, the average treatment effect intensifies up to almost one kilometer (column 5). The estimates of the per-meter effect of distance are smaller in the samples that include more boundaries that contribute little to the variation, i.e., in columns 1 and 2. In contrast, the per-meter effect is reasonably constant in all samples that exclude (almost irrelevant) boundaries with jumps of 300 meters or higher. The resulting tradeoff between sample size and the exclusion of boundaries that offer little variation motivates the choice to use the 300 meter threshold as the baseline. Moreover, including many boundaries with small variation "on average" introduces more misclassification of local postcodes into the *Fast* and *Slow* categories.

4.4.2 Sensitivity to School ICT

Table 4: Fuzzy SRD Estimates: Impact of Exchange Distance on Student Performance with School ICT Controls.

Dependent variable: National KS3 exam scores at age 14 (in percentile ranks)					
	(1)	(2)	(3)	(4)	(5)
Nonparametric	0.00122**	0.00124***	0.00124***	0.00103**	0.00092**
Estimates	(0.00047)	(0.00047)	(0.00047)	(0.00047)	(0.00046)
School Proximity to LE Station		0.00047***	0.00047***	0.00058***	
		(0.00010)	(0.00010)	(0.00013)	
Months since ADSL School Upgrade		0.00666	0.00728		
		(0.00652)	(0.03590)		
School ADSL Upgrade Year Dummies	No	No	Yes	No	No
School LE Station Fixed Effect	No	No	No	Yes	No
School Fixed Effect	No	No	No	No	Yes
Observations	183,892	183,712	183,712	183,638	183,598

Notes: The table shows the fuzzy SRD nonparametric estimates for average test scores on the KS3. Columns report different specifications. Each coefficient comes from a separate regression, where the running variable is the residential distance to the LE station boundary. The window size for the residential distance to the LE station boundary is $\pm$ 300 meters. We only include boundaries where the distance jumps by at least 300 meters. Standard errors clustered at the segment-by-year level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4 estimates the effects on test scores, including several important school technological controls. The positive effect of home broadband on student performance may be partially driven by other technological factors that might be correlated with home broadband. For instance, if the effects are mediated by the fact that the school has faster internet connection, not controlling for this variable would bias our baseline estimate upward. Table 4 compares our baseline estimates (column 1) with four different specifications. We find that student test scores are higher if the school is closer to an LE station and as the number of months since the ADSL upgrade increases (column 2). Although this is not a causal estimate based on an SRD, it is still prone to *passive* sorting. However, our main result is robust to the inclusion of this school-level control. In column 3, we add information on the exact year when the LE that the school is located in upgraded to ADSL technology. Again, the main coefficient does not move. In column 4, we include secondary school LE station fixed effects to absorb any variation that is common to the school LE station. In this specification, the point estimate is lower than the previous ones, but it is significant at the 5% level, and moreover, we cannot reject the hypothesis that it is identical to the baseline estimates at conventional levels of statistical significance.

Finally, we add secondary school fixed effects (column 5). Note that this is possible because secondary school attract students from different neighbourhoods. As a result, there exist a sufficient number of students on both sides of each boundary segment attending different secondary schools. This highly saturated specification nevertheless places high demands on computing power. From this specification, we estimate an effect of crossing the boundary of 0.00092 per meter. The point estimate is slightly lower than the previous ones, but we cannot reject the hypothesis that they are identical at conventional levels of statistical significance. The coefficient also remains significantly different from zero at the 5% level. Moreover, the school fixed effects might absorb possible interactions between home- and school-level technology. In conclusion, we cannot reject the null hypothesis that our baseline estimates are different from any of those obtained with the inclusion of these control variables. Hence, the results indicate that our baseline estimates are not biased by other mediating school technological factors but represent estimates of the causal effect of home broadband.

4.5 Heterogeneous Effects

4.5.1 By Subject

We next explore heterogeneity in the impact of broadband on student test scores. First, we analyze separate regressions for each subject (i.e., English, maths and science). This is motivated by the fact that subject differences are often found in the literature on education interventions.[26] Table 5 shows the fuzzy SRD nonparametric estimates on the average student percentile rank score by subject. Column 1 of Table 5 is equivalent to the estimate shown in column 5 of Panel B in Table 3, based on our preferred specification. Splitting the results up by subject, it becomes evident that the effect is the strongest for English, at 0.00141 percentiles per meter. In contrast, the effect is 0.00075 per meter for maths and 0.117 per meter for science. However, the confidence intervals of these coefficients overlap, and we cannot reject the hypothesis of equality at conventional levels.

Table 5: Fuzzy SRD Estimates: Impact of Exchange Distance on Student Performance by Subject.

Dependent variable: National KS3 exam scores at age 14 (in percentile ranks)				
	Average Score (1)	**English** (2)	**Maths** (3)	**Science** (4)
Nonparametric Estimates	0.00122** (0.00047)	0.00141*** (0.00051)	0.00075 (0.00048)	0.00117** (0.00050)
Observations	183,892	183,892	183,892	183,892

Notes: The table shows the fuzzy SRD nonparametric estimates for average test scores on the KS3 by subject. Each coefficient comes from a separate regression, where the running variable is the residential distance to the LE station boundary segment. The window size for the residential distance to the LE station boundary segment is ± 300 meters, and only boundaries where the exchange distance jumps by at least 300 meters are included. Standard errors clustered at the segment-by-year level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

[26] See, for instance, Vigdor et al. (2014) or Falck et al. (2018), who find differential effects between maths and reading and maths and science, respectively. Other empirical evidence, such as that in Malamud et al. (2019) or Cristia et al. (2017), cannot reject the hypothesis that effects on maths and reading are significantly different at conventional levels, but the estimates in both papers have different statistical precision. Machin et al. (2007) find a similar pattern across subjects in their study of the payoff of ICT technology in English primary schools.

4.5.2 By Student Background Characteristics

We analyze heterogeneity along the following student predetermined characteristics: student gender, prebroadband test scores, ethnicity and free school meal eligibility. Some interesting patterns emerge. First, the effect is not driven by only one gender. The effect per meter of distance is both more positive and more significant for girls than for boys (0.00125 vs. 0.00098, respectively). However, both estimates are positive. Second, the positive effects are strongest for high achievers based on a median split of the prebroadband KS1 test scores. The effect for high achievers is 0.00156 percentiles per exchange distance meter (column 4) and 0.00096 for low achievers (column 5). Third, nonwhite pupils have a larger estimate (0.00211 per meter, column 7), but this effect is also less precisely estimated than that for white pupils (0,00114, column 6). Last but not least, only students who are ineligible for free school meals benefit significantly, with an estimated effect of 0.00132 percentiles per meter in comparison to 0.00073 per meter (approximately half the size) for students with free school meal eligibility.

These resulting coefficients are not distinct from each other at conventional levels of statistical significance. However, the overall pattern is of interest, with the most positive effects for girls, high achievers and students who are not eligible for free school meals. This result is consistent with that of Dettling et al. (2018), who find that the impact of high-speed internet on college applications is concentrated among white students with more educated parents and mainly located in urban and high-income areas. Moreover, Malamud et al. (2019) find no significant effects of home internet access on student achievement for students enrolled in low-achieving primary schools in Peru. Potentially, the smaller effect on the below-median prebroadband KS1 test score and free school meal-eligible groups may partly reflect the trouble that struggling students have in developing effective study strategies for learning (Angrist & Lavy, 2009; Fryer Jr, 2011), in our case in a home online environment. While none of the groups have negative effect estimates, these results still speak to the hypothesis that home broadband access might exaggerate existing educational inequality by achievement and family affluence.[27]

[27] Note that we also explore potential non-linearity in the effect of available broadband speed. We split the sample into quartiles based on the distance to the LE station on the slow side of each boundary. Unfortunately, other observable characteristics also differ across these quartiles. As a consequence, we could not generate credible estimates that identify non-linearity net of heterogeneity using this approach.

