CERGE - EI
Center for Economic Research and Graduate Education –
Economics Institute

Essays on Economics of Education

Bohdana Kurylo

Dissertation

Prague 2026

Dissertation Committee:

Nikolas Mittag (CERGE-EI, chair)
Randall K. Filer (City University of New York)
Štěpán Jurajda (CERGE-EI)

Referees:

Paul Devereux (University College Dublin)
Steven Rivkin (University of Illinois Chicago)

Contents

Acknowledgements

I am deeply grateful to my supervisor, Nikolas Mittag, for his invaluable guidance, encouragement, and insightful feedback throughout my PhD studies. His support has been instrumental in shaping both this dissertation and my development as a researcher. I would also like to sincerely thank Randall K. Filer, Štěpán Jurajda, Daniel Münich, Mariola Pytliková, and Jan Zapal for their thoughtful comments and suggestions, which have greatly strengthened this work.

My thanks extend to the CERGE-EI Academic Skills Center—Deborah Nováková, Gray Krueger, and Andrea Downing—for their meticulous editing, helpful comments on my presentations, and the teaching skills I gained through their training. I am also grateful to the Study Affairs Office (SAO) for their continuous help and support throughout my studies. I wish to thank my co-author, Davit Adunts, for his collaboration and contributions.

Finally, I am indebted to my family and parents for their unwavering understanding, encouragement, and belief in me during this journey. I gratefully acknowledge financial support from the Charles University Grant Agency (GAUK) project No. 314119.

Prague, Czech Republic   Bohdana Kurylo
August 2025

Abstract

This dissertation investigates how teacher-student demographic match, policy-induced changes in migration opportunities, and ordinal rank influence students' educational outcomes. The first chapter examines whether the benefits of having a same-race teacher extend beyond test scores to non-test academic outcomes. Using the random assignment of teachers in the Measures of Effective Teaching (MET) project, I find that Black students assigned to Black teachers not only improve their math performance but also report more effective teacher–student communication. This paper contributes direct evidence on a potential mechanism - improved communication effectiveness - that may help to explain the long-term gains from same-race teachers observed in prior studies. Although communication does not explain short-run test score improvements in my setting, educational research shows that teacher–student rapport is strongly associated with outcomes such as high school graduation and college enrolment. I find that the communication effect is driven by more effective instructional alignment between Black teachers and Black students, consistent with the literature on culturally relevant pedagogy. The second chapter (jointly with Davit Adunts) examines how expanded international migration opportunities influence gender differences in STEM field choices. We study the impact of a 2017 visa liberalization policy between the European Union and Ukraine, which lifted visa requirements for Ukrainian citizens holding biometric passports. Using comprehensive administrative data on university applications from all Ukrainian high school graduates, we analyze how this policy shift affected male and female students' preferences for STEM programs. Employing a difference-in-differences approach and leveraging regional variation in pre-policy emigration rates, we find that the gender gap in selecting a STEM field as a first-choice preference widened by approximately 12.2 percent after the reform, driven primarily by a stronger response among male students to migration opportunities. These insights are particularly relevant for policymakers aiming to reduce gender imbalances in STEM and retain globally-mobile talent to support economic growth. The third chapter examines the long-term academic effects of students' ordinal rank within their kindergarten classroom and how incomplete peer data can bias the estimates of these effects. Using data from Project STAR—a large-scale randomized controlled trial with near-complete test score coverage—I estimate the impact of reading- and math-specific classroom rank on high school

GPA, graduation, SAT/ACT participation, and ACT scores. Students ranked near the top of their kindergarten classroom experience lasting academic advantages, particularly in GPA and test participation. These effects are highly nonlinear and are concentrated among top-ranked students. To assess the consequences of data limitations, I simulate varying levels of peer test score observability under a missing completely at random (MCAR) assumption. The results show that measurement error due to partial peer data disproportionately attenuates estimates at the top of the rank distribution, precisely where the true effects are largest. This alignment between effect heterogeneity and bias severity suggests that studies relying on incomplete peer data may systematically understate the role of top-ranked status in shaping students' educational trajectories. The findings underscore the importance of accurate rank measurement for both empirical validity and the interpretation of early academic dynamics.

