English Language Requirement and Educational Inequality: Evidence from 16 Million College Applicants in China

Hongbin Li* Lingsheng Meng† Kai Mu‡ Shaoda Wang§

February 15, 2024

Abstract

This paper studies the unintended effect of English language requirement on educational inequality by investigating how the staggered rollout of English listening tests in China's high-stakes National College Entrance Exam (NCEE) affected the rural-urban gap in college access. Leveraging administrative data covering the universe of NCEE participants between 1999 and 2003, we find that the introduction of English listening tests significantly lowered rural students' exam score percentile ranks relative to their urban counterparts, resulting in a 30% increase in the rural-urban gap in college access. Our back-of-the-envelope calculations suggest that, as a result of this policy change, more than 54,000 rural students lost college seats to their urban peers between 1999 and 2003, and another 11,000 rural students who elite colleges could have admitted ended up in non-elite colleges, causing them significant future income losses.

Keywords: Globalization, English Education, Human Capital, Inequality

JEL Codes: I24; I28; F69

*Stanford University, hongbinli@stanford.edu
†Stanford University, lmeng@stanford.edu
‡University of Hong Kong, kaimu@connect.hku.hk
§University of Chicago and NBER, shaoda@uchicago.edu

1 Introduction

The economic globalization in the past few decades has been accompanied by the globalization of English education, as many non-English-speaking countries have established English curricula and relevant assessment and selection criteria in their educational systems.[1] As of today, there are more than 1.5 billion English learners across the globe, accounting for more than 20% of the world population; this is four times the total number of native English speakers worldwide. While English language skills have been shown to have positive returns in the labor market,[2] rich anecdotal evidence also indicates that imposing English language requirements for higher education might exacerbate inequality in educational opportunities. Learning English as a second language, especially listening and speaking skills, typically requires rich extracurricular resources that are less available to disadvantaged students.[3] As a result, requiring English language skills in educational assessment and selection might hurt students from disadvantaged backgrounds relative to their peers and thus reduce social mobility.

While the relationship between English language requirements and educational inequality is frequently mentioned in policy discussions worldwide,[4] little rigorous evidence exists on this topic. In this paper, we formally examine this relationship by estimating how the introduction of English listening tests in China's National College Entrance Exam (NCEE) exacerbated the disadvantages in educational opportunities of rural students relative to their urban peers. Owing to the pre-existing lack of resources for English learning, rural students tend to underperform in the new listening tests. Because college admission in China is determined almost solely by the NCEE scores, underperformance in these tests would further translate into lower access to a college education.

Leveraging novel administrative data covering the universe of NCEE participants between 1999 and 2003 and exploiting the staggered rollout of the NCEE English listening

[1] For more detailed discussions of this trend in English language policy, see Cameron (2002), Kubota (2002), Shin (2007), and Hornberger and Vaish (2009).

[2] Existing work has documented the economic value of English language ability in various settings for both low-skilled and high-skilled individuals; for example, see McManus et al. (1983), Grenier (1984), McManus (1985), Kossoudji (1988), Tainer (1988), Chiswick (1991), Dustmann and Soest (2001), Berman et al. (2003), Bleakley and Chin (2004), Bleakley and Chin (2010), Azam et al. (2013).

[3] For example, students from better socioeconomic backgrounds tend to have better access to interactive tutoring, radio programs, movies, etc., which are important for the development of listening and speaking skills and can hardly be replaced by non-interactive learning materials.

[4] For example, see Ping (2010), Ye and Zhao (2011), Butler (2014) on China; Jeon (2012) on South Korea; Chinh et al. (2014) on Vietnam; Mattheoudakis and Alexiou (2009) on Greece.

tests during this period, we find that the new language requirement significantly enlarged the rural-urban gap in access to higher education. The newly introduced English listening test, which accounted for 20% of the total score of the English subject and 4% of the total score of the entire NCEE, lowered rural students' average percentile rank in the English part of the exam by two percentage points and their average percentile rank in the total NCEE score by 1.1 percentage points. As a result, the rural students' chances of college admission were reduced by roughly two percentage points, which amounts to nearly 30% of the baseline rural-urban gap in college admission.

A back-of-the-envelope calculation suggests that, due to the introduction of the NCEE English listening tests, more than 54,000 rural students lost college seats to their urban peers between 1999 and 2003. Additionally, even among those admitted to college, nearly 11,000 rural students who would have been admitted to an elite college lost their elite college seats to their urban peers due to the introduction of English listening. Linking this number to estimates of the returns to college education in China that are documented in the literature, we calculate that the rural students who lost their college seats due to the English listening tests later experienced a reduction in their starting wage of more than 40%, which is equivalent to a yearly income transfer from rural to urban students in the amount of 450 million RMB. Together, these calculations suggest that the English language requirements in the high-stakes educational selection have economically significant equity implications, which should be carefully taken into account when designing language policies.

This paper speaks to three strands of literature. First, it sheds light on the socioeconomic consequences of language policies. With the global expansion of English language education over the past few decades, a long-standing literature has investigated the labor market returns to foreign language skills for both low-skilled and high-skilled individuals (McManus et al., 1983; McManus, 1985; Grenier, 1984; Kossoudji, 1988; Tainer, 1988; Chiswick, 1991; Dustmann and Soest, 2001; Berman et al., 2003; Bleakley and Chin, 2004, 2010). Specifically, it has been shown that globalization and trade liberalization have increased the returns to English language skills in the developing world (Munshi and Rosenzweig, 2006; Levinsohn, 2007; Oster and Millett, 2010; Shastry, 2012; Azam et al., 2013). However, the socioeconomic costs of English language education are largely neglected. Our paper fills in this gap by providing the first rigorous empirical evi-

dence on how compulsory English tests could exacerbate existing rural-urban educational inequality, which highlights an unintended consequence of English language education.

Second, this paper adds to the large literature on the relationship between globalization and socioeconomic inequality. Existing work has focused mostly on the direct consequences, such as the impacts of globalization on wage inequality (Cragg and Epelbaum, 1996; Feenstra and Hanson, 1997; Harrison and Hanson, 1999; Attanasio et al., 2004; Goldberg and Pavcnik, 2004; Han et al., 2012) and unemployment (Autor et al., 2013, 2014; Acemoglu et al., 2016; Dell et al., 2019). Our paper complements this line of work by revealing a subtle channel through which globalization indirectly affects inequality: the educational policies adopted by many countries to prepare their labor force for a globalizing world, such as compulsory English language education, could differentially affect different social classes and run the risk of exacerbating existing socioeconomic inequalities.

Third, this paper contributes to the broader literature on the role of educational policies in shaping social mobility (Corak, 2013). This literature has so far considered the impacts of primary and secondary education (Restuccia and Urrutia, 2004; Bailey and Dynarski, 2011; Chuard and Schmiedgen-Grassi, 2020), test-taking for college admissions (Bulman, 2015; Goodman, 2016; Goodman et al., 2020), information (Hoxby et al., 2013), financial aid (Hoxby and Avery, 2012; Londoño-Vélez et al., 2020), government transfers to colleges (Capelle, 2019), and access to different kinds of colleges (Zimmerman, 2019; Chetty et al., 2020; Mountjoy, 2022). A new subset of this literature examines the role of design features of standardized testing (Riehl, 2019; Duquennois, 2022). Our research complements the existing works by documenting how language requirements in very high-stakes exams can substantially affect disadvantaged students' future educational opportunities.

The remainder of this paper is organized as follows. In Section 2, we briefly introduce the institutional background. In Section 3, we describe our data. We present the empirical analyses in Section 4. We evaluate the economic significance of our findings in Section 5. Section 6 concludes.

2 Background

China's NCEE is an extremely high-stakes, closed-book exam that millions of high school graduates take to compete for college admission every year. All colleges in China admit students based on their provincial rankings in the NCEE.[5] For most students, conditional on their own stated college preferences, the provincial ranking of their NCEE scores in the same cohort within the track of their choice (STEM or humanities) is the sole determinant of admission outcomes.[6] Taking the NCEE is therefore regarded by many as a life-changing opportunity to gain upward mobility—that is why students spend years preparing for the exam.

College admission in China follows a centralized system, in which students first learn about their own score, then submit a ranked list of preferred colleges, after which the colleges admit students solely based on their submitted college lists and the provincial rankings of their NCEE scores.[7] Due to the highly competitive nature of this matching market, even a marginal improvement in the provincial NCEE score ranking would typically allow a student to include better colleges in the ranked list and have more desirable admission outcomes. Even within the same college, popular majors such as economics, finance, and computer science are typically available only to students with higher provincial rankings. Therefore, students at any part of the score distribution have strong incentives to increase their NCEE scores, even by just a small margin.

In the NCEE, all students are tested on Chinese, Math, and English, with each subject accounting for 150 points. In addition, students on the STEM track are tested on physics, chemistry, and biology, while those on the humanities track are tested on history, politics, and geography. The track-specific exam contents account for another 300 points.[8] The grading and admission processes are implemented independently by each province; therefore, the NCEE scores are inter-personally comparable only within the same cohort-track-province cluster.

[5]Students take the NCEE in their home province as determined by *Hukou*, i.e., household registration determining permanent residence as well as urban or rural residency status.