Table 6: Impact of Exchange Distance on Student Performance by Student Characteristics.

Dependent variable: National exam scores at age 14 (KS3)

	Student Gender		**KS1 Test Score**		**Ethnicity**		**FSM Eligibility**	
	Boys	**Girls**	**High**	**Low**	**White**	**Nonwhite**	**Eligible**	**Ineligible**
	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Nonparametric	0.00098	0.00125*	0.00156**	0.00096	0.00114**	0.00211	0.00073	0.00132***
Estimates	(0.00067)	(0.00068)	(0.00068)	(0.00066)	(0.00051)	(0.00133)	(0.00149)	(0.00050)
Observations	91,106	91,856	89,229	93,763	144,194	38,443	27,496	155,252

Notes: The table shows the fuzzy SRD nonparametric estimates for average test scores on the KS3 for different subgroups of students. FSM stands for Free School Meal. Each coefficient comes from a separate regression, where the running variable is the residential distance to the LE station boundary segment. The window size for the residential distance to the LE station boundary segment is $\pm$ 300 meters, and only boundaries where the exchange distance jumps by at least 300 meters are included. Standard errors clustered at the segment-by-year level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.6 Effects in the longer-run: age-16 KS4 scores

We exploit our student register data to examine if the effects are lasting or fading out in the medium run. For this, we consider the national age-16 Key Stage exams that our cohorts are taking in the years 2007 to 2010. Notably, this period is now extending into post-2008 when alternative technologies became available, potentially diluting the treatment. There are in particular smartphones and fire cable connections that do not show the distance-speed relations of the ADSL technology.

Note that in order to compute the age-16 student performance variables we had to categorize exams into the three subjects English, maths and science. This is because, unlike in the KS3 exam, students can choose additional exams, such as in "further maths". We used the resulting grades to generate an overall percentile grade for each subject, following Lavy et al. (2012). Another reasons that the resulting test scores are less precise compared to the finely-measured age-14 KS3 scores, is that the individual exams are graded on a coarse letter-scale.

Table 7 shows the resulting fuzzy-RDD estimates. We still find statistically significant effects on average scores, comparable in size to the baseline KS3 effects but less precisely estimated. The subject-specific results are in line with those of KS3. These results show that the effect of broadband speed persist beyond KS3, and do not fade out two years later.

Table 7: Impact of Exchange Distance on age-16 KS4 Student Performance

Dependent variable: National KS4 exam scores at age 16 (in percentile ranks)				
	Average Score (1)	**English** (2)	**Maths** (3)	**Science** (4)
Nonparametric Estimates	0.00111* (0.00059)	0.00106* (0.00059)	0.00070 (0.00060)	0.00108* (0.00061)
Observations	138,680	138,680	138,680	138,680

Notes: The table shows the fuzzy SRD nonparametric estimates for average test scores on the KS4 overall, and by subject. Each coefficient comes from a separate regression, where the running variable is the residential distance to the LE station boundary segment. The window size for the residential distance to the LE station boundary segment is $\pm$ 300 meters, and only boundaries where the exchange distance jumps by at least 300 meters are included. Standard errors clustered at the segment-by-year level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5 Broadband Internet Speed and Student Test Scores

Up until this point, we have focused on using the fuzzy SRD to obtain a causal estimate of the effect of distance from the LE station on test scores, that is, β from equation (1). This corresponds to the estimation of LATE. However, we are ultimately interested in the relation between broadband speed and test scores. Unfortunately, postcode-level speed data for our study period do not exist. We use out-of-sample-period data to reverse engineer the distance-speed relation present during our study period.

To investigate this relation, we combine the linked postcode-to-exchange station telecom network data with data from the UK's telecommunication regulator (Ofcom) on the average internet speed experienced by households in 2012 and 2013. These data are explained in Section 2.2.3. The Ofcom dataset is based on the speeds of broadband connections operated by the main operators.[28] To replicate average internet broadband speeds that are plausible for our period of analysis (2005–2008), we use only postcodes with average measure speeds of up to 10 Mbit/s per second, thus excluding superfast broadband connections, which were not available before 2008.

The average internet speed in our resulting dataset is 5.5 Mbit/s and not far from UK consumers' actual average download speeds of 4.7 Mbit/s in 2008 (OfCom, 2009b).[29]

[28]This dataset includes over 13 million connections in 2012 and over 19 million connections in 2013. Annex 1 of the respective Ofcom infrastructure reports provides detailed descriptions (OfCom, 2012; OfCom, 2013).

[29]OfCom (2009b) reports that consumers living in urban areas received average download speeds of 4.3 Mbit/s in comparison with 3.3 Mbit/s among those living in rural areas. Average internet speed varies greatly from country to country. The highest averages can be observed in Asian countries (e.g.,

The dataset provides us with the average speed (modem sync speeds) recorded in speed tests across individual households, which we can geolocate at the level of their full residential postcode. We use this information to link the speed microdata to the telecom network database discussed above to estimate the relationship between residential exchange distances and available internet speed using the fuzzy SRD design.[30]

Table 8: Fuzzy SRD Estimates: Impact of Exchange Distance on Average Broadband Speed.

Dependent variable: Average internet speed (Mbit/s)	(1)	(2)	(3)
Nonparametric Estimates	0.00083***	0.00080***	0.00089***
	(0.00015)	(0.00016)	(0.00018)
Months since ADSL Upgrade			0.0002
			(0.004)
Log Number of Premises			0.167*
			(0.085)
Log Area Square Meters			-0.190**
			(0.087)
Distance Bins	Yes	Yes	Yes
Area & Amenities Controls	No	Yes	Yes
Observations	20,274	20,274	20,274

Notes: The table shows the fuzzy SRD nonparametric estimates for the effect of exchange distance on average internet speed (Mbit/s). The variable "Months since ADSL Upgrade" refers to the number of months since an ADSL upgrade occurred in the LE station catchment area. The variable "Log Number of Premises" refers to the logarithm of the number of premises in the LE station catchment area. The variable "Log Area Square Meters" refers to the logarithm of the square meters of the LE station catchment area. Each coefficient is from a separate regression, where the running variable is the residential distance to the LE station boundary segment. The window size for the residential distance to the LE station boundary is $\pm$ 300 meters. Standard errors clustered at the segment-by-year level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8 shows the fuzzy SRD estimates of the effect of exchange distance on postcode average (download) speed. These estimates are based on an identical empirical strategy as the one used to derive the effects of distance on test scores, as discussed in Section 3.3. Various specifications that include different sets of technological controls result in very similar estimates based on the boundary jumps: for each additional 100 meters in distance to the connected exchange, speed decreases by 0.089 Mbit/s. This

South Korea, Hong Kong, Singapore, and Japan) and Scandinavian countries (e.g., Norway, Sweden, Finland, and Denmark). See Inc. (2015).