Abstrakt

Tato disertační práce zkoumá, jak demografická shoda mezi učiteli a žáky, změny v migračních příležitostech vyvolané politikou a pořadové hodnocení ovlivňují vzdělávací výsledky žáků. První kapitola zkoumá, zda mezi přínosy učitele stejné rasy patří nejen lepší výsledky testů, ale i jiné akademické výsledky. Na základě náhodného přidělování učitelů v rámci projektu Measures of Effective Teaching (MET) zjišťuji, že černošští žáci přidělení černošským učitelům nejen zlepšují své výsledky v matematice, ale také uvádějí efektivnější komunikaci mezi učiteli a žáky. Tato práce přináší přímé důkazy o potenciálním mechanismu – zlepšené efektivitě komunikace –, který může pomoci vysvětlit dlouhodobé přínosy učitelů stejné rasy pozorované v předchozích studiích. Ačkoli komunikace nevysvětluje krátkodobé zlepšení výsledků testů ve zkoumaném prostředí, výzkum v oblasti vzdělávání ukazuje, že vztah mezi učitelem a žákem je silně spojen s výsledky, jako je absolvování střední školy a zápis na vysokou školu. Zjišťuji, že komunikační efekt je způsoben efektivnějšími výukovými postupy černošských učitelů vůči černošským studentům, což je v souladu s literaturou o kulturně relevantní pedagogice. Druhá kapitola (společně s Davitem Aduntsem) zkoumá, jak rozšířené možnosti mezinárodní migrace ovlivňují genderové rozdíly ve výběru oborů STEM. Studujeme dopad politiky liberalizace vízového režimu mezi Evropskou unií a Ukrajinou z roku 2017, která zrušila vízovou povinnost pro ukrajinské občany držící biometrické pasy. Na základě komplexních administrativních údajů o přihláškách na vysoké školy od všech ukrajinských absolventů středních škol analyzujeme, jak tato změna politiky ovlivnila preference mužských a ženských studentů pro obory STEM. Pomocí přístupu „difference-in-differences“ a s využitím regionálních rozdílů v míře emigrace před zavedením této politiky zjišťujeme, že genderová nerovnost ve výběru oboru STEM jako první volby se po reformě zvýšila přibližně o 12,2 %, což bylo způsobeno především silnější reakcí mužských studentů na migrační příležitosti. Tyto poznatky jsou zvláště relevantní pro tvůrce politik, kteří se snaží snížit genderovou nerovnováhu v oborech STEM a udržet globálně mobilní talenty na podporu ekonomického růstu. Třetí kapitola zkoumá dlouhodobé akademické účinky pořadového umístění žáků v jejich mateřské škole a to, jak neúplné údaje o vrstevnících mohou zkreslit odhady těchto účinků. Na základě údajů z projektu STAR – rozsáhlé randomizované kontrolované studie s téměř úplným pokrytím testových výsledků – odhaduji dopad pořadového

umístění v třídě v čtení a matematice na průměrný prospěch na střední škole, absolvování studia, účast na testech SAT/ACT a výsledky testů ACT. Studenti, kteří se umístili na předních místech ve své mateřské škole, mají trvalé akademické výhody, zejména v GPA a účasti na testech. Tyto účinky jsou vysoce nelineární a soustředí se mezi studenty s nejvyšším hodnocením. Abych posoudila důsledky omezených údajů, simuluji různé úrovně pozorovatelnosti výsledků testů vrstevníků za předpokladu zcela náhodného chybějícího údaje (MCAR). Výsledky ukazují, že chyba měření způsobená částečnými údaji o vrstevnících neúměrně oslabuje odhady v horní části rozložení pořadí, tedy přesně tam, kde jsou skutečné účinky největší. Tato shoda mezi heterogenitou účinků a závažností zkreslení naznačuje, že studie založené na neúplných údajích o vrstevnících mohou systematicky podceňovat roli nejlepších žáků při formování vzdělávací trajektorie studentů. Zjištění podtrhují význam přesného měření pořadí jak pro empirickou validitu, tak pro interpretaci rané akademické dynamiky.

Introduction

Educational outcomes are shaped not only by individual effort and ability, but also by structural features of schools and society that can either widen or narrow disparities in opportunity. This dissertation investigates three such dimensions: teacher–student demographic matching, expanded international migration opportunities, and classroom peer composition. Each chapter employs experimental and quasi-experimental methods to identify causal effects, drawing on variation from randomized assignments, policy reforms, and naturally occurring differences in peer environments. By examining mechanisms that operate within classrooms, across education systems, and through broader social contexts, the analysis contributes to the economics of education literature and to our understanding of the drivers of educational inequality.