[6]Rare exceptions include winners of national/international Olympiad contests, students who win sports scholarships, students with exceptional artistic talents, etc.

[7]Each college distributes its admission quota annually at the provincial-track level. The students' *Hukou* status (rural vs. urban) is not taken into account.

[8]Some provinces had different total scores in certain years, which does not affect our analysis because we use a student's provincial percentile rank (rather than the score itself) as the main outcome variable.

When English was first included in the NCEE in 1978, the test material was limited to reading comprehension and essay writing. In 1999, citing the importance of English communication skills for China's integration into the global economy, the Ministry of Education (MOE) mandated that English listening be incorporated into the NCEE nationwide by 2003. In response to this requirement, provinces across China started to introduce a listening section in the NCEE English test, which was worth 30 points (20% of the total score for English and 4% of the total score for the entire NCEE). The English listening test is conducted in the first 20 minutes of the two-hour English exam. Several English conversations are played through speakers in each exam room. At the end of each conversation, several questions regarding the conversation are played, and students need to record the answers on their exam papers.

The introduction of the NCEE English listening test has been riddled with controversy since its very beginning. Immediately after the MOE's announcement in 1999, heated debates erupted in the popular media, worrying that such a policy would harm students from disadvantaged socioeconomic backgrounds because extracurricular resources would be needed to develop English listening skills. Specifically, it has been pointed out that students from affluent socioeconomic backgrounds usually have more opportunities to engage with native English speakers, enroll in interactive English teaching programs, and gain exposure to original English radio programs and movies. Such resources are believed to be pivotal in acquiring English listening skills and can hardly be substituted by non-interactive learning materials.[9]

Due to the mounting concerns that the NCEE English listening test could exacerbate the rural-urban divide in college access, in 2005, after the English listening test was rolled out nationwide, the MOE issued a follow-up policy allowing each province to decide for itself whether to keep or abolish the test in future NCEEs. Over the following decade, more than half of the provinces eventually removed the English listening test from the NCEE, often citing "fairness for rural students" as a key motivation behind such decisions.[10]

[9]See, for example, `https://gaokao.eol.cn/yy_2876/20120608/t20120608_788386.shtml`. Appendix Section B discusses in more detail the disadvantages rural students face in English listening learning.

[10]See, for example, `https://gaokao.chsi.com.cn/gkxx/ss/201309/20130918/512733390.html`.

3 Data

3.1 NCEE English Listening Rollout

We collected information on the year each province first introduced English listening into the NCEE from the *China Education and Examination Yearbooks*, which we cross-validated with information from various news archives. Appendix Figure A.1 illustrates the staggered rollout of the NCEE English listening test between 1999 and 2003.

3.2 NCEE Administrative Data

The main data used in this paper come from a novel administrative dataset maintained by the MOE and covering more than 22 million NCEE test takers between 1999 and 2003. For each exam taker, we have detailed information on basic demographics, exam performance in each subject, and college admission outcomes. Importantly, the data allow us to categorize each student as "urban" or "rural" based on the *Hukou* status. Compared to rural residency, urban residency is often associated with substantially better socioeconomic conditions and superior public and private educational inputs.

Table 1 shows the descriptive statistics for urban and rural NCEE takers. We restrict our sample to NCEE participants who chose English as their foreign language.[11] Among the 22,608,392 NCEE participants between 1999 and 2003, 48.5% had urban residency. Roughly 54% of urban test takers and 63% of rural test takers were male. For both urban and rural test takers, the average age was roughly 19 years old; 94% were ethnic Han, and around 0.5% were Communist Party members. The sample contains individuals who took the NCEE in multiple years (henceforth "repeaters"). Roughly 21% of the urban test takers and 29% of the rural test takers were repeaters.[12]

Panels B and C present descriptive statistics for score percentile ranks and college

[11] A small share of students chose languages other than English as their foreign languages of study, such as Russian, German, Spanish, or Japanese, and were assigned to different tracks for college admission.

[12] In our main empirical analyses, we restrict the sample to Han Chinese first-time exam takers. Ethnic minority students are often entitled to various bonus points in NCEE admissions. Some minority students also had the option to choose other foreign languages instead of English in the NCEE, which might introduce extra noise in our analysis. Summary statistics for this restricted sample are presented in Appendix Table A.2. As demonstrated in Appendix Figure A.4, our baseline findings are quantitatively similar if we include minority students or repeaters. Furthermore, to the extent that the estimated effects are larger when repeaters are included, we argue that this could result from the asymmetric selection of relatively high-quality urban students into repeaters.

admission outcomes, respectively. Urban students outperformed rural students in English and Chinese, while rural students were better at Math. Overall, there is a salient rural-urban gap in access to a college education. Urban students had consistently higher admission rates than their rural counterparts across the board, from admission to any college (including both 3-year community colleges and 4-year regular colleges) to admission to top schools such as "Project 211" and "Project 985" colleges.[13]

4 Empirical Analyses

In this section, we first discuss the baseline empirical results on how the introduction of English listening tests affected rural-urban gaps in NCEE scores and college admission, followed by the associated event study analyses. We then show the heterogeneous impacts on admissions to colleges of various tiers, followed by mechanism tests and robustness checks.

4.1 Baseline Effects on Rural-Urban Gaps in Exam Performance

Our baseline identification strategy exploits the staggered rollout of the NCEE English listening test across different provinces and investigates its differential impacts on rural vs. urban students within the same province-cohort-track cluster. Specifically, we estimate the following triple-difference (DDD) model:

$$y_{irpst} = \beta \cdot listening_{pt} \cdot rural_i + \delta_{pst} + \lambda_{rt} + \varepsilon_{iprst} \tag{1}$$

where y_{irpst} is the outcome of interest (score percentile rank or admission outcome) for student i, of *Hukou* type r, in track s, in province p, and in year t.

On the right-hand side, $listening_{pt}$ is an indicator that NCEE English listening was included in province p in year t, and $rural_i$ is a dummy variable that equals one if student i has rural *Hukou*, and zero otherwise. The interaction term of these two variables thus identifies the differential impacts of the English listening test on rural vs. urban students. We control for province-track-year fixed effects (δ_{pst}) to account for any province-track-

[13]Project 211 was a project of National Key Universities launched in 1995 by China's MOE. "211" colleges roughly translate into the top 100 universities in China, while "985" colleges are the best 39 universities among the "211" colleges.

specific shock common to rural and urban students, such as changes in admission quotas, and rural-year fixed effects (λ_{rt}) to account for any national trend of urban-rural gaps in NCEE performance. Standard errors are clustered at the province-track level.

The baseline results presented in Table 2 (Columns 1 and 2) suggest that introducing English listening tests into the NCEE significantly lowered rural students' English score percentile ranks by more than two percentage points or a more than 60% increase of the baseline rural-urban gap in NCEE English exam performance. In contrast, in Appendix Table A.3, we conduct a placebo test by investigating the impacts of introducing English listening tests on Chinese and Math scores and find precisely estimated null effects. As shown in Columns 3 and 4 of Table 2, we estimate that introducing NCEE English listening tests enlarged the rural-urban gap in overall NCEE performance by a magnitude of more than one percentage point.

The effects on NCEE scores further translated into a widened rural-urban gap in college admission. As shown in Columns 5 and 6 of Table 2, when English listening tests were introduced into the NCEE, rural students' chances of college admission dropped by two percentage points relative to their urban peers. Given that the baseline urban-rural gap in college admission in our sample is 7.2 percentage points (see Panel C of Table 1), our estimates suggest that the introduction of English listening in the NCEE widened the urban-rural college admission gap by roughly 30 percent.[14]

4.2 Event Study Estimates for Rural vs. Urban Students

To examine the validity of our research design and understand the dynamic impacts of English listening tests, in addition to the baseline DDD specification, we also estimate event study models separately for the subsamples of rural and urban students:

$$y_{ipst} = \sum_{k \neq -1} D_{pt}^{k} \cdot \beta_k + \delta_p + \theta_{st} + \varepsilon_{ipst} \tag{2}$$

[14] In Appendix Table A.4, we zoom in on a subset of more selective colleges, such as four-year general-purpose universities, Project 211 universities, and Project 985 universities. Across the board, we see an enlarged rural-urban gap in access to these elite colleges. However, since only the top-ranked students would potentially apply to the elite colleges, the population average treatment effects become increasingly underpowered as we focus on more selective schools.

where y_{ipst} is the score percentile rank for student i who took the NCEE in province p, year t choosing track s. D_{pt}^{k} are event dummies indicating the kth year until/since province p's first adoption of English listening relative to year t. We choose the year before each province's first adoption of English listening (i.e., event time -1) as the reference period. The province fixed effects δ_p account for any province-specific, time-invariant determinants of NCEE performance. The track-year fixed effects θ_{st} control for time-varying track-specific factors common to all provinces. To avoid putting negative weights on the average treatment effect of certain groups in conventional two-way fixed effects models, we follow the recent econometrics literature and adjust the conventional event study approach with an "interaction-weighted" estimator (Sun and Abraham, 2021).