[30] After 2008, the broadband internet infrastructure in the UK was updated to significantly less distance-sensitive technologies, such as coaxial cables or fibres. Due to this feature of our dataset, we regard our estimates of the effect of crossing the invisible catchment area boundary from the *Slow* to the *Fast* side on the available internet connection speed as a lower bound (for the test score estimation period). However, in 2012, 75% of broadband subscriptions were still for ADSL technologies, which is sensitive to spatial decay.

coefficient is statistically significant at the 1% level. This provides a causal estimate of the impact of proximity to the LE station on average speed, which we label β_{Speed}. Reassuringly, this speed-distance decay estimate is very similar to that obtained by Ahlfeldt et al. (2017) for ADSL technologies in 2009 and 2010.[31]

By combining the estimate of the impact of proximity to the LE on test scores from equation (1), $\widehat{\beta}$, with $\widehat{\beta_{Speed}}$, we can compute a Wald estimate for the effect of available speed on test scores, ω:

$$\omega = \frac{\beta}{\beta_{Speed}} \tag{6}$$

Applying equation (6) to compute ω yields an effect of $\widehat{\omega} = 1.37$. This implies that for each increase in broadband speed of one Mbit/s (corresponding to an increase of 20% with respect to the baseline average speed) test scores increase on average by 1.37 percentile points. This average effect is equivalent to approximately 5% of a standard deviation in the national test score distribution. The standard error of $\widehat{\omega}$ is obtained using the Delta method and estimated with high precision (i.e., the p-value is < 0.01).[32] Even ω, which is the policy-relevant parameter, can still be interpreted as a reduced-form effect, as it reflects the effect of broadband availability – and not usage – on test scores. [33]

Finally, to assess the magnitude of the effect of broadband usage on test scores, we complement this Wald estimate with survey data evidence on broadband usage and information about the extent to which students used the internet for school work for our study period. Using data from Oxford Internet Surveys (OxIS) in several years, we compute broadband take-up for the group of interest in our sample years. We exploit this database because it allows us to be precise in the definition of the relevant population under study: English households with home broadband that had children aged

[31] Ahlfeldt et al. (2017) use broadband speed tests carried out by individuals in 2009 and 2010 from a private company to estimate the speed-distance relationship for different internet technologies. The estimates of Ahlfeldt et al. (2017) imply that actual download speed decreases 0.085 Mbit/s for each additional 100 meters further from the LE station for ADSL technologies (for distances ranging 1000-3000 meters from the LE). In addition, this estimate is not dissimilar to estimates provided by Ofcom using survey data in 2009 (OfCom, 2009b)

[32] See Appendix Section A.4 for more details.

[33] One of the channels through which we can surmise that broadband speed affects test scores is via broadband subscriptions take-up if it was unbalanced across boundary-segments. Using Ofcom postcode-level data for 2013, we test whether the effect of distance to LE on postcode take-up rates (the number of active broadband lines divided by the number of premises) changes across boundaries. We find precisely estimated zeroes.

between 14 and 17 years old.[34] We find that the weighted average of broadband usage among these families between 2005 and 2009 is 69.5%. These figures square with the finding reported by Livingstone & Bober (2005) that the vast majority of children used the internet at home, and most of them devoted the use time to do work for school or college. These authors' figures show that in the UK, 84% of 9- to 19-year-olds used the internet daily or weekly in 2005. Among those, 90% declared that they used the internet to do work for school or college.

Using these calculations, which refer to the exact age group that we study, scaling our Wald estimate by broadband take-up shows that for each increase in broadband speed of one Mbit/s, test scores increase on average by 1.97 percentile points. This increase is equivalent to approximately 7% of a standard deviation in the national test score distribution.

The size of our headline estimate, an effect of 5% of a standard deviation (7% if we scale our estimate by broadband take-up), can be interpreted as a cumulative effect of having faster internet speed over six years on average. Assuming that effects are linear over time, the average effect is equivalent to an increase of 0.8% of a standard deviation in the national test score distribution (1.22% if we scale our estimate by broadband take-up) per year of exposure to faster internet speed. Our estimates are not small in the context of the education literature. For instance, teachers have been identified as one of the most important factors in test performance, with a one standard deviation increase in teaching quality improving test scores by 0.1–0.2 standard deviations (Rivkin et al., 2005; Slater et al., 2012; Chetty et al., 2014). Peer effects are smaller at between 0.01 and 0.08 standard deviations (e.g., in the English context Lavy et al., 2012; Gibbons & Telhaj, 2016), and various dimensions of neighborhood quality have even smaller effects on test scores (e.g., Jacob, 2004; Sanbonmatsu et al., 2006; Gibbons et al., 2013, 2017). Therefore, in the context of the education literature, our estimates represent medium-sized effects.

[34] More aggregated datasets used in other studies, such as Eurostat or Ofcom data, produced similar numbers. The drawback of using these databases is that they, unlike OxIS, do not allow us to refine our calculations to the specific population of interest.

6 Mechanisms

The theoretical time taken to perform various online activities changes substantially with the internet connection speed (OfCom, 2010).[35] Faster internet speed increases the amount of information obtainable from the web or shared via the network per unit of time, affecting education performance in two ways. First, students might become more productive per unit of time invested in studying if online learning resources enhance learning. For this to be feasible, students need to have access to relevant study productivity-improving online tools, be aware of them, and use them appropriately. Between 2005 and 2008, multiple online resources offering educational support became available to teenagers. For instance, *YouTube* was launched in 2005, and soon after, educational institutions and individuals began uploading educational content with keen interest from the platform to help educators. Multiple popular education-oriented channels existed in this period (e.g., *Khan Academy*, *Smarter Every Day*, *TED Talks*).[36] On the other hand, with better connections, the possibility of engaging in social media and other distracting activities increases, potentially decreasing the amount of time spent doing school work. Students could also make use of early social media tools (such as *Microsoft Messenger*) to share information and study together or visit sites such as *Sparknotes* to download essay-writing content. In fact, amongst the list of most viewed websites at the time were *Google, MSN Windows, Wikipedia* or *YouTube* (OfCom, 2007).

During our period of study, English children had access to and used the internet from a computer at home. Livingstone & Bober (2005) show that in autumn 2004, 75% children aged 9-19 years old had access to the internet at home. OfCom (2006) and OfCom (2008) provide similar figures for 2005 and 2007, with 68% and 77% of children aged 12-15 having access to the internet at home, respectively. According to OfCom

[35] For instance, downloading a 250 Kbit/s web page takes 1 and 0.3 seconds with a provider of 2 Mbit/s and 8 Mbit/s, respectively. Besides, this difference grows exponentially with the online activity provided. Downloading a DVD-quality film (4GB) takes 4 hours and 48 minutes and 1 hour and 11 minutes with a provider of 2 Mbit/s and 8 Mbit/s, respectively. See Figure 4.1 of OfCom (2010).

[36] The founder of the *Khan Academy* started posting videos in 2006 and created the academy channel in 2008. His videos became very popular, attracting tens of thousands of viewers every month. *Smarter Every Day* started its educational videos in 2007, and *TED Talks* began sharing talks in 2006. The demand for educational videos led to the creation of *YouTube EDU* in 2009 as a repository for its educational content. The number of articles in encyclopedia-type sites, such as *Wikipedia*, also increased exponentially since 2004. This free online encyclopedia has been found to be a relevant tool to improve learning (Derksen et al., 2022).

(2008), the internet consumption for children aged 12-15 at home doubled between 2005 and 2007, from 7.1 to 13.8 average weekly hours. Hence, students were accessing to the internet at home, where they could use study-enhancing online tools.

Table 9: SRD Estimates: Impact of Exchange Distance on Broadband Take-up and Computer Use at Home.

	A. Internet and Broadband Take-up (Household level)			B. Computer Use (Individual level)			
	Computer at home	**Internet Use**	**Broadband at home**	**Work**	**Playing Games**	**Hobbies**	**Educational Work**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fast Side	-0.907	-0.257	-0.101	0.423	1.411	1.499*	0.200
	(0.735)	(0.875)	(0.530)	(0.713)	(0.954)	(0.866)	(0.668)
Age 15-18				-0.036	0.136	0.198*	0.346***
				(0.043)	(0.151)	(0.101)	(0.091)
Fast Side*Age 15-18				-0.018	0.155	-0.015	0.274**
				(0.083)	(0.185)	(0.163)	(0.126)
Observations	556	538	438	426	426	426	426

Notes: The table shows the SRD nonparametric estimates for the internet and broadband take-up (columns 1-3) and computer use (columns 4-6) for the specification of Equation (4). Columns 4 to 6 show the SRD nonparametric estimates for different computer use and interacted with the relevant population (individuals between 15-18 years old) conditional on having broadband at home. Columns report different specifications. Each coefficient comes from a separate regression, where the running variable is the residential distance to the LE station's boundary. The window size for the residential distance to the LE station boundary is ± 1000 meters. We only include boundaries where the distance jumps by at least 300 meters. Standard errors clustered at the boundary-by-year level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 9 shows the estimates of the effect of exchange distance on broadband take-up and computer use at home. These estimates are based on the empirical strategy discussed in Section 3.5. Panel A (columns 1-3) shows that households on the *Fast* side do not significantly enjoy higher Internet use nor more broadband take-up at home (relative to those on the *Slow* side). This result is reassuring to the fact that our paper estimates changes in the speed of the broadband connection (the intensive margin) and not simply access to high-speed internet (the extensive margin).