The first chapter investigates the causal effect of having a same-race teacher on a central non-test academic outcome: teacher–student communication. While prior work has documented positive effects of racial matching on student perceptions, engagement, and behavior (Dee, 2005; Gershenson et al., 2016; Egalite & Kisida, 2018), as well as on long-run outcomes such as graduation and college enrolment (Gershenson et al., 2022), communication has typically been treated as an inferred mechanism rather than directly measured. Using data from the Measures of Effective Teaching Project with randomized classroom assignments, I compare Black students assigned to Black versus non-Black teachers, restricting the sample to students who remained in their initially assigned classrooms to avoid bias from post-assignment sorting. The results show that same-race teachers significantly improve student-reported communication, particularly through clearer explanations of material, consistent with the "culturally relevant pedagogy" hypothesis (Ladson-Billings, 1995; Dee & Penner, 2017). Importantly, these gains occur without reducing other students' outcomes or overall test performance. By providing direct evidence on communication, this chapter helps explain how teacher-student demographic match can generate the long-run benefits documented in other studies and informs debates on teacher workforce diversification.

The second chapter shifts the focus from classroom interactions to policy-induced changes in educational incentives. I examine how the 2017 EU–Ukraine visa liberalization, which expanded short-term mobility to the Schengen Area, influenced Ukrainian students' field-of-study choices at university. While the policy did not directly grant study rights abroad, it plausibly lowered search and signaling costs for pursuing opportunities in EU labour markets, where STEM skills may be more transferable than those in other fields. Using administrative data on the universe of Ukrainian university applicants, I implement a difference-in-differences design that exploits pre-policy variation in regional emigration rates. I find that the reform increased the likelihood that male applicants—particularly in high-emigration regions—chose STEM fields, with little effect on female applicants. These results are consistent with migration opportunities affecting education choices through expected returns, and they contribute to the literature linking international mobility, skill specificity, and educational investment. Policy implications include the potential for migration policy to influence domestic skill supply, particularly in STEM-intensive sectors.

The third chapter examines how ordinal rank within a classroom influences student achievement, and how measurement error in the rank—arising from missing data on peers' test scores—can bias the estimated effects. Using data from Tennessee's Project STAR experiment, I replicate the widely used methodology of ranking students by within-class test scores and show that incomplete peer data leads to systematic attenuation bias in the estimated rank effects. I then develop a Monte Carlo simulation framework to quantify this bias, demonstrating that even moderate levels of missing data can lead to substantial underestimation of the rank effects. These findings contribute to the literature on peer effects and relative standing (Murphy & Weinhardt, 2020; Denning et al., 2021) by highlighting the importance of measurement precision, and they provide methodological guidance for future research using rank-based measures in education settings.

Together, these chapters provide new evidence on how student outcomes are shaped by factors that operate beyond curriculum content and test scores. The first chapter shows how teacher–student demographic match can improve a process-oriented outcome—communication—that supports long-run gains. The second chapter demonstrates how policy-induced changes in mobility can shift educational investments toward fields with higher expected external returns. The third chapter clarifies how peer comparisons affect achievement and how data limitations can distort empirical conclusions. By integrating the insights from each setting, this dissertation advances our understanding of how relational, incentive-based, and social-contextual factors influence educational pathways, with implications for teacher assignment, migration, and school organization policies.

1. Beyond Test Scores: Same-Race Teachers and Teacher-Student Communication Effectiveness

1.1 Introduction

Disparities in cognitive and socio-emotional skills between minority and non-minority students often arise in the period of early childhood[1] and have been shown to have important long-lasting impacts on student well-being (Todd & Wolpin, 2007). One approach to diminish the preschool disadvantage is attracting more effective teachers, who may significantly improve student performance (Rockoff, 2004) and long-term outcomes (Chetty, Friedman, & Rockoff, 2014). However, as previous studies suggest, teacher effectiveness differs across contexts and depends on teacher-school matches (Jackson, 2013; Delgado, 2025) and teacher-class matches (Aucejo et al., 2022; Graham, Ridder, Thiemann, & Zamarro, 2020).

One particular case of matches relates to teachers sharing the same identity with students. Despite the extensive evidence on the positive effects of being matched with same-identity teachers (same-race and same-gender teachers), on student test scores[2], considerably less is known about whether these effects extend beyond achievement to other academically relevant dimensions of the classroom experience, such as student engagement, classroom behavior, and communication effectiveness[3]. Examining the impact of same-identity teachers on these non-test outcomes is important because they capture socio-emotional and interactional processes that are closely linked to learning and are predictive of longer-run educational attainment. Consistent with this view, Jackson (2018) shows that a teacher's effects on student behaviors, including attendance and

[1] Bond and Lang (2018) find that the black-white test gap evolution does not have a racial component in human capital acquisition, but can be explained by differences in socioeconomic characteristics from childhood. However, this finding does not exclude the possibility that future investments may mitigate the initial disadvantage.