As we can see in Figure 1, for all three main outcome variables, the rural and urban students followed almost identical trends before the introduction of English listening tests, lending support to the validity of our baseline triple-difference approach. In stark contrast, after the introduction of the NCEE English listening tests, we see an immediate divergence in trends between rural and urban students, with the urban students significantly outperforming their rural counterparts in English scores, aggregate scores, and college admission rate.[15] These patterns suggest a causal interpretation of the relationship between English listening tests and widening rural-urban gaps in exam performance and college access.[16]

4.3 Heterogeneous Impacts on Admission to Selective Colleges

According to the existing literature, not only does college education *per se* have a large return in China's labor market (Li et al., 2012a), there is also a particularly steep return to attending more selective colleges (Jia and Li, 2021). To paint a complete picture of the consequences of the NCEE English listening tests, we need to consider their impacts on admissions to elite colleges.

Estimating the population average treatment effects using the entire sample would

[15] For college admission, we observe a negative common shock to both rural and urban students in the last period. While it does not affect our interpretation that focuses on the rural-urban gap, we hypothesize that this common shock comes from a later NCEE reform ("3+X"), which allowed students to select some of their testing subjects. Qualitatively, in anticipation of this pending change in exam format, many repeating NCEE takers decided to accept their college offers instead of retaking the exam again, which reduced the college seats for first-time rural and urban NCEE takers included in our sample.

[16] The absence of pre-trends and immediate changes in English scores after introducing listening tests suggests minimal anticipatory adjustments in test preparation.

mask important heterogeneities and suffer from reduced statistical power when attempts are made to detect the impacts of NCEE English listening tests on elite college admissions since student ability has a wide distribution and only those above a certain ability/score threshold would be eligible for admissions to more selective colleges.

To uncover the underlying heterogeneities related to student abilities, we split our sample into quartiles based on the sum of Chinese and Math scores and separately estimate the baseline specification for each quartile.[17]

As shown in Figure 2, English listening tests' effect on college admissions varies with academic ability. For below-median students, the English listening test mainly affected the rural-urban gap in admissions to any college. For students in the second and third quartiles, the English listening test affected the rural-urban gap in admissions to four-year regular colleges rather than three-year community colleges. For students in the top quartile, the English listening test impacted the rural-urban gap in admissions to the most elite schools, namely those "Project 211/985" colleges.

4.4 Private vs. Public Input in English Learning?

The rural-urban gap in English learning resources could come from either public or private input: urban-Hukou students tend to attend more resource-rich high schools while also being more likely to afford private English tutoring and learning materials.

To evaluate the relative importance of these two possible channels, we estimate a more saturated econometric model that controls for high school fixed effects, so we can compare the gaps between rural and urban students attending the same school before and after the introduction of the English listening tests.[18] As shown in Appendix Table A.5, controlling for high school fixed effects eliminates any impact of English listening on urban-rural gaps in NCEE scores or college admission outcomes. These results indicate that once rural students are given the same public English learning resources as their urban peers, they are no longer disproportionately hurt by the introduction of English listening tests. This suggests that a disparity in public rather than private input is the

[17] We proxy student ability with Chinese and Math scores because they are not directly affected by the introduction of English listening and because Chinese and Math are universally tested in all years, for all tracks, and across all provinces. The results remain qualitatively similar if we instead use the Chinese score alone, the Math score alone, or the total non-English (including track-specific subjects) score as stratifying variables.

[18] There are on average 15 high schools in each county.

main driving force behind the rural-urban gap in English listening test performance.

4.5 Can Longer Preparation Time Mitigate the Rural-Urban Gap in English Listening Performance?

Since many provinces announced that English listening tests would be included in the NCEE only a few months before the exam, one might think that such short notice (and thus lack of ample preparation) could have contributed to the enlarged rural-urban gap in NCEE performance. If that is the case, an English listening requirement might not increase social inequality as long as the disadvantaged students have enough time to prepare for it.

We exploit another institutional feature of the reform to evaluate the relevance of this hypothesis. During our sample period, 10 provinces conducted one or two years of a pilot English listening test before its formal introduction in the NCEE.[19] In the pilot tests, English listening questions were tested, but their scores were not counted into either the English subject score or the aggregate NCEE score.[20]

If students from provinces where pilot listening tests were conducted in previous years were better prepared for the NCEE English listening exam, we might expect the urban-rural gaps in NCEE scores and college admissions to be less affected by the formal introduction of English listening tests. We test for this preparation effect by comparing treatment effects of the formal introduction of English listening tests in "prepared" relative to "unprepared" provinces. As shown in Appendix Table A.6, we find no significant preparation effects.[21] We also control for these pilot listening exams as placebo treatments to further probe the validity of our baseline triple difference approach. Reassuringly, as shown in Appendix Table A.7, pilot listening exams do not affect NCEE scores or college admission outcomes. In contrast, the estimated effects of actual English listening exams remain largely unchanged.

[19] See Appendix Table A.1 for detailed information on the variation in pilot listening exams.

[20] In an English exam with a pilot listening test, non-listening scores were re-scaled to 150 points.

[21] The results of this test lend further credence to the notion of minimal prior test preparation. The presence of a pilot exam serves as a more explicit signal of upcoming formal assessments, theoretically increasing the incentive for early preparation. However, our empirical results show no statistically significant effects of pilot listening tests on preparation activities. This implies that any overall anticipation effect, presumably less pronounced than that observed in provinces with pilot tests, is likely to be minimal and thus unlikely to substantially bias our primary findings. We discuss this test in greater detail in Appendix Section C.

Our examination of the pilot listening exam lends further credence to the notion of minimal advance test preparation. The presence of a pilot test serves as a more explicit signal of upcoming formal assessments, theoretically heightening the incentive for early preparation. Yet, our empirical findings reveal no statistically significant effects of pilot listening exams on preparation activities. This implies that any overall anticipation effect, presumably less pronounced than that observed in provinces with pilot tests, is likely minimal and thus unlikely to substantially skew our primary findings.

4.6 Selective Participation in the NCEE

A potential alternative interpretation of our baseline DDD results is that the introduction of English listening in the NCEE led to differential participation in NCEE between rural and urban students. It could be that the introduction of English listening discouraged more rural students with relatively low academic performance from participating in the NCEE, expecting that the policy change would worsen their exam performance and college admission prospects.

Such an interpretation is unlikely to have driven our findings substantially. The bottom two panels of Figure 2 show the differential impact of the listening test across student ability levels, with students in the higher quartiles significantly more affected in terms of admission to 985 and 211 colleges, respectively. In contrast, students in the lowest quartile were barely affected. Furthermore, we formally probe this alternative interpretation by investigating whether the proportions of urban and rural "potential cohorts" who eventually participated in the NCEE are systematically correlated with the introduction of English listening.[22] As shown in Appendix Table A.8, there is no systematic correlation between the introduction of English listening exams and differential urban-rural NCEE participation rates. See more details of this test in Appendix Section D.

[22]For each province-year-*Hukou* cluster, we construct the "potential cohort" by weighting the population of relevant birth cohorts from the National Population Census 1990 by the share of birth cohorts observed in our NCEE dataset.

5 Economic Significance

We have found that the introduction of English listening in the NCEE benefited urban students at the cost of their rural peers in terms of college admission. In this section, we perform a series of simple back-of-the-envelope calculations to shed light on our reduced-form findings' policy and welfare implications. Specifically, we leverage our baseline estimates to answer the following questions: (1) how many rural students lost their college admission seats to their urban peers due to the introduction of English listening tests? And (2) what is the magnitude of the implicit transfer of future income from rural to urban students due to the reform?

5.1 Rural Admission Loss Due to English Listening Test

Our baseline empirical results from Section 4.1 suggest that, on average, English listening tests reduce the chance of rural students' college admission by two percentage points relative to their urban peers. Applying this estimate to the rural test takers in our sample, we calculate that, between 1999 and 2003, more than 54,000 rural students lost their college seats to their urban peers due to the introduction of English listening tests.[23] Extrapolating beyond our sample period, our calculations suggest that for every year that the NCEE listening is held nationwide, more than 20,000 rural students would lose college seats to their urban peers.[24]

We also calculate the number of elite college seats lost by rural students to their urban peers. Since English listening affects the rural-urban gap in elite college admission mainly among the top-performing students (see Figure 2), we focus on students in the top quartile, for which our estimated treatment effect is 1.5 percentage points. We then calculate that nearly 11,000 rural students lost elite college seats to their urban peers due to English listening between 1999 and 2003.[25] Extrapolating beyond our sample period,

[23] The proportion of rural NCEE takers in our sample is close to 50%, which means that our estimated treatment effect of a two percentage point increase in the urban-rural admission rate gap corresponds to a one percentage point drop in rural students' admission rate. We multiply the estimated effect on the rural admission by the number of rural NCEE takers in province-years in which English listening was tested: $1\% \cdot \sum_{pt} RuralN_{pt} \times Listening_{pt} = 54134.72$.

[24] There were on average 2,328,325 rural NCEE takers each year between 1999 and 2003, which is a lower bound for the post-2003 period. Therefore, the conservative estimate for the average number of rural students losing college seats each year is $1\% \times 2328325 = 23283.25$.