Panel B displays the effects on different margins of computer use at home at the individual level for respondents of different ages. Access to the Internet and broadband take-up are not the only important dimensions, but also teenagers' internet use. Parents perceive children's internet use with concern, with the vast majority of parents (75%) having rules in the household about access and internet content (Livingstone & Bober, 2005). Columns (5) and (6) of Table 9 show that teenagers of ages 15-18 are using the computer more often than the other individuals in the household for potential distraction activities, such as playing games and hobbies. However, teenagers on the *Fast* side do not significantly use the computer more on potential distraction activities (relative to those on the *Slow* side). This result seems to rule out the hypothesis of these students decreasing the amount of time spent doing school work due to the more frequent engagement in social media and other distracting activities. Moreover, column (7) shows that teenagers use the computer more often than individuals aged 19-55 for educational work. We find that teenagers just on the *Fast* side are 27% more likely to use their computer for educational work (relative to those on the *Slow* side). Results are statistically significant at the 5% level. Parents of children aged between 12 and 15 years old self-report the internet as the media platform most beneficial for their kids (OfCom, 2008). In addition, in 2004, 73% of parents believed that the internet could help their children to do better at school and help them learn (Livingstone & Bober, 2005).

These results support the hypothesis that the mechanisms behind the positive effects on student test scores seem to be a more education-oriented use of the internet combined with a controlled use of distraction activities.

7 Conclusion

This paper investigates the impact of ICT quality at home on educational outcomes, related to human capital accumulation relevant to labour market performance. We use a fuzzy SRD approach to present estimates of the causal net effect of available broadband speed on the test scores of 14-year-old pupils in England. We estimate that an increase in available home broadband speed of 1 Mbit/s (corresponding to an increase of 20% with respect to the baseline average speed) leads to an increase in student test scores of approximately 5% of a standard deviation. The estimate increases to 7% when we adjust the effects for the population of interest. We find larger responses from students of more advantaged socio-economic background. We rule out that the impacts are driven by school characteristics or biased by other mediating technological factors. Results do not fade out two years later, as we show significant effects on age-16 test scores for the same cohort of students. An examination of the mechanisms behind our findings shows that the beneficial effects of faster internet speed are driven by a higher education-oriented use of the internet combined with no increase in the use of distraction activities.

A direct implication from this heterogeneity in the effects is that better broadband connections at home contribute to educational inequalities. This highlights the importance of policies aimed at improving the internet learning environment at home, particularly for low-SES households. Programs that focus on improving both access to fast connections and digital literacy with a special focus on low-income families, such as *BroadbandUSA*, *Internet Essentials* or the *ConnectHome* program, will likely be the most beneficial to improve student learning and reduce education inequalities.[37] Providing low-SES parents' with more appropriate tools to understand the risks and benefits of access to the internet seems key to improving their ability to help their children take full advantage of the EduTech potential.

Our findings are even more pertinent in the context of the recent COVID-19 pandemic, in which many countries closed schools for several months and home online learning became a major substitute for in-classroom teaching. The pandemic brought the consequences of unequal access to technology to the spotlight, as families with fewer resources could not provide their children with the appropriate infrastructure to engage in digital school work (Andrew et al., 2020; Bacher-Hicks et al., 2021; Stantcheva, 2022). Our results highlight the value added to broadband investments and the importance of ensuring universal access and digital skills to mitigate increases in inequalities in educational opportunity in times of regular schooling.

[37] In addition, these type of interventions can have positive direct effects on improving low-income families labor market prospects. For instance, Zuo (2021) shows that the *Internet Essentials* program increased employment rates and earnings for qualifying low-income families. Gallego et al. (2020) show that children whose parents receive information about their internet use spend less time online.

References

Ahlfeldt, G., Koutroumpis, P., & Valletti, T. (2017). Speed 2.0: Evaluating access to universal digital highways. *Journal of the European Economic Association*, 15(3), 586–625.

Akerman, A., Gaarder, I., & Mogstad, M. (2015). The skill complementarity of broadband internet. *The Quarterly Journal of Economics*, 130(4), 1781–1824.

Amaral-Garcia, S., Nardotto, M., Propper, C., & Valletti, T. (2019). Mums Go Online: Is the Internet Changing the Demand for Healthcare? *The Review of Economics and Statistics*, (pp. 1–45).

Andrew, A., Cattan, S., Costa Dias, M., Farquharson, C., Kraftman, L., Krutikova, S., Phimister, A., & Sevilla, A. (2020). Inequalities in children's experiences of home learning during the covid-19 lockdown in england. *Fiscal studies*, 41(3), 653–683.

Angrist, J. & Lavy, V. (2002). New Evidende on Classroom Computers and Pupil Learning. *Economic Journal*, 112(482), 735–765.

Angrist, J. & Lavy, V. (2009). The effects of high stakes high school achievement awards: Evidence from a randomized trial. *American economic review*, 99(4), 1384–1414.

Angrist, J. D., Lavy, V., Leder-Luis, J., & Shany, A. (2019). Maimonides' rule redux. *American Economic Review: Insights*, 1(3), 309–24.

Angrist, N., Djankov, S., Goldberg, P. K., & Patrinos, H. A. (2021). Measuring human capital using global learning data. *Nature*, 592(7854), 403–408.

Bacher-Hicks, A., Goodman, J., & Mulhern, C. (2021). Inequality in household adaptation to schooling shocks: Covid-induced online learning engagement in real time. *Journal of Public Economics*, 193, 104345.

Barro, R. J. (2001). Human capital and Growth. *American Economic Review*, 91(2), 12–17.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of political economy*, 70(5, Part 2), 9–49.

Bellou, A. (2015). The Impact of Internet Diffusion on Marriage Rates: Evidence from the Broadband Market. *Journal of Population Economics*, 28(2), 265–297.

Belo, R., Ferreira, P., & Telang, R. (2014). Broadband in school: Impact on student performance. *Management Science*, 60(2), 265–282.

Belo, R., Ferreira, P., & Telang, R. (2016). Spillovers from wiring schools with broadband: the critical role of children. *Management Science*, 62(12), 3450–3471.

Beuermann, D. W., Cristia, J., Cueto, S., Malamud, O., & Cruz-Aguayo, Y. (2015). One laptop per child at home: Short-term impacts from a randomized experiment in peru. *American Economic Journal: Applied Economics*, 7(2), 53–80.

Bhuller, M., Havnes, T., Leuven, E., & Mogstad, M. (2013). Broadband Internet: An Information Superhighway to Sex Crime? *Review of Economic Studies*, 8(4), 1237–1266.

Black, S. E. (1999). Do Better Schools Matter? Parental Valuation of Elementary Education. *Quarterly Journal of Economics*, 114(2), 577–599.

Bulman, G. & Fairlie, R. W. (2016). Technology and Education: Computers, Software, and the Internet. *Handbook of the Economics of Education*, 5, 239–280.

Burgess, S. M. (2016). Human capital and education: The state of the art in the economics of education.

Campante, F., Durante, R., & Sobbrio, F. (2018). Politics 2.0: The multifaceted effect of broadband internet on political participation. *Journal of the European Economic Association*, 16(4), 1094–1136.

Cattaneo, M. D., Jansson, M., & Ma, X. (2018). Manipulation testing based on density discontinuity. *The Stata Journal*, 18(1), 234–261.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9), 2593–2632.

Clement, J. (2020). Global digital population as of july 2020.

Commission, S. M. et al. (2021). State of the nation 2021: Social mobility and the pandemic.