[2] For instance, Dee (2004); Egalite, Kisida, and Winters (2015); Fairlie, Hoffmann, and Oreopoulos (2014); Joshi, Doan, and Springer (2018); Lusher, Campbell, and Carrell (2018); Penney (2017a, 2017b) have shown positive effects of same-race teachers on student test scores.

[3] There is scant evidence on the positive effects of a same-race teacher on behavioral and other non-test academic outcomes (Dee, 2005; Egalite & Kisida, 2018; Gershenson, Holt, & Papageorge, 2016; Holt & Gershenson, 2019; Lindsay & Hart, 2017).

disciplinary outcomes in ninth grade, are more strongly associated with high school graduation and college enrollment than the same teacher's effects on test scores.

This paper examines the effects of a same-race teacher[4] on student test scores and teacher-student communication effectiveness as a potential underlying mechanism. Communication effectiveness, i.e., how well a student understands their teacher's explanations, feels heard, and engages in dialogue, is an essential but often overlooked component of the classroom experience. Improved communication may not always translate into immediate test score gains, but it reflects enhanced instructional clarity, student motivation, and classroom rapport, which are likely to influence educational persistence and identity formation.

Moreover, research shows that communication effectiveness may have long-term implications. Hamre and Pianta (2001), for example, demonstrate that kindergarten students who experience poor teacher relationships—marked by weak communication and conflict - have worse academic and behavioural outcomes through eighth grade. In the context of teacher–student demographic match, Gershenson et al. (2022) provide evidence that improved communication is a likely mechanism through which same-race teachers increase college enrolment. Their findings suggest that the effect accumulates with repeated exposure and is not simply a one-time "role model" boost.

To identify the effects of a same-race teacher, I exploit the random assignment of teachers to classes within the U.S. Measures of Effective Teaching (MET) project, which enables me to address the issues related to the systematic sorting of students and teachers. I use the information on student-level perceptions of teaching practices from the Student Perception Survey (SPS) and administrative data to measure teacher-student communication effectiveness and test scores, respectively. I find that being taught by a same-race teacher improves the performance of Black students on math test scores. However, the effects of a same-race teacher on English test scores are small and insignificant. These findings are consistent with previous findings[5] of the randomized STAR study (Dee, 2004) and more recent evidence from observational studies (Joshi et al., 2018). Beyond the effect of a same- race teacher on test scores, I find that matched Black

[4] I limit my analysis to the impact of a same-race teacher and cannot examine the effects of a same-gender teacher on student performance, as most teachers in my sample are female.

[5] The evidence from previous papers on the effects of a same-race teacher is based on the Tennessee STAR project from the 1980s (Dee, 2004), which may be drastically different in terms of the school environment and administrative data from the specific school district or state; e.g., Florida (Egalite et al., 2015).

students report more effective communication with their same-race teachers than do their unmatched schoolmates.

While student-reported communication is often viewed as a positive classroom dynamic, I acknowledge that such subjective measures may not directly reflect academic content mastery or objective learning gains. In this study, I interpret higher communication ratings as indicative of greater teacher engagement, cultural alignment, or classroom rapport, rather than assuming they are inherently beneficial. Crucially, the results show that improved student-reported communication with same-race teachers is not associated with diminished academic achievement among students of other races. Nor do I find evidence of negative effects on standardized test performance overall. This suggests that any improvements in subjective teacher–student rapport do not come at the expense of academic content delivery or peer learning.

Although teacher–student communication does not appear to explain the short-term test score gains associated with same-race teachers, it may still play an important role in shaping students' long-term academic engagement and success. This is particularly relevant in contexts where students experience repeated exposure to demographically similar teachers. Consistent with findings from Gershenson et al. (2022), my results support the view that communication is a key relational outcome that may help explain the persistent, longer-run benefits of teacher–student demographic congruence.

To understand the possible underlying explanations behind the effect of a same- race teacher on communication effectiveness, I examine the effects of a same-race teacher on separate dimensions of communication. The findings indicate that Black students report better understanding of explanations made by same-race teachers than those of White teachers. These findings suggest that the effect of a same-race teacher on communication may be explained by a shared cultural background and culturally aligned instructions, which are in line with the hypothesis of studies about culturally relevant pedagogy (Dee & Penner, 2017; Ladson-Billings, 1995).