[25] Following similar steps, $1.5\% \times 0.55 \times \sum_{pt} RuralN_{pt} \times Listening_{pt} = 10773.881$.

our calculation suggests that for every year that the NCEE listening is held nationwide, more than 3,000 elite college seats would be reallocated from rural students to their urban peers.[26]

5.2 Economic Inequality Caused by English Listening Test

In China's labor market, the return to college education is huge. Li et al. (2012a) estimate a 40% income return to college education, while Jia and Li (2021) estimate an additional 40% return to elite colleges. Combining these estimates with our calculations in Section 5.1, we can measure the labor market implications of switching college seats from rural students to urban students due to the introduction of the NCEE English listening tests.

According to the 2010 Chinese College Students Survey (CCSS) (Li et al., 2012b), the average monthly starting wage for students graduating from non-elite colleges is 1,900 RMB.[27] Linking this number to the estimates of Li et al. (2012a), the counterfactual wage for non-college graduates is $1900/140\% = 1357$; linking this number to the estimates of Jia and Li (2021), the counterfactual wage for elite college graduates is $1900 \times 140\% = 2660$. Thus, we can calculate that rural students lost more than 450 million RMB of future *annual* income to their urban peers between 1999 and 2003 due to the introduction of English listening tests in the NCEE.[28]

Our calculation of the future income loss is conservative, as it only captures the "infra-marginal" treatment effects, i.e., college vs. no college and elite college vs. non-elite college. Because the introduction of English listening tests worsens the relative performance of rural students across the board (Appendix Figures A.2 and A.3), rural students far away from the (elite) college cutoffs are also more likely to end up in marginally worse colleges, or less popular majors within the same college. Accounting for these marginal treatment effects would lead to even larger estimates of future income losses.

[26] There are on average 369,269 rural NCEE takers in the top performance quartile each year, which is a lower bound for the post-2003 period. Therefore, the conservative estimate for the number of rural students losing elite college seats each year is $1.5\% \times 0.55 \times 369269 = 3046.47$.

[27] Wage in CCSS is defined as the highest offer received by a student at the time of the survey, which is upon graduation.

[28] $[54135 \times (1900 - 1357) + 10774 \times (2660 - 1900)] \times 12 = (29,395,305 + 8,188,240) \times 12 = 451,002,540$.

6 Conclusion

This paper provides the first piece of rigorous empirical evidence on how English language requirements in education contribute to educational inequality in a non-English-speaking country. Leveraging novel administrative data and exploiting the staggered introduction of English listening tests in China's National College Entrance Exam between 1999 and 2003, we find that the policy significantly lowered rural students' exam scores relative to their urban peers, thereby widening the rural-urban gap in college access by 30%.

A simple back-of-the-envelope calculation suggests that, due to the introduction of English listening tests in the NCEE during this period, more than 54,000 rural students lost access to college education altogether, and nearly 11,000 rural students whom elite colleges could have admitted ended up in non-elite-colleges. Together, the loss of better educational opportunities corresponds to a transfer of future income from rural to urban students in the amount of 450 million RMB per year.

Our findings suggest nuance in modifying high-stakes exams that have significant implications for students' educational paths and career prospects. The addition of new exam components, while aimed at encouraging the development of valuable skills, may inadvertently widen the gap between urban and rural students if implemented too rapidly. The alternative of gradual integration of new exam material, with a slow increase in its significance, could avoid this risk. If the government raises support for rural education by allocating more resources to rural schools, this could enable more rural students to acquire new, valuable English language skills that are beneficial in the broader labor market. This would enable rural students to adapt to and excel in acquiring new, valuable English language skills that are beneficial in the broader labor market.

References

Acemoglu, D., Autor, D., Dorn, D., Hanson, G.H., Price, B., 2016. Import competition and the great US employment sag of the 2000s. Journal of Labor Economics 34, S141–S198.

Arrow, K., Bowles, S., Durlauf, S.N., 2018. Meritocracy and economic inequality. Princeton University Press.

Attanasio, O., Goldberg, P.K., Pavcnik, N., 2004. Trade reforms and wage inequality in Colombia. Journal of Development Economics 74, 331–366.

Autor, D.H., Dorn, D., Hanson, G.H., 2013. The China syndrome: Local labor market effects of import competition in the United States. American Economic Review 103, 2121–68.

Autor, D.H., Dorn, D., Hanson, G.H., Song, J., 2014. Trade adjustment: Worker-level evidence. The Quarterly Journal of Economics 129, 1799–1860.

Azam, M., Chin, A., Prakash, N., 2013. The returns to English-language skills in India. Economic Development and Cultural Change 61, 335–367.

Bailey, M.J., Dynarski, S.M., 2011. Gains and gaps: Changing inequality in US college entry and completion. Technical Report. National Bureau of Economic Research.

Berman, E., Lang, K., Siniver, E., 2003. Language-skill complementarity: returns to immigrant language acquisition. Labour Economics 10, 265–290.

Bleakley, H., Chin, A., 2004. Language skills and earnings: Evidence from childhood immigrants. Review of Economics and Statistics 86, 481–496.

Bleakley, H., Chin, A., 2010. Age at arrival, English proficiency, and social assimilation among US immigrants. American Economic Journal: Applied Economics 2, 165–92.

Bulman, G., 2015. The effect of access to college assessments on enrollment and attainment. American Economic Journal: Applied Economics 7, 1–36.

Butler, Y.G., 2014. Parental factors and early English education as a foreign language: A case study in Mainland China. Research papers in education 29, 410–437.

Cameron, D., 2002. Globalization and the teaching of 'communication skills'. Globalization and language teaching 67, 82.

Capelle, D., 2019. The great Gatsby goes to college: Tuition, inequality and intergenerational mobility in the us.

Chetty, R., Friedman, J.N., Saez, E., Turner, N., Yagan, D., 2020. Income segregation and intergenerational mobility across colleges in the United States. The Quarterly Journal of Economics 135, 1567–1633.

Chinh, N.D., Quynh, T.H., Ha, N.T., et al., 2014. Inequality of access to English language learning in primary education in Vietnam: A case study, in: Equality in education. Brill Sense, pp. 139–153.

Chiswick, B.R., 1991. Speaking, reading, and earnings among low-skilled immigrants. Journal of Labor Economics 9, 149–170.

Chuard, P., Schmiedgen-Grassi, V., 2020. Switzer-land of opportunity: Intergenerational income mobility in the land of vocational education. Available at SSRN 3662560 .

Corak, M., 2013. Income inequality, equality of opportunity, and intergenerational mobility. Journal of Economic Perspectives 27, 79–102.

Cragg, M.I., Epelbaum, M., 1996. Why has wage dispersion grown in Mexico? is it the incidence of reforms or the growing demand for skills? Journal of Development Economics 51, 99–116.

Dell, M., Feigenberg, B., Teshima, K., 2019. The violent consequences of trade-induced worker displacement in Mexico. American Economic Review: Insights 1, 43–58.

Duquennois, C., 2022. Fictional money, real costs: Impacts of financial salience on disadvantaged students. American Economic Review 112, 798–826.

Dustmann, C., Soest, A.v., 2001. Language fluency and earnings: Estimation with misclassified language indicators. Review of Economics and Statistics 83, 663–674.

Feenstra, R.C., Hanson, G.H., 1997. Foreign direct investment and relative wages: Evidence from Mexico's maquiladoras. Journal of international economics 42, 371–393.

Goldberg, P.K., Pavcnik, N., 2004. Trade, inequality, and poverty: What do we know? Evidence from recent trade liberalization episodes in developing countries. Technical Report. National Bureau of Economic Research.

Goldin, C.D., 2001. Meritocracy and economic inequality (review).

Goodman, J., Gurantz, O., Smith, J., 2020. Take two! SAT retaking and college enrollment gaps. American Economic Journal: Economic Policy 12, 115–58.

Goodman, S., 2016. Learning from the test: Raising selective college enrollment by providing information. Review of Economics and Statistics 98, 671–684.

Grenier, G., 1984. The effects of language characteristics on the wages of Hispanic-American males. Journal of Human Resources , 35–52.

Han, J., Liu, R., Zhang, J., 2012. Globalization and wage inequality: Evidence from urban China. Journal of International Economics 87, 288–297.

Harrison, A., Hanson, G., 1999. Who gains from trade reform? some remaining puzzles. Journal of Development Economics 59, 125–154.

Hornberger, N., Vaish, V., 2009. Multilingual language policy and school linguistic practice: globalization and english-language teaching in India, Singapore and South Africa. Compare 39, 305–320.

Hoxby, C., Turner, S., et al., 2013. Expanding college opportunities for high-achieving, low-income students. Stanford Institute for Economic Policy Research Discussion Paper 12, 7.

Hoxby, C.M., Avery, C., 2012. The missing" one-offs": The hidden supply of high-achieving, low-income students. Technical Report. National Bureau of Economic Research.

Hu, J., 2007. Developing English education in rural environment. Hunan Normal University Journal of Educational Science 6, 101–103.

Hu, R., 2002. English listening instruction in rural schools. Vocational Education Forum , 42–42.