Cristia, J., Ibarrarán, P., Cueto, S., Santiago, A., & Severín, E. (2017). Technology and child development: Evidence from the one laptop per child program. *American Economic Journal: Applied Economics*, 9(3), 295–320.

Czernich, N., Falck, O., Kretschmer, T., & Woessmann, L. (2011). Broadband infrastructure and economic growth. *The Economic Journal*, 121(552), 505–532.

Das, J., Singh, A., & Chang, A. Y. (2022). Test scores and educational opportunities: Panel evidence from five low-and middle-income countries. *Journal of Public Economics*, 206, 104570.

Derksen, L., Michaud-Leclerc, C., & Souza, P. C. (2022). Restricted access: How the internet can be used to promote reading and learning. *Journal of Development Economics*, (pp. 102810).

Dettling, L. J., Goodman, S., & Smith, J. (2018). Every little bit counts: The impact of high-speed internet on the transition to college. *Review of Economics and Statistics*, 100(2), 260–273.

Escueta, M., Quan, V., Nickow, A. J., & Oreopoulos, P. (2017). *Education technology: An evidence-based review*. Technical report, National Bureau of Economic Research.

Faber, B., Sanchis-Guarner, R., & Weinhardt, F. (2015). ICT and Education: Evidence from Student Home Addresses. NBER Working Paper Num. 21306.

Fairlie, R. W. & Robinson, J. (2013). Experimental Evidence on the Effects of Home Computers on Academic Achievement among Schoolchildren. *American Economic Journal: Applied Economics*, 5(3), 211–240.

Falck, O., Gold, R., & Heblich, S. (2014). E-Lections: Voting Behavior and the Internet. *American Economic Review*, 107(7), 2238–2265.

Falck, O., Mang, C., & Woessmann, L. (2018). Virtually no effect? different uses of classroom computers and their effect on student achievement. *Oxford Bulletin of Economics and Statistics*, 80(1), 1–38.

Fryer Jr, R. G. (2011). Financial incentives and student achievement: Evidence from randomized trials. *The Quarterly Journal of Economics*, 126(4), 1755–1798.

Gallego, F. A., Malamud, O., & Pop-Eleches, C. (2020). Parental monitoring and children's internet use: The role of information, control, and cues. *Journal of Public Economics*, 188, 104208.

Gavazza, A., Nardotto, M., & Valletti, T. (2019). Internet and politics: Evidence from uk local elections and local government policies. *The Review of Economic Studies*, 86(5), 2092–2135.

Gelman, A. & Imbens, G. (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, 37(3), 447–456.

Geraci, A., Nardotto, M., Reggiani, T., & Sabatini, F. (2022). Broadband internet and social capital. *Journal of Public Economics*, 206, 104578.

Gibbons, S., Silva, O., & Weinhardt, F. (2013). Everybody needs good neighbours? evidence from students' outcomes in england. *The Economic Journal*, 123(571), 831–874.

Gibbons, S., Silva, O., & Weinhardt, F. (2017). Neighbourhood Turnover and Teenage Attainment. *Journal of the European Economic Association*, 15(4), 746–783.

Gibbons, S. & Telhaj, S. (2016). Peer effects: Evidence from secondary school transition in england. *Oxford Bulletin of Economics and Statistics*, 78(4), 548–75.

Goolsbee, A. & Guryan, J. (2006). The Impact of Internet Subsidies in Public Schools. *Review of Economics and Statistics*, 88(2), 336–347.

Hanushek, E. A. & Woessmann, L. (2012). Do better schools lead to more growth? cognitive skills, economic outcomes, and causation. *Journal of economic growth*, 17(4), 267–321.

Hanushek, E. A. & Woessmann, L. (2015). *The knowledge capital of nations: Education and the economics of growth*. MIT press.

Inc., A. T. (2015). *Akamai's [state of the internet]. Q1 2015 report*. Technical Report 1.

Jacob, B. A. (2004). Public housing, housing vouchers, and student achievement: Evidence from public housing demolitions in chicago. *American Economic Review*, 94(1), 233–258.

Lakdawala, L. K., Nakasone, E., & Kho, K. (2023). Dynamic impacts of school-based internet access on student learning: Evidence from peruvian public primary schools. *American Economic Journal: Economic Policy*.

Lavy, V., Silva, O., & Weinhardt, F. (2012). The Good, the Bad, and the Average: Evidence on Ability Peer Effects in Schools. *Journal of Labor Economics*, 30(2), 367 – 414.

Leuven, E. & Løkken, S. A. (2020). Long-term impacts of class size in compulsory school. *Journal of Human Resources*, 55(1), 309–348.

Livingstone, S. & Bober, M. (2005). Uk children go online: Final report of key project findings.

Machin, S., McNally, S., & Silva, O. (2007). New Technology in Schools: Is There a Payoff? *Economic Journal*, 117(522), 1145–1167.

Malamud, O., Cueto, S., Cristia, J., & Beuermann, D. W. (2019). Do children benefit from internet access? experimental evidence from peru. *Journal of Development Economics*, 138, 41–56.

Malamud, O. & Pop-Eleches, C. (2011). Home Computer Use and the Development of Human Capital. *Quarterly Journal of Economics*, 126(2), 987–1027.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics*, 142(2), 698–714.

McKeay, M. (2017). *"Q2 2017 state of the internet security report*. Technical report, Akamai, Tech. Rep., 2017.[Online]. Available: https://www. akamai. com/us/en . . . .

Mincer, J. (1974). Schooling, experience, and earnings. human behavior & social institutions no. 2.

Nardotto, M., Valletti, T., & Verboven, F. (2015). Unbundling the Incumbent: Evidence from UK Broadband. *Journal of the European Economic Association*, 13(2), 330–362.

OfCom (2006). *Media Literacy Audit: Report on media literacy amongst children*. Technical report.

OfCom (2007). *The Communications Market: Broadband*. Research document.

OfCom (2008). *Media Literacy Audit: Report on UK children's media literacy*. Research document.

OfCom (2009a). *UK broadband speeds 2008. Consumer experience of broadband performance: initial findings*. Research document, OfCom.

OfCom (2009b). *UK broadband speeds 2009. Consumers' experience of fixed-line broadband performance*. Research report, OfCom.

OfCom (2010). *UK fixed broadband speeds, November/December 2010*. Research document, OfCom.

OfCom (2012). *2012 UK Communications Infrastructure Report*. Report.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, 73(2), 417–458.

Roser, M., Ritchie, H., & Ortiz-Ospina, E. (2020). Internet. *OurWorldInData.org*.

Rouse, C. E. & Krueger, A. B. (2004). Putting computerized instruction to the test: a randomized evaluation of a "scientifically based" reading program. *Economics of Education Review*, 23(4), 323–338.

Ryan, C. & Sibieta, L. (2010). Private schooling in the uk and australia.

Sanbonmatsu, L., Kling, J., Duncan, G., & Brooks-Gunn, J. (2006). Neighborhoods and academic achievement: Results from the moving to opportunity experiment. *The Journal of Human Resources*, XLI.

Slater, H., Davies, N. M., & Burgess, S. (2012). Do teachers matter? measuring the variation in teacher effectiveness in england*. *Oxford Bulletin of Economics and Statistics*, 74(5), 629–645.

Stantcheva, S. (2022). *Inequalities in the Times of a Pandemic*. Technical report, National Bureau of Economic Research.