Additionally, I do not find evidence supporting two alternative explanations for the positive effect of a same-race teacher on communication effectiveness, including higher general communication ability of Black teachers and more teacher attention directed towards matched students. The latter suggests that gains for matched students are not at the expense of non-matched students.

This paper adds to previous studies on the impacts of same-race teachers on student perceptions and behavioral outcomes (Egalite & Kisida, 2018; Gershenson et al., 2016; Holt & Gershenson, 2019; Lindsay & Hart, 2017) and long-term outcomes (Gershenson et al., 2022) by providing direct evidence on the effects of a same-race teacher on teacher-student communication effectiveness. This finding emphasizes the importance of matching students to a same-race teacher for improving non-test academic outcomes, which may help to explain the improved long-term outcomes documented by Gershenson et al. (2022). This paper also relates to a broader strand of literature on teacher effects and match effects (Aucejo et al., 2019; Graham et al., 2020; Jackson, 2013; Wedenoja, Papay, & Kraft, 2020) by shedding light on the importance of matching minority students with a same-race teacher. Furthermore, the findings align with the hypothesis about culturally relevant pedagogy (Dee & Penner, 2017; Irvine, 1989; Ladson-Billings, 1995), according to which Black teachers are better at instructing same-race students, for instance, they often use more relevant examples, thanks to a higher degree of shared cultural background.

Overall, the findings of this paper may imply that the effects of a same-race teacher extend beyond test scores to non-test academic outcomes, which can help to explain the positive long-term effects of a same-race teacher. In particular, I show that same-race teachers improve the effectiveness of communication with Black students. The evidence further supports that this effect is driven by the higher effectiveness of Black teachers at instructing same-race students in particular, which aligns with the literature on culturally relevant pedagogy.

The remainder of this paper is as follows. Section 2 reviews previous related literature. Section 3 describes data. Section 4 discusses identification and estimation strategies. Section 5 provides evidence of the effects of a same-race teacher on test scores and non-test academic outcomes. Section 6 concludes.

1.2 Literature

A large literature examines how student–teacher racial matching affects educational outcomes. While early work emphasized standardized achievement (e.g., Dee, 2004; Ehrenberg, Goldhaber, & Brewer, 1995), more recent studies highlight non-test academic outcomes—such as engagement, perceptions of instructional quality, attitudes toward learning, and classroom behavior—that are strong predictors of long-run success (Jackson, 2018). These outcomes are not only intrinsically important but also plausible mechanisms through which racial matching influences later educational attainment.

Evidence on non-test outcomes is broadly consistent. Dee (2005) finds that Black students assigned to Black teachers receive more favourable behavioural evaluations. Gershenson et al. (2016) show that same-race teachers hold higher expectations for students' attainment. Egalite and Kisida (2018) report that minority students perceive their teachers' instructional practices more positively when matched by race, while Holt and Gershenson (2019) and Lindsay and Hart (2017) document reductions in absenteeism and disciplinary incidents. Reviews and meta-analyses (Redding, 2019; Egalite, 2024) conclude that racial matching systematically improves student-reported experiences and teacher perceptions, with especially pronounced effects for Black students.

In contrast, evidence on test-score impacts is mixed. While some studies find positive effects (e.g., Dee, 2004; Gershenson et al., 2022), others report null or heterogeneous impacts. Penney (2023), for instance, finds no consistent achievement gains overall, though effects are larger for lower-achieving students. This pattern suggests that racial matching may yield its primary benefits through non-test channels such as communication, motivation, or engagement.

The most direct evidence on long-run effects comes from Gershenson et al. (2022), who use experimental (Tennessee STAR) and quasi-experimental (North Carolina) data to show that exposure to at least one Black teacher in early grades increases high school graduation and college enrolment rates for Black students. By comparing students with one versus multiple same-race teachers, they infer that role model effects influence aspirations and test-taking, while improved communication and rapport support persistence into higher education. However, communication is not measured directly; instead, it is inferred from patterns of repeated exposure.

This chapter complements the previous papers on the same-race teacher effects by providing direct evidence on whether same-race teachers improve student–teacher communication—a non-test outcome hypothesized to mediate long-run gains. In contrast to Egalite and Kisida (2018), who compare minority and non-minority students and risk upward bias from cross-group differences in reporting, I focus on within-group comparisons among Black students. This design, combined with randomized teacher assignment and the exclusion of students who switched classes or schools, minimizes bias from both reporting heterogeneity and post-assignment sorting.