Jeon, M., 2012. English immersion and educational inequality in South Korea. Journal of Multilingual and Multicultural Development 33, 395–408.

Jia, R., Li, H., 2021. Just above the exam cutoff score: Elite college admission and wages in China. Journal of Public Economics 196, 104371.

Kossoudji, S.A., 1988. English language ability and the labor market opportunities of Hispanic and East Asian immigrant men. Journal of Labor Economics 6, 205–228.

Kubota, R., 2002. The impact of globalization on language teaching in Japan, in: Globalization and language teaching. Routledge, pp. 23–38.

Levinsohn, J., 2007. Globalization and the returns to speaking English in South Africa, in: Globalization and poverty. University of Chicago Press, pp. 629–646.

Li, H., Liu, P.W., Zhang, J., 2012a. Estimating returns to education using twins in urban China. Journal of Development Economics 97, 494–504.

Li, H., Meng, L., Shi, X., Wu, B., 2012b. Does attending elite colleges pay in China? Journal of Comparative Economics 40, 78–88.

Liu, J., 2008. English Listening Instruction in Rural Junior High Schools. Ph.D. thesis.

Liu, J., 2009. Barriers to English listening for rural high school students. Talents .

Londoño-Vélez, J., Rodríguez, C., Sánchez, F., 2020. Upstream and downstream impacts of college merit-based financial aid for low-income students: Ser Pilo Paga in Colombia. American Economic Journal: Economic Policy 12, 193–227.

Mattheoudakis, M., Alexiou, T., 2009. 10. early foreign language instruction in Greece: Socioeconomic factors and their effect on young learners' language development, in: The age factor and early language learning. De Gruyter Mouton, pp. 227–252.

McManus, W., Gould, W., Welch, F., 1983. Earnings of Hispanic men: The role of English language proficiency. Journal of Labor Economics 1, 101–130.

McManus, W.S., 1985. Labor market costs of language disparity: An interpretation of Hispanic earnings differences. The American Economic Review 75, 818–827.

Mountjoy, J., 2022. Community colleges and upward mobility. American Economic Review 112, 2580–2630.

Munshi, K., Rosenzweig, M., 2006. Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy. American Economic Review 96, 1225–1252.

Oster, E., Millett, M.B., 2010. Do call centers promote school enrollment? Evidence from India. Technical Report. National Bureau of Economic Research.

Ping, K., 2010. Necessity and feasibility of English listening tests in NCEE. Theory and Practice of Education 5, 17–18.

Restuccia, D., Urrutia, C., 2004. Intergenerational persistence of earnings: The role of early and college education. American Economic Review 94, 1354–1378.

Riehl, E., 2019. Fairness in college admission exams: From test score gaps to earnings equality .

Shastry, G.K., 2012. Human capital response to globalization education and information technology in India. Journal of Human Resources 47, 287–330.

Shin, H., 2007. English language teaching in Korea, in: International handbook of English language teaching. Springer, pp. 75–86.

Song, X., 2017. Causes for weak English listening ability among rural high school students. Course Education Research 26.

Sun, L., Abraham, S., 2021. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of Econometrics 225, 175–199.

Tainer, E., 1988. English language proficiency and the determination of earnings among foreign-born men. Journal of Human Resources , 108–122.

Tan, K.P., 2008. Meritocracy and elitism in a global city: Ideological shifts in Singapore. International Political Science Review 29, 7–27.

Wang, Y., 2000. Obstacles to English listening instruction in rural high schools. Teaching and Management 8.

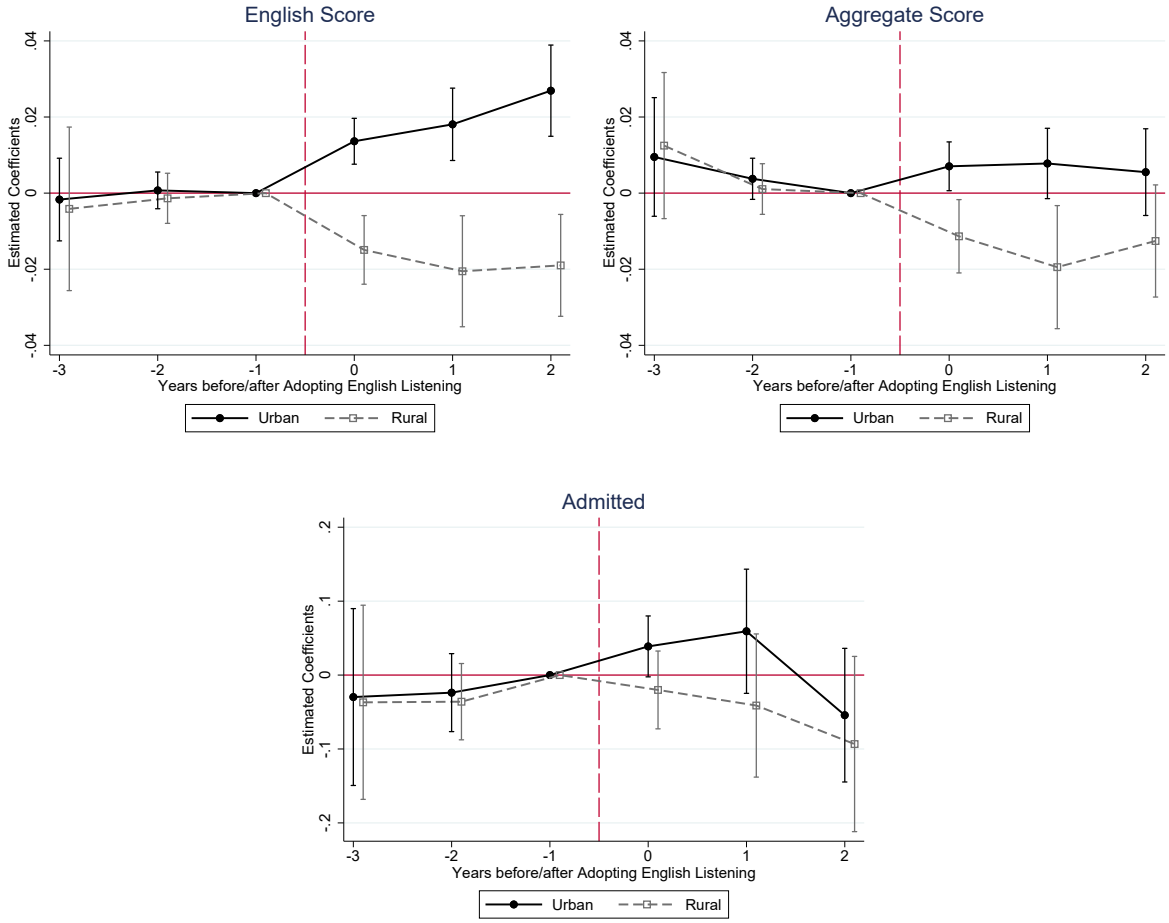
Ye, F., Zhao, P., 2011. Analysis on difficulties of English listening instruction in rural secondary education. Education Innovation , 20–21.

Zhang, L., 2018. Weak links in rural English education. Shandong Agricultural Engineering Institute Journal 2.

Zhu, Y., 2014. Causes to and countermeasure for low performance of English listening among rural high school students. Teaching Review 5.

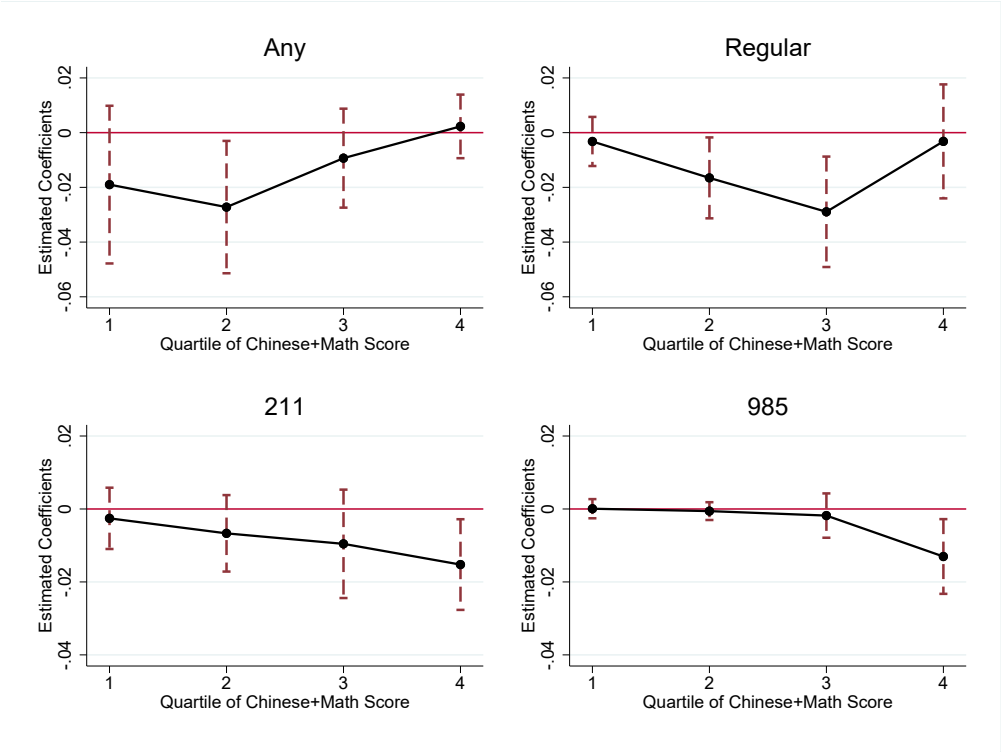
Zimmerman, S.D., 2019. Elite colleges and upward mobility to top jobs and top incomes. American Economic Review 109, 1–47.