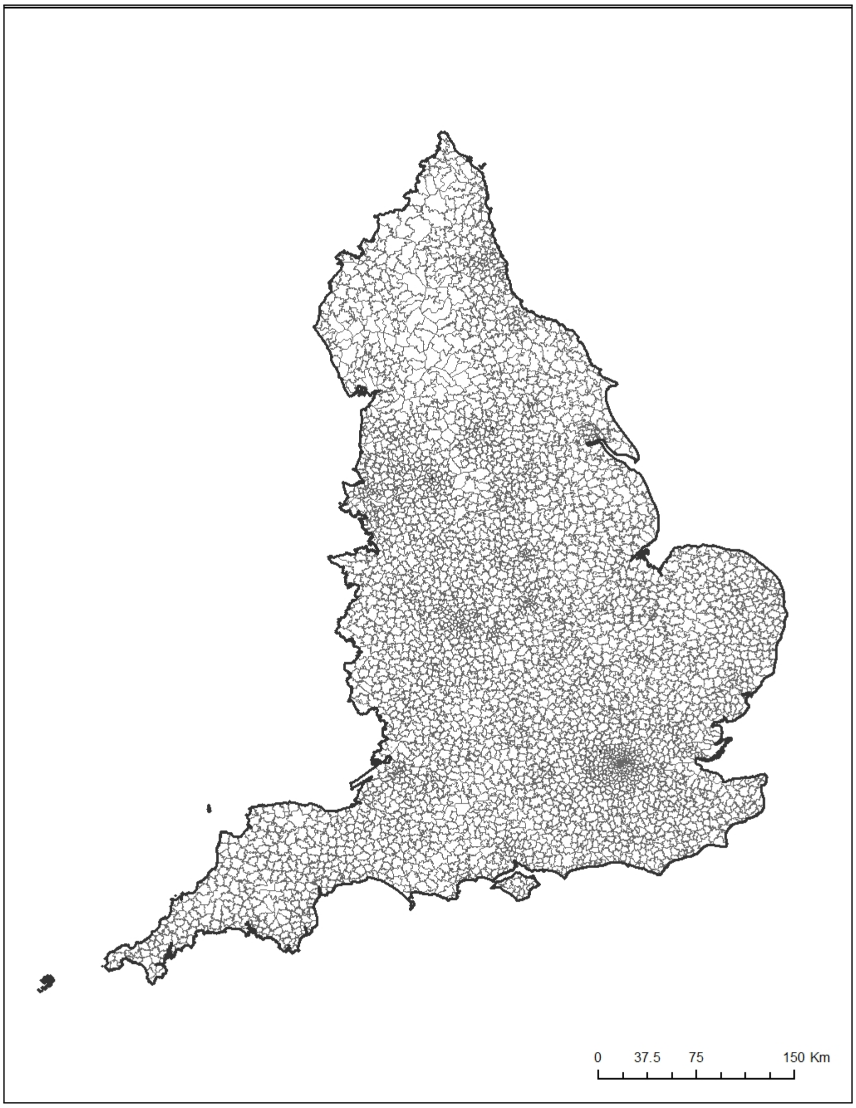
Vigdor, J., Ladd, H., & Martinez, E. (2014). Scaling the Digital Divide: Home Computer Technology and Student Achievement. *Economic Inquiry*, 52(3), 1103–1119.

Zuo, G. W. (2021). Wired and hired: Employment effects of subsidized broadband internet for low-income americans. *American Economic Journal: Economic Policy*, 13(3), 447–82.

Appendix Figures and Tables

A.1 Telephone Local Exchange Areas in England

Figure A.1: Telephone Local Exchange Areas in England



Notes: There are 3,978 local exchange areas in England (of which 3,925 are completely contained in England). The average area is of 34 square kilometres, and serves an average of 5,830 premises (with 93% of them residential).

A.2 Validity of the Research Design

Table A.1: McCrary (2008) and Cattaneo et al. (2018) Test for Manipulation of the Forcing Variable for the Different Treatment subamples.

A. McCrary Test				
Distance (meters)	**Log Difference in frequency bins**	**Z-stat**	**Bandwidth**	**Bin size**
250 meters	.267 (0.206)	1.29	40.14	.556
300 meters	-.024 (0.068)	.347	44.92	.604
350 meters	-.043 (0.048)	.88	50.22	.65
400 meters	-.038 (0.04)	.936	54.5	.69
450 meters	-.038 (0.035)	1.07	59.69	.73
500 meters	-.041 (0.031)	1.33	66.7	.77
B. Cattaneo et al. (2018): RD Manipulation Test using local polynomial density estimation:				
Bias-corrected Density Estimate to the left of the cutoff	.00035 (0.000017)			
Bias-corrected Density Estimate to the right of the cutoff	.00033 (0.000018)			
T-test for bias-corrected density test	-0.7886			
P-value for bias-corrected density test	0.4303			

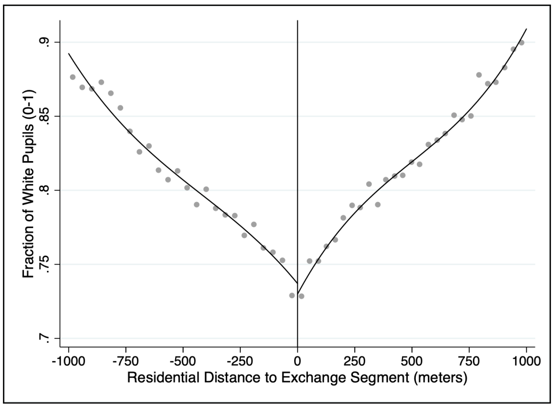
Notes: This table show the McCrary (2008) (Panel A) and Cattaneo et al. (2018) test for manipulation of the forcing variable. The McCrary (2008) test is performed separately for each treatment sample. The table columns of Panel A show the estimated discontinuity in the density function of the assignment variable at the threshold, its standard error (in parentheses), the associated z-statistic, the estimated optimal bandwidth, bin size and the number of observations. The optimal bandwidth and bin size are obtained using the selection procedure proposed by McCrary (2008). The table rows of Panel B present the bias corrected density estimate to the left and right of the LE boundaries, the t-test for the bias-corrected density test and the p-value of the test.

Table A.2: Sharp-SRD Estimates: The Impact Exchange Distance on Student Performance

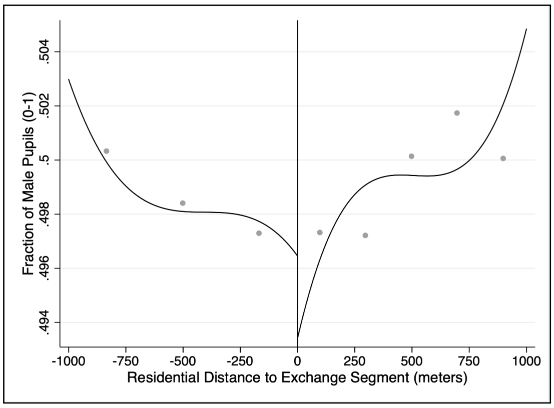
	(1)	(2)	(3)	(4)	(5)
Effect of exchange distance on national age-14 KS3 exam scores (in meters)					
Non-Parametric	0.00043***	0.00034**	0.00027**	0.00025*	0.00024*
Estimates	(0.00014)	(0.00014)	(0.00013)	(0.00013)	(0.00013)
Student Pre-Internet Score KS1	Yes	Yes	Yes	Yes	Yes
Student Controls	No	Yes	Yes	Yes	Yes
Time-Variant Area & Amenities	No	No	Yes	Yes	Yes
Time-Invariant Area & Amenities	No	No	No	Yes	Yes
School Controls	No	No	No	No	Yes
Observations	183,892	183,892	183,892	183,892	183,892

Notes: The table shows the sharp SRD non-parametric estimates for test scores in KS3. Columns report different specifications. Each coefficient comes from a separate regression, where the running variable is the residential distance to the LE station boundary. The window size for the residential distance to the LE station boundary is ± 300 meters. We only include boundaries where the distance jumps by at least 300 meters. Standard errors clustered at the segment-by-year level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

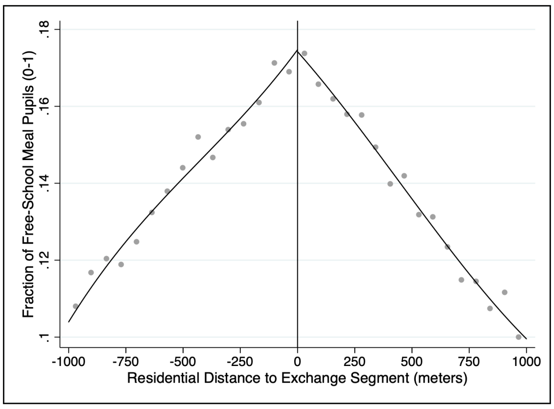
Figure A.2: Students Characteristics.