The results show that Black students with same-race teachers report significantly higher communication effectiveness, especially in the clarity of explanations. This finding aligns with the culturally relevant pedagogy framework (Ladson-Billings, 1995; Dee & Penner, 2017), which posits that shared cultural background enables teachers to connect content to students' experiences and use more relevant examples. Importantly, these communication gains do not come at the expense of other students' outcomes or overall test performance, indicating that rapport and rigor can coexist.

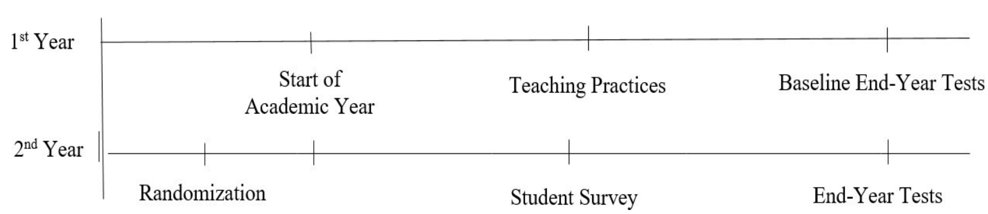
More broadly, this work extends the literature on teacher–student match effects (Aucejo et al., 2019; Jackson, 2013) by shifting the focus from achievement gains to a relational process—communication—that is central to learning yet rarely measured directly. By quantifying this mechanism in a causal framework, the chapter helps explain how racial matching shapes the student experience and potentially underpins the long-run benefits documented in prior research.

1.3 Data

1.3.1 The MET Project

The Measures of Effective Teaching (MET) project was initiated by the Bill & Melinda Gates Foundation to identify and measure effective teaching practices through a combination of classroom observations, student surveys, and achievement gains (Kane, McCaffrey, Miller, & Staiger, 2013). The project took place across six large urban school districts[6] in the United States during the 2009–2011 academic years. In the first year of the study, researchers collected comprehensive baseline data on teaching practices, teacher characteristics, and student achievement based on end-of-year standardized tests (Figure 1.1).

Figure 1.1 The timeline of the MET project



[6] In particular, the districts include New York City Department of Education, Charlotte-Mecklenburg Schools, Denver Public Schools, Memphis City Schools, Dallas Independent School District, and Hillsborough County Public Schools.

In the second year, teachers were randomly assigned to classrooms within schools to generate exogenous variation in teacher–student matching. Schools volunteered to participate, and principals identified "exchange groups" of teachers who taught the same subject and grade level in the upcoming school year. These exchange groups served as the foundation for the randomization procedure: within each group, the MET research team randomly assigned student rosters to teachers, creating "randomization blocks" defined by subject, grade, and instructional period (Kane et al., 2013). Teachers without eligible peers in their subject–grade combination (singletons) were excluded from randomization.

Random assignment procedures were coordinated and overseen centrally to ensure consistent implementation across participating school districts. Each school submitted standardized scheduling spreadsheets identifying eligible teachers, instructional periods, and student rosters, which were systematically reviewed and verified by the MET research team prior to assignment. Through this process, the project established 668 randomization blocks across 284 schools, yielding 1,591 successfully randomized teachers. The design effectively eliminated potential sources of self-selection: students were not permitted to select teachers or classes, and all teacher–student matches within a given subject and grade were determined by the experimental protocol.

In practice, however, full compliance with random assignment was not achieved. The randomization was carried out during the summer of 2010, before schools could confirm which teachers or students would actually be present at the start of the school year. Following random assignment, some students transferred to other schools or switched teachers within the same school, while some teachers left their positions or were reassigned to different courses or grade levels. In certain cases, schools chose not to implement the assigned rosters at all. As a result, a considerable share of students were ultimately taught by a different "actual" teacher than the one to whom they had been randomly assigned. Compliance varied substantially across districts and schools: the highest rates were observed in Dallas (66%), Charlotte-Mecklenburg (63%), and Hillsborough (56%), where most deviations stemmed from students moving to classes outside their original randomization block (Kane et al., 2013). Hence, this paper uses the sample of students who did not deviate from the randomly assigned teachers.

1.3.2 Sample and Descriptive Statistics

The main analytical sample includes elementary school students in 4-5th grades and secondary school students in 6-8th grades whose teachers are randomly assigned[7] and participated in the MET project until the end of the 2010-2011 academic year. Furthermore, I restrict the sample to students who complied with random assignment to classrooms, and about whom there is available information on socio-demographic characteristics, student perceptions of teaching practices, and test scores. The resulting sample includes students from five school districts[8]. In most cases, primary-school students have one general elementary teacher and the same peers in both subjects for the school year. Secondary-school students have two subject specialist teachers: one each for Math and English. Table A1.1 presents the summary statistics for the main analytical sample. The sample consists of 21 % Black students, 28 % White students, 40 % Hispanic students, and 11 % other-race students, including Asian, American Indian, and non-specified-race students. The sample is gender-balanced; 48 % of students are Male. More than half (60 %) qualify for the free or reduced-price lunch (FRL) program. 14 % of students are English Language Learners (ELL), 10 % are 'gifted' and 7 % are classified as having special educational needs.