Figure 1: Event Study on Urban and Rural Students' NCEE Scores and Admission



Notes: This figure plots point estimates, and their 95% confidence intervals from augmenting Equation 2 with the "interaction-weighted" estimator proposed in Sun and Abraham (2021), for urban and rural students separately. Since there is no "never-treated" group in our sample, the "control group" for the Sun-Abraham method is the "last-treated" group. As a result, the relative timing to treatment (i.e. the horizontal axis) only ranges from -3 to 2. In the upper row, outcome variables are the percentile ranks of English and aggregate NCEE scores, calculated within each province-year-track cluster. In the lower row, the outcome variable is a dummy for admission into any college. The regressions control for county fixed effects. Standard errors are clustered at the province-track level.

Figure 2: Heterogeneity of Treatment Effects Stratified with Student Performance



Notes: This figure shows how the effects of English listening on urban-rural gaps of NCEE outcomes vary by student performance, which we proxy with the sum of Chinese and Math scores. We split our sample into four quartiles of performance (1st quartile being bottom performers and 4th quartile being top performers). For each quartile, we separately estimate our baseline DDD model in Equation 1 (county fixed effects included). The outcome variables are indicators for admission into different types of colleges. The "regular colleges" refer to colleges offering 4-year degree programs, in contrast to 3-year community colleges. The "211 colleges" refer to colleges included in the 211 Program, which are broadly regarded as the top 100 colleges in China. The "985 colleges" are a more select group within the 211 colleges, participating in the 985 Program and representing the top 39 colleges in the country. Dashed lines around point estimates are 95% confidence intervals. Standard errors are clustered at the province-track level.

Table 1: Summary Statistics

	Urban		Rural	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
	A. Demographic Information			
Male	.537	.499	.626	.484
Age	18.886	1.073	19.246	1.185
Han	.937	.243	.940	.237
CCP Member	.005	.069	.005	.069
Repeater	.212	.409	.288	.453
	B. NCEE Score Percentile Ranks			
Total Score	.501	.299	.499	.278
Chinese Score	.518	.292	.482	.283
Math Score	.490	.297	.510	.280
English Score	.516	.298	.484	.277
	C. College Admission Dummies			
Any College	.651	.477	.579	.494
Regular Colleges	.330	.470	.261	.439
Project 211 Colleges	.114	.318	.068	.252
Project 985 Colleges	.046	.209	.024	.153
Observations	10,966,764		11,641,628	

Notes: The sample is restricted to NCEE participants who chose English as the foreign language to be tested. For each variable, the sample size with non-missing values may vary. NCEE Score percentile ranks are calculated within each province-year-track cluster. In Panel C, the "regular colleges" refer to colleges offering 4-year degree programs, in contrast to 3-year community colleges. The "211 colleges" refer to colleges included in the 211 Program, which are broadly regarded as the top 100 colleges in China. The "985 colleges" are a more select group within the 211 colleges, participating in the 985 Program and representing the top 39 colleges in the country.

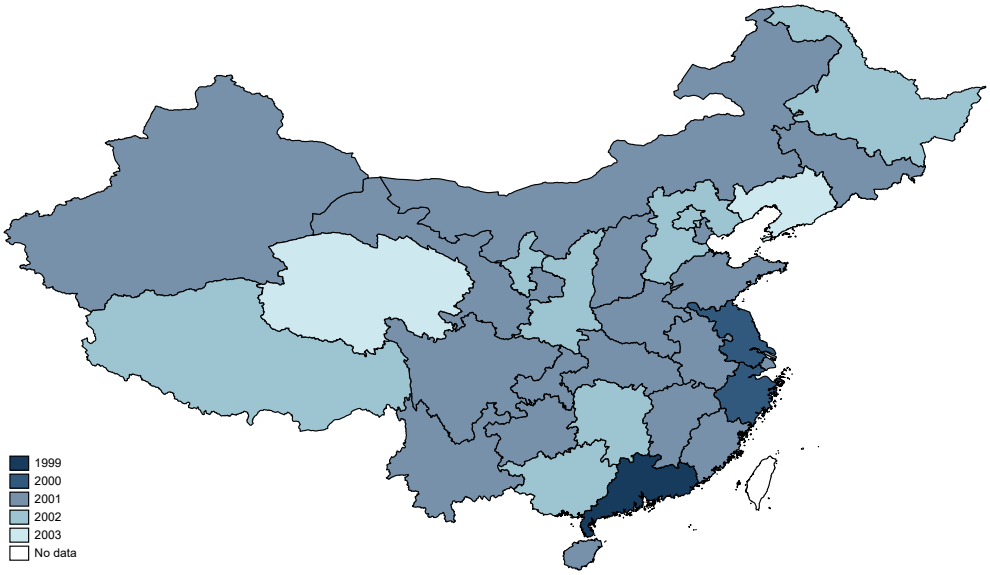
Table 2: Average Effects of English Listening on Urban-Rural Gaps in NCEE Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
	English Score		Aggregate Score		Admitted	
Listening×Rural	-.020**	-.021***	-.011*	-.012**	-.020**	-.017**
	(.008)	(.007)	(.007)	(.006)	(.008)	(.008)
Rural-Year FE	✓	✓	✓	✓	✓	✓
Province-Year-Track FE	✓	✓	✓	✓	✓	✓
County FE	✓	✓	✓	✓	✓	✓
Trimming		✓		✓		✓
Obs.	15,825,397	14,243,935	15,825,260	14,243,798	15,825,397	14,243,935
R^2	.062	.060	.060	.050	.129	.139

Notes: This table reports regression results from estimating the baseline triple difference model in Equation 1. The sample consists of first-time Han Chinese exam takers. Outcome variables are the percentile ranks of English and aggregate NCEE scores (calculated within province-year-track clusters), as well as an indicator for admission into any college. In Columns 2, 4, and 6, we trim our sample by dropping observations with aggregate score percentile rank lower than 0.1. Standard errors in parentheses are clustered at the province-track level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