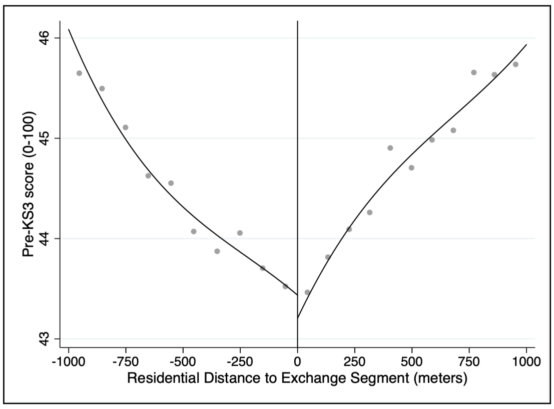
(a) Fraction of White Pupils (0-1)



(b) Fraction of Male Pupils (0-1)



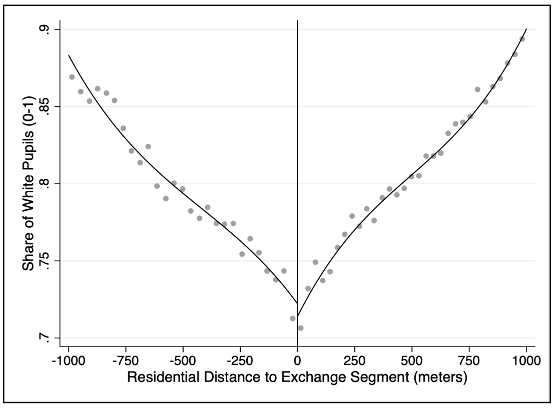
(c) Fraction of Free-School Meal Pupils (0-1)



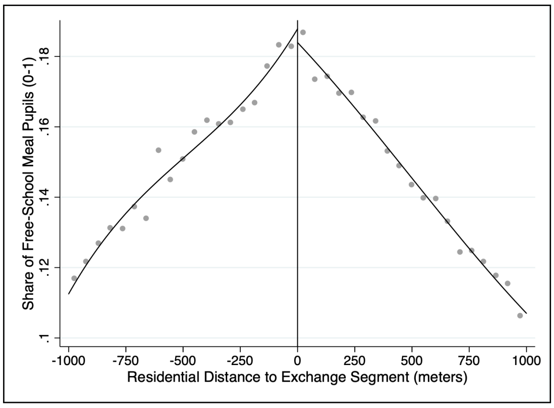
(d) Pre-KS3 score (0-100)

Notes: The dots represent the fraction of white pupils, the fraction of male pupils, the fraction of free-school meal pupils and the pre-internet score per interval of residential distance to the exchange segment boundaries. The solid lines are fitted values from a third order polynomial approximation, which is estimated separately on both sides of the cutoffs. "Residential Distance to Exchange Segment (meters)" refers to the residential distance to the exchange segment boundaries. Black vertical lines identify the LE station boundaries.

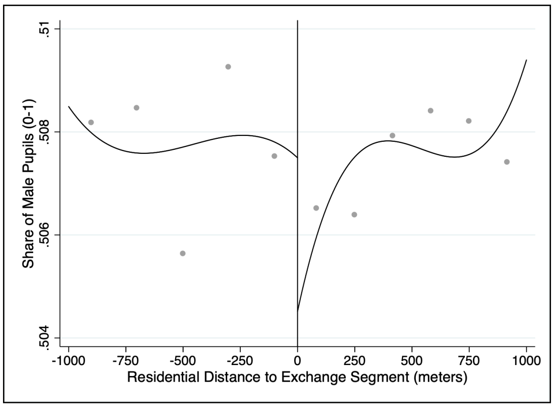
Figure A.3: Density & Area Socio-Economics.



(a) Share of White Pupils (0-1)



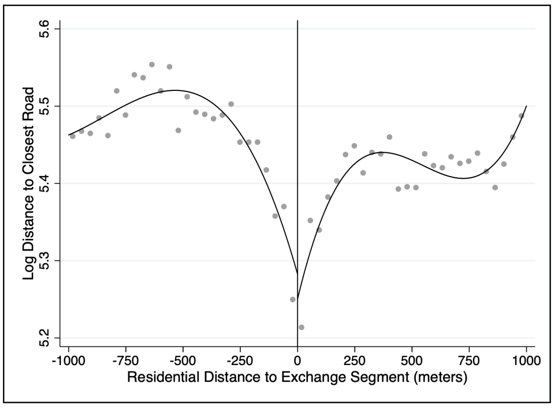
(b) Share of Free-School Meal Pupils (0-1)



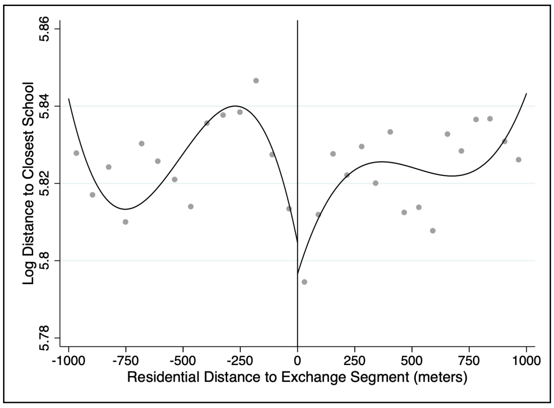
(c) Share of Male Pupils (0-1)

Notes: The dots represent the share of white pupils, share of free-school meal and share of male pupils per interval of residential distance to the exchange segment boundaries.. The solid lines are fitted values from a third order polynomial approximation, which is estimated separately on both sides of the cutoffs. "Residential Distance to Exchange Segment (meters)" refers to the residential distance to the exchange segment boundaries. Black vertical lines identify the LE station boundaries

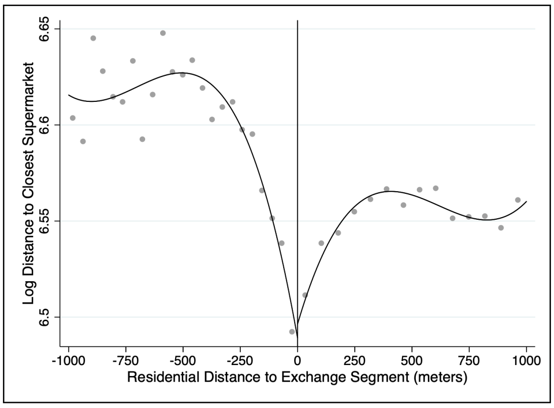
Figure A.4: Amenities.



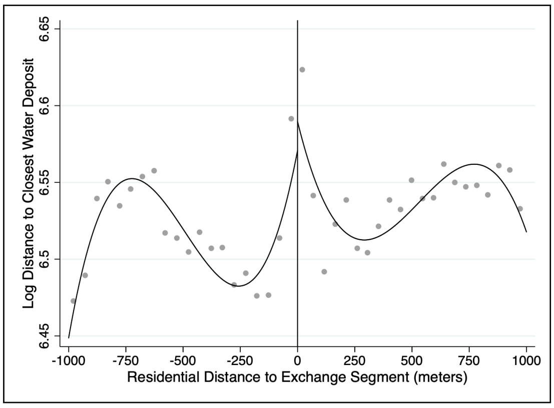
(a) Log Distance to Closest Road



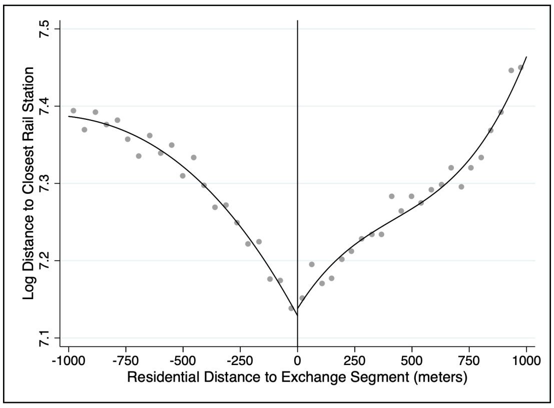
(b) Log Distance to Closest School



(c) Log Distance to Closest Supermarket



(d) Log Distance to Closest Water Deposit



(e) Log Distance to Closest Rail Station

Notes: The dots represent the logarithm distance to closest road, school, supermarket, water deposit and rail station per interval of residential distance to the exchange segment boundaries. The solid lines are fitted values from a third order polynomial approximation, which is estimated separately on both sides of the cutoffs. "Residential Distance to Exchange Segment (meters)" refers to the residential distance to the exchange segment boundaries. Black vertical lines identify the LE station boundaries.