The racial representation of teachers in the sample includes 74 % who are White and 26 % who are Black[9]. The majority are female (83 %). The overall fraction of students matched to same-race teachers is 41 %; however, there is considerable heterogeneity across racial groups. White students have a considerably higher probability of being taught by same-race teachers at 84 %, while Black students are matched at 51 %. Table A1.2 presents the mean of Black and White teacher characteristics. Black teachers in the sample have, on average, about two years less experience than their White counterparts, however, the p-value from the Kolmogorov-Smirnov test indicates that the difference is insignificant.

Black teachers are more likely to have a Master's or higher degree than their White colleagues. In terms of prior observed teaching practices related to communication according to FFT protocol

---

[7] From the core sample of the second year of the MET project (2,086 teachers from 310 schools), 1,559 randomly assigned teachers from 284 schools continued in the study and 184 teachers dropped out between the random assignment and the start of the school year. Specifically, the number of randomized teachers of grades 4-5 is 470.

[8] Initially, schools from six school districts participated in the MET project, however, I do not observe free and reduced-price lunch eligibility of students and prior observed teaching practices in one of the districts.

[9] I restrict the sample to White and Black teachers, as the fraction of Hispanic and other-race teachers is negligible in the original sample.

and prior value-added in Math and English, there are no significant differences in these measures of teacher effectiveness between the Black and White teachers.

1.3.3 Measures of Student Performance and Other Non-Test Academic Outcomes

I exploit the end-of-year achievement tests and student perceptions of teaching practices to measure student performance and teacher-student communication effectiveness, respectively. The scores from the end-of-year achievement tests are standardized within the school district, such that test scores have zero mean and unit standard deviation. I use the second-year standardized test scores in Math and English as my outcome variables. To measure teacher-student communication effectiveness, I use information on student perceptions of teaching practices from the Student Perception (or Tripod) Survey. The Survey contains information on how students evaluate seven dimensions of classroom instruction: Care, Control, Clarify, Challenge, Captivate, Confer, and Consolidate[10]. There are two versions of the survey for elementary-school and secondary-school students. In classes for general teachers, a randomly selected half of the class filled out the survey while thinking about their English class, and the other half completed the survey while thinking about their Math class. Most questions on a Tripod survey use Likert-type response options with a 5-point scale (Totally Untrue to Totally True).

The MET researchers created the composite measure of teacher-student interactions exploiting the factor analysis. Six dimensions of teaching practices as perceived by students load on one factor (care, captivate, consolidate, clarify, confer and challenge) with the exception for control. Appendix Table A1.3 shows the correlation between the different dimensions of classroom instruction evaluated by students and the loadings on each dimension after performing an oblique rotation of the factors. In comparison to the given measure of teacher-student communication, I create alternative measure of the teacher-student communication effectiveness using information on fourteen underlying questions (Table A1.4) related to communication

[10] According to the description of instruments on the Measures of Effective Teaching Longitudinal Database website, these seven dimensions are defined as follows. "Care measures student perceptions of whether the classroom is a safe place. Clarify measures student perceptions of teacher behaviors that help students to better understand the content being taught. Challenge measures student perceptions of classroom rigor and required effort. Captivate measures student perceptions of how well the teacher captures the attention and interest of students. Confer measures student perceptions of how much a teacher takes students' points of view into account when teaching. Consolidate measures student perceptions of how much the teacher helps students cognitively represent what they have learned in a connected way and how well the teacher promotes student understanding of the interconnectedness of different curriculum topics".

between teacher and students from Student Survey. The questions are identically formulated across two versions of survey for elementary- and secondary-school students. I conduct the factor analysis to construct the measure of teacher-student communication effectiveness.

1.4 Identification and Estimation

1.4.1 Identification Assumption and Related Issues

The main identification assumption of the impact of same-race teachers is that the probability of being matched with a same-race teacher is not correlated with student characteristics, conditional on school-grade-subject fixed effects. I perform a range of balance tests to verify that my identification assumption holds. Columns 1 and 2 of Table A1.5 provide evidence that exposure to a same-race teacher for Black students within the randomization blocks does not depend on students' observed characteristics, including prior test scores, gender, English language learner (ELL) status, eligibility for free or reduced-price lunch (FRL), 'gifted' or special education needs (SPED) status, or enrollment in English and Math classes, respectively.