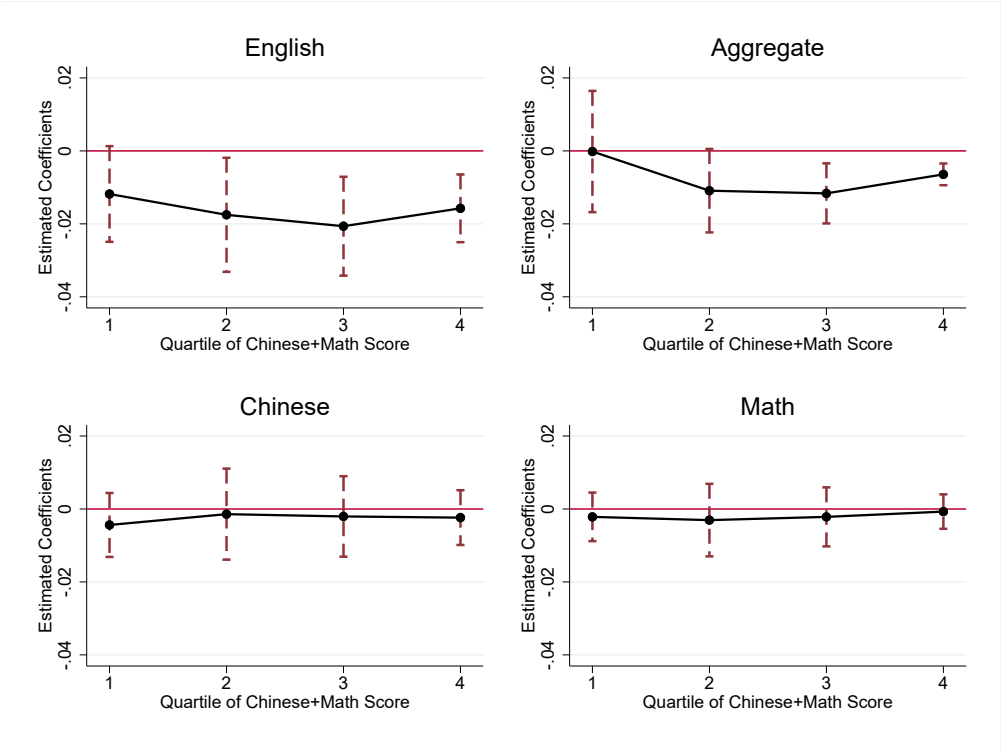
A Appendix Figures and Tables

Figure A.1: Timing of English Listening Adoption by Province



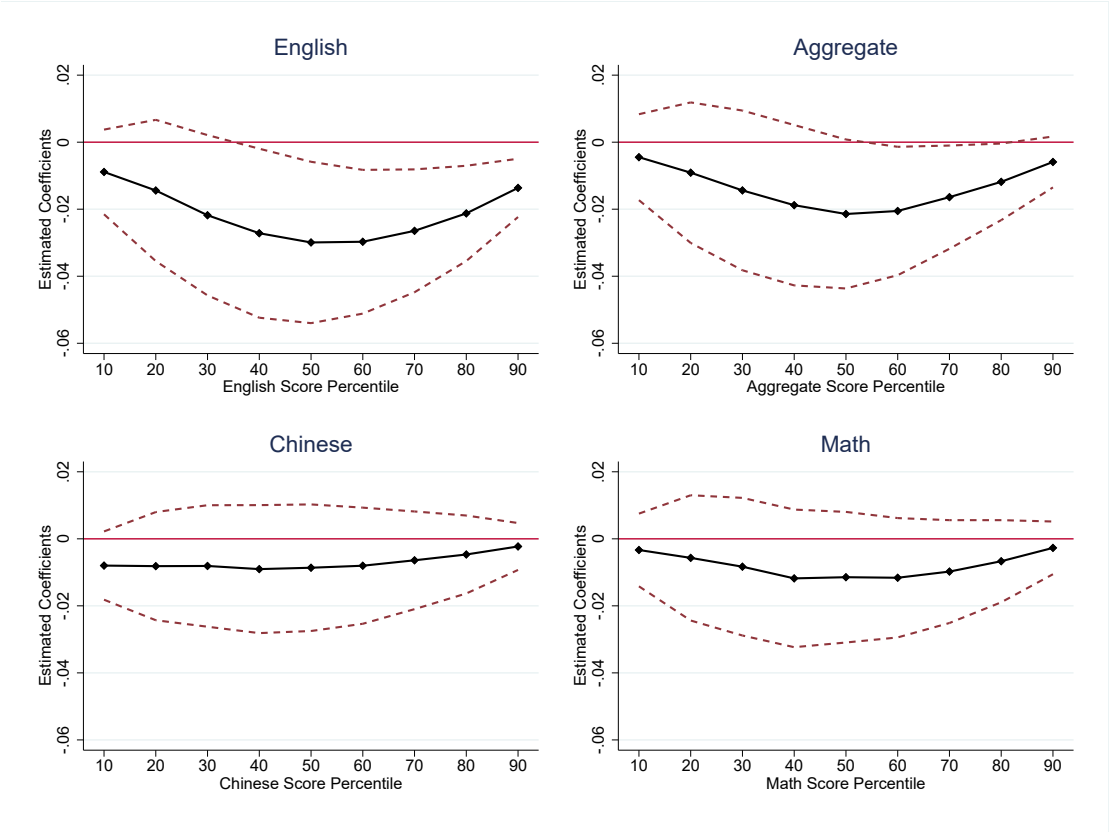
Notes: Information in this map on the timing of introduction of English listening is collected from *China Education and Examination Yearbooks*. We classify English listening to be adopted only if the scores of English listening tests were included in the aggregate score.

Figure A.2: Heterogeneity of Treatment Effects: NCEE Scores



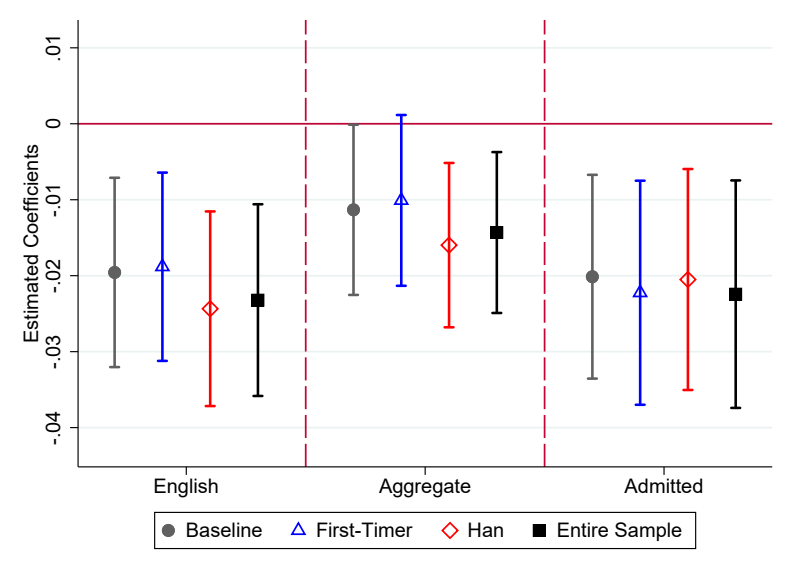
Notes: This figure shows the results from the same stratified analysis as shown in Figure 2, with the outcome variables being NCEE score percentile ranks. For students in the top 3 quartiles, the introduction of the English listening test significantly enlarged the rural-urban gaps in English and aggregate scores. Reassuringly, the introduction of NCEE listening tests has no effect on urban-rural gaps in Chinese or Math scores for any of the four performance quartiles. Dashed lines around point estimates are 95% confidence intervals. Standard errors are clustered at the province-track level.

Figure A.3: DID in CDF of Score Percentile Ranks



Notes: This figure illustrates the Difference-in-Differences in cumulative distribution function of score percentile ranks. Specifically, for each subject, we replicate the baseline DDD specification 9 times, with outcome variables being indicators of the subject score being higher than each decile. Dashed lines around point estimates are 95% confidence intervals. Standard errors are clustered at the province-track level.

Figure A.4: Robustness of Baseline Results to Alternative Samples



Notes: This figure shows the robustness of our baseline DDD estimates for Equation 1 to four alternative subsamples. For each of the three main outcome variables, the subsamples used for estimation are (from left to right, see legend) the baseline (Han first-timers), first-timers, Han, and the entire sample. Solid lines around point estimates are 90% confidence intervals. Standard errors are clustered at the province-track level.

Table A.1: Timing of the Introduction of English Listening Exam in NCEE

	1999	2000	2001	2002	2003
Guangdong	30	30	30	30	30
Jiangsu	0	30	30	30	30
Zhejiang	0	30	30	30	30
Jilin	0	Pilot	30	30	30
Jiangxi	0	Pilot	30	30	30
Inner Mongolia	0	0	30	30	30
Shanghai	0	0	30	30	30
Anhui	0	0	30	30	30
Shandong	0	0	30	30	30
Hainan	0	0	30	30	30
Yunnan	0	0	30	30	30
Tianjin	Pilot	Pilot	20	30	30
Shanxi	0	Pilot	20	30	30
Henan	0	Pilot	20	30	30
Fujian	0	0	20	30	30
Hubei	0	0	20	30	30
Chongqing	0	0	20	30	30
Sichuan	0	0	20	30	30
Guizhou	0	0	20	30	30
Gansu	0	0	20	30	30
Xinjiang	0	0	20	30	30
Beijing	0	0	Pilot	30	30
Tibet	0	0	Pilot	30	30
Shaanxi	0	0	Pilot	30	30
Ningxia	0	0	Pilot	30	30
Hebei	0	0	0	30	30
Heilongjiang	0	0	0	30	30
Hunan	0	0	0	30	30
Guangxi	0	0	0	30	30
Liaoning	0	0	Pilot	Pilot	30
Qinghai	0	0	0	0	30

Notes: This table shows detailed information on the roll-out of the English listening exam between 1999 and 2003. Pilot denotes that English listening was tested without being counted into the final English subject score. 20 and 30 denote that English listening was tested and counted into the final English subject score for 20 and 30 points, respectively.

Table A.2: Summary Statistics for Preferred Sample

	Urban		Rural	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
	A. Demographic Information			
Male	.524	.499	.615	.487
Age	18.655	.887	18.911	.985
CCP Member	.004	.062	.004	.062
	B. NCEE Score Percentile Ranks			
Total Score	.500	.299	.500	.278
Chinese Score	.517	.292	.482	.283
Math Score	.490	.297	.510	.280
English Score	.515	.298	.484	.277
	C. College Admission Dummies			
Any College	.636	.481	.530	.499
Regular Colleges	.317	.465	.230	.421
Project 211 Colleges	.117	.322	.064	.244
Project 985 Colleges	.050	.218	.024	.154
Observations	8,042,128		7,783,937	

Notes: The preferred sample is restricted to ethnic Han first-time NCEE participants who chose English as the foreign language to be tested. For each variable, sample size with non-missing values may vary. NCEE Score percentile ranks are calculated within each province-year-track cluster. In Panel C, the "regular colleges" refer to colleges offering 4-year degree programs, in contrast to 3-year community colleges. The "211 colleges" refer to colleges included in the 211 Program, which are broadly regarded as the top 100 colleges in China. The "985 colleges" are a more select group within the 211 colleges, participating in the 985 Program and representing the top 39 colleges in the country.