A.3 Robustness of Baseline Estimates

Table A.3: Sample Choice Based on Distance to Segment-Boundary

	Distance (meters)						
	100	**150**	**200**	**300**	**500**	**750**	**1000**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Avg. jump in distance to connected exchange when crossing boundary to the *Fast* side (in m)							
Non-Parametric	-779.4***	-794.1***	-805.9***	-816.6***	-820.3***	-828.5***	-832.9***
Estimates	(19.7)	(18.4)	(16.7)	(15.9)	(15.4)	(15.3)	(15.3)
B. Effect of exchange distance on national age-14 KS3 exam scores (in percentile ranks)							
Non-Parametric	0.00157**	0.00154**	0.00134***	0.00122**	0.00092**	0.00101**	0.00100**
Estimates	(0.00066)	(0.00060)	(0.00051)	(0.00047)	(0.00045)	(0.00043)	(0.00043)
Observations	49,303	85,182	119,55	183,892	305,404	431,709	530,773

Notes: The table shows the fuzzy SRD non-parametric estimates for the average distance to LE station (Panel A) and test scores in KS3 (Panel B) with different samples based on the distance to segment-boundary. Columns report different specifications. Each coefficient comes from a separate regression, where the running variable is the residential distance to the LE station boundary. We only include boundaries where the distance jumps by at least 300 meters. Standard errors clustered at the segment-by-year level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4: Sample Choice Based on Magnitude of Jump across Segment-Boundary

	Minimum Boundary Jump (in meters)					
	100	**200**	**250**	**300**	**400**	**500**
	(1)	(2)	(3)	(4)	(5)	(6)
A. Avg. jump in distance to connected exchange when crossing boundary to the *Fast* side (in m)						
Non-Parametric	-653.788***	-731.578***	-775.172***	-816.649***	-915.704***	-999.089***
Estimates	(13.477)	(14.451)	(15.254)	(15.927)	(17.194)	(18.996)
B. Effect of exchange distance on national age-14 KS3 exam scores (in percentile ranks)						
Non-Parametric	0.00064	0.00080*	0.00116**	0.00122**	0.00131***	0.00135***
Estimates	(0.00051)	(0.00049)	(0.00048)	(0.00047)	(0.00045)	(0.00046)
Observations	671,601	601,366	561,921	530,773	463,692	396,354

Notes: The table shows the fuzzy SRD non-parametric estimates for the average distance to LE station (Panel A) and test scores in KS3 (Panel B) with different samples based on the minimum boundary jump across the segment-boundary. Columns report different specifications. Each coefficient comes from a separate regression, where the running variable is the residential distance to the LE station boundary. The window size for the residential distance to the LE station boundary is $\pm$ 300 meters. Standard errors clustered at the segment-by-year level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: Fuzzy-SRD Estimates: The Impact of Crossing the Broadband Boundary on Student Performance by Clustering choices for standard errors.

	(1)	(2)	(3)	(4)	(5)
Non-Parametric	0.00122**	0.00122**	0.00122**	0.00122***	0.00122**
Estimates	(0.00047)	(0.00032)	(0.00052)	(0.00047)	(0.00051)
Segment-Year Level	Yes	No	No	No	No
Segment and Year Level	No	Yes	No	No	No
Segment Level	No	No	Yes	No	No
Boudary-Year Level	No	No	No	Yes	No
Boundary Level	No	No	No	No	Yes
Observations	183,892	183,892	183,892	183,892	183,892

Notes: The table shows the fuzzy SRD non-parametric estimates for test scores in KS3 by different clustering choice for standard errors. Columns report different specifications. Each coefficient comes from a separate regression, where the running variable is the residential distance to the LE station boundary. The window size for the residential distance to the LE station boundary is $\pm$ 300 meters. We only include boundaries where the distance jumps by at least 300 meters. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.4 Delta Method Calculations

We use the test statistics:

$$\frac{\frac{\widehat{\beta}}{\widehat{\beta_{Speed}}}}{SD(\frac{\widehat{\beta}}{\widehat{\beta_{Speed}}})} \sim N(0,1) \tag{A.1}$$

Assuming that our estimates are independent, we can compute:

$$VAR(\frac{\widehat{\beta}}{\widehat{\beta_{Speed}}}) = \frac{1}{(\widehat{\beta_{Speed}})^2} * (VAR(\widehat{\beta}) + (\frac{\widehat{\beta}}{\widehat{\beta_{Speed}}})^2 * VAR(\widehat{\beta_{Speed}})) \tag{A.2}$$

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

1978	Luca Macedoni John Morrow Vladimir Tyazhelnikov	Firms in product space: Adoption, growth and competition
1977	Gonzalo Nunez-Chaim Henry G. Overman Capucine Riom	Does subsidising business advice improve firm performance? Evidence from a large RCT
1976	Julian Alves Bruno Serra Jason Greenberg Yaxin Guo Ravija Harjai John Van Reenen	Labour market power: New evidence on Non-Compete Agreements and the effects of M&A in the UK
1975	Virginia Minni	Global managers, local workers: Wage setting inside a multinational firm
1974	Jonathan Colmer Suvy Qin John Voorheis Reed Walker	The changing nature of pollution, income and environmental inequality in the United States
1973	Italo Colantone Gianmarco I.P. Ottaviano Piero Stanig	The social footprint of globalization: Towards the introduction of strategic industries in quantitative trade models
1972	Viet Nguyen-Tien Robert J.R. Elliott Eric Strobl Chengyu Zhang	Estimating the longevity of electric vehicles: What do 300 million MOT test results tell us?
1971	Michael Amior Jan Stuhler	Immigration, monopsony and the distribution of firm pay
1970	Jonathan Colmer David Lagakos Martin Shu	Is the electricity sector a weak link in development?

1969	Gaia Dossi Marta Morando	Political ideology and innovation
1968	Natalie Irmert Jan Bietenbeck Linn Mattisson Felix Weinhardt	Autonomous schools, achievement and segregation
1967	Stephen B. Billings Adam Soliman	The erosion of homeownership and minority wealth
1966	Pawel Bukowski Pawel Chrostek Filip Novokmet Marek Skawiński	Income inequality in the 21^{st} century Poland
1965	Benny Kleinman Ernest Liu Stephen J. Redding Motohiro Yogo	Neoclassical growth in an interdependent world
1964	Hanwei Huang Gianmarco I.P. Ottaviano	Rethinking revealed comparative advantage with micro and macro data
1963	Natalie Chen Dennis Novy Carlo Perroni Horng Chern Wong	Urban-biased structural change
1962	Robin Kaiji Gong Yao Amber Li Kalina Manova Stephen Teng Sun	Tickets to the global market: First US patent awards and Chinese firm exports
1961	Antonin Bergeaud Arthur Guillouzouic	Proximity of firms to scientific production

The Centre for Economic Performance Publications Unit
Tel: +44 (0)20 7955 7673 Email info@cep.lse.ac.uk
Website: http://cep.lse.ac.uk Twitter: @CEP_LSE