Table A.3: Baseline DDD Results: Chinese and Math Scores as Placebo

	(1)	(2)	(3)	(4)
	Chinese		Math	
Listening×Rural	-.006	-.005	-.006	-.005
	(.007)	(.007)	(.007)	(.007)
Rural-Year FE	✓	✓	✓	✓
Province-Year-Track FE	✓	✓	✓	✓
County FE	✓	✓	✓	✓
Trimming		✓		✓
Obs.	15,825,353	14,243,927	15,825,204	14,243,790
R^2	.051	.048	.054	.043

Notes: This table shows regression results from estimating Equation 1, with Chinese and Math score percentile ranks as the outcome variables. In Columns 2 and 4, we trim our sample by dropping observations with aggregate score percentile rank lower than 0.1. Standard errors are clustered at the province-track level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: Baseline DDD Results: Admission to More Selective Colleges

	(1)	(2)	(3)	(4)	(5)	(6)
	Regular		211		985	
Listening×Rural	-.017*	-.019*	-.008	-.010	-.003	-.004
	(.009)	(.011)	(.006)	(.007)	(.003)	(.003)
Rural-Year FE	✓	✓	✓	✓	✓	✓
Province-Year-Track FE	✓	✓	✓	✓	✓	✓
County FE	✓	✓	✓	✓	✓	✓
Trimming		✓		✓		✓
Obs.	15,825,397	14,243,935	15,825,397	14,243,935	15,825,397	14,243,935
R^2	.099	.107	.059	.063	.035	.037

Notes: This table shows regression results from estimating Equation 1 with outcome variables being indicators for admission into 4-year regular, Project 211, or Project 985 colleges, respectively. "Regular" refers to colleges offering 4-year degree programs, in contrast to 3-year community colleges. "211" refers to colleges included in the 211 Program, which are broadly regarded as the top 100 colleges in China. "985" refers to a more select group within the 211 colleges, participating in the 985 Program and representing the top 39 colleges in the country. In Columns 2, 4, and 6, we trim our sample by dropping observations with aggregate score percentile rank lower than 0.1. Standard errors are clustered at the province-track level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Private vs. Public Input: Absorbing High School Fixed Effects

	(1)	(2)	(3)
	English	Aggregate	Admitted
Listening×Rural	-.001	.006	.004
	(.008)	(.008)	(.006)
Rural-Year FE	✓	✓	✓
Province-Year-Track FE	✓	✓	✓
High School FE	✓	✓	✓
Obs.	15,702,031	15,701,894	15,702,031
R^2	.285	.320	.190

Notes: This table presents regression results from estimating Equation 1 while absorbing high school fixed effects. In Columns 1 and 2, outcome variables are percentile ranks for English and aggregate NCEE score, respectively. In Column 3, the outcome variable is an indicator for admission into any college. Standard errors are clustered at the province-track level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: Pilot English Listening Exams as Preparation

	(1)	(2)	(3)
	English	Aggregate	Admitted
Listening×Rural	-.022***	-.014*	-.020***
	(.008)	(.007)	(.007)
Prepared×Listening×Rural	.012	.015	.009
	(.013)	(.013)	(.023)
Rural-Year FE	✓	✓	✓
Province-Year-Track FE	✓	✓	✓
County FE	✓	✓	✓
Obs.	15,825,397	15,825,260	15,825,397
R^2	.062	.060	.129

Notes: This table presents regression results from estimating Equation C.1. In Columns 1 and 2, outcome variables are percentile ranks for English and aggregate NCEE score, respectively. In Column 3, the outcome variable is an indicator for admission into any college. Standard errors are clustered at the province-track level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.7: Pilot English Listening Exams as Placebo

	(1)	(2)	(3)
	English	Aggregate	Admitted
Listening×Rural	-.021**	-.013*	-.017**
	(.008)	(.007)	(.007)
Pilot×Rural	-.005	-.006	.012
	(.011)	(.012)	(.013)
Rural-Year FE	✓	✓	✓
Province-Year-Track FE	✓	✓	✓
County FE	✓	✓	✓
Obs.	15,825,397	15,825,260	15,825,397
R^2	.062	.060	.129

Notes: This table presents regression results from estimating Equation C.2. In Columns 1 and 2, outcome variables are percentile ranks for English and aggregate NCEE score, respectively. In Column 3, the outcome variable is an indicator for admission into any college. Standard errors are clustered at the province-track level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.8: Probing Selective NCEE Participation

	(1)	(2)	(3)	(4)
Dep. Var.	NCEE Share Urban		Δ Part. Rate	
Listening	-.008	-.006	-.002	-.006
	(.015)	(.015)	(.029)	(.028)
Cohort Share Urban		.794***		-1.245***
		(.246)		(.473)
Province FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Obs.	142	142	142	142
R^2	.933	.937	.918	.923

Notes: Columns 1 and 2 show regression results from estimating Equation D.1 and its variant, with the outcome variable being the proportion of urban students among all NCEE takers in a given province-year. Columns 3 and 4 show regression results from estimating Equation D.2 and its variant, with the outcome variable being the gap in shares of urban and rural "potential cohort" who eventually took NCEE in a given province-year. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

B Rural Disadvantages in English Listening Learning

English instruction, especially listening and speaking, is riddled with various human capital and logistical constraints in rural secondary education. First of all, teachers in rural junior and senior high schools typically possess limited English listening and speaking proficiency themselves, making it difficult to ensure the quality of English listening instruction for their students (Hu, 2002; Zhu, 2014). Relative to teachers from better-developed urban regions, rural teachers may also perform poorer in adapting to new exam requirements in NCEE (such as the addition of English listening), which requires new instruction methods ((Hu, 2007)).

Secondly, logistical constraints among rural high schools are particularly salient in the relatively poor performance of English listening among rural students. Compared with urban schools, rural schools lack multi-media devices such as computers, projectors, speakers, and related software that are critical for English listening instruction (Hu, 2002; Zhu, 2014).

Besides the aforementioned constraints faced by rural schools, rural students are also disadvantaged in private inputs required for English listening learning. Firstly, due to budget and information constraints, as well as peer effects, rural students rarely have access to extracurricular English learning, which can help build English proficiency as well as confidence. Rural students are also less exposed to an "English learning environment," such as parents who know English themselves or English movies and TV programs (Zhu, 2014).

Moreover, information and cultural barriers also constrain rural English listening proficiency. Materials in English listening instruction and exams are often related to foreign countries' international news or cultural practices, to which rural students have less access (Wang, 2000; Liu, 2008, 2009; Song, 2017). Besides, Zhang (2018) argues that the perceived labor market returns to English listening skills among rural students are lower than their urban counterparts, which gives rural students less incentive to put effort into acquiring English listening skills.

C Pilot Listening Exams

We probe for potential "preparation effects" of pilot listening exams by estimating the following equation:

$$\begin{aligned} y_{irpst} = \beta \cdot listening_{pt} \times rural_i + \gamma \cdot prepared_p \times listening_{pt} \times rural_i \\ + \theta \cdot prepared_p \times rural_i + \lambda_{rt} + \delta_{pst} + \varepsilon_{ipst} \end{aligned} \tag{C.1}$$

where $prepared_p$ is an indicator for province p to have ever conducted a pilot listening exam between 1999 and 2003.

Appendix Table A.6 shows that there are no statistically detectable preparation effects of pilot listening exams.

We also leverage the pilot listening exams as a placebo test to our baseline DDD approach by estimating the following equation:

$$y_{irpst} = \beta \cdot listening_{pt} \times rural_i + \gamma \cdot pilot_{pt} \times rural_i + \lambda_{rt} + \delta_{pst} + \varepsilon_{ipst} \tag{C.2}$$

where $pilot_{pt}$ is an indicator for province p in year t to have conducted a pilot listening exam.

As shown in Appendix Table A.7, pilot listening exams do not affect NCEE scores or college admission, while the estimated effects of English listening exams remain largely unchanged.

D Selective NCEE Participation

This section discusses robustness checks on whether our treatment of interest (the introduction of English NCEE listening exams) affects sample selection. As discussed in Section 4.6, for each province-year-*Hukou* cluster, we construct the "potential cohort" by weighing the population of relevant birth cohorts from the National Population Census 1990 with the share of birth cohorts observed in our NCEE dataset.

The first specification we estimate is:

$$NCEEShareUrban_{pt} = \beta \cdot listening_{pt} + \gamma \cdot CohortShareUrban_{pt} + \theta_p + \eta_t + \varepsilon_{pt} \tag{D.1}$$

where $NCEEShareUrban_{pt}$ is the proportion of urban students among all NCEE takers in province p, year t. $listening_{pt}$ is an indicator for English listening to be included in province p in year t. We control for $CohortShareUrban_{pt}$, the proportion of urban population among the "potential cohort" who would have taken NCEE in province p in year t. We also control for province fixed effects θ_p and year fixed effects η_t.

For robustness, we also estimate the following alternative specification:

$$\Delta PartRate_{pt} = \beta \cdot listening_{pt} + \gamma \cdot CohortShareUrban_{pt} + \theta_p + \eta_t + \varepsilon_{pt} \tag{D.2}$$

where $\Delta PartRate_{pt}$ is the share of urban "potential cohort" who eventually took NCEE minus the share of rural "potential cohort" who eventually took NCEE in province p in year t.

As shown in Appendix Table A.8, for both specifications, there is no systematic correlation between the introduction of English listening and the urban-rural gap in NCEE participation.