The impact of these variables on being assigned to same-race teachers is jointly insignificant (p-value of 0.61 and 0.29 for each subject, respectively). Columns 3 and 4 of Table A5 similarly show that White student's exposure to a same-race teacher does not depend on student characteristics. Hence, there is no evidence that the main identification assumption of the model does not hold.

Because teachers are randomly assigned to a classroom within the randomization block, the exogeneity of being assigned to a same-race teacher is ensured in the case of perfect randomization. However, there was teacher attrition from the first to the second year of the MET project. Attrition occurred because teachers were not scheduled to teach the grade and subject, or they chose not to participate (Kane et al., 2013). One hundred eighty-four teachers dropped out of the study between the random classroom assignment and the start of the second school year.

Despite the teacher attrition, Kane and Staiger (2012) show that samples of teachers participating in the first and second years do not have different characteristics in terms of race and prior teaching experience. Assignments to a same-race teacher may still be endogenous due to non-random student sorting into classes. If parents of Black students whose parents are more involved are more likely to choose a school/class with same-race teachers, or if high-ability Black students systematically sort into classes taught by Black teachers, the effects of same-race teachers may be overstated. I test whether classroom and teacher characteristics predict non-compliance of students to class assignment. Columns 1 and 2 of Table A1.6 show that classroom and teacher characteristics do not predict non-compliance of students to classes in Math and English. The exception is in the case of very experienced Math teachers, which is positively correlated with students actively enrolling in their classes, thus potentially violating the assumption of random sorting. However, the impact of classroom and teacher characteristics is jointly insignificant in Math (p-values is 0.74). Hence, systematic student sorting into classes in terms of observable characteristics is not likely to violate the identification assumption and affect the results.

Another potential identification issue is reverse causality between communication and student test scores. Students who earned higher test scores on state exams may report better communication, and/or students who report effective communication may have a higher level of innate ability. However, the possibility of reverse causality is eliminated by the timing of student reporting on communication and taking state exams. The MET researchers administer the Student Perception Survey in the fall semester (the end of October/ the beginning of November), while state exams were administered at the end of the academic year (April-June). Hence, teacher evaluations of student performance and state exam scores did not influence student reports of their perceptions of teaching practices, which I use to measure teacher-student communication.

1.4.2 Model Specifications

To estimate the effect of a same-race teacher on student outcomes[11], I estimate a linear model:

$$\boldsymbol{Y^l}_{\boldsymbol{isgk}} = \boldsymbol{\alpha_0} + \boldsymbol{\alpha_1 BS} \times \boldsymbol{BT_i} + \boldsymbol{\alpha_2 W\,S} \times \boldsymbol{BT_i} + \boldsymbol{\alpha_3 W\,S} \times \boldsymbol{W\,T_i} + \boldsymbol{\alpha_4 X_i} + \boldsymbol{\theta_{sgk}} + \boldsymbol{\varepsilon_{isgk}} \quad \textbf{\textit{(1.1)}}$$

where i*, s, g*, and k index students, school, grade, and subject, respectively. Upper index l denotes the set of student outcomes, including standardized test scores, teacher-student communication, teacher expectations, and student beliefs. $BS \times BT_i$ is a binary variable equal to one if a Black student i was taught by a Black teacher. $W\,S \times BT_i$ is a dummy variable equal to one if a White student is taught by a Black teacher and $W\,S \times W\,T_i$ is a dummy variable equal

[11] Student outcomes include both test scores and one of non-test academic outcomes, particularly, teacher-student communication effectiveness.

to one if a White student is taught by a White teacher. In the full regression, I also include the binary variables for racial interactions of Black and White teachers with Hispanic and other-race students. The comparison group is Black students taught by White teachers. The within-group comparison of Black students is particularly important for estimating the impact of a same-race teacher on non-exam academic outcomes since it allows me to address the differences in perceptions of teaching practices between students of different racial groups. $\boldsymbol{X_i}$ is a vector of predetermined characteristics of students and teachers, including student prior test scores, gender, English language learner (ELL) status, 'gifted' or special educational needs status (SPED), free or reduced- price lunch eligibility (FRLS), and teacher gender: prior teacher effectiveness is measured by value-added, teacher experience, prior teaching practices are measured according to classroom-based protocol, Framework For Teaching (FFT*).* $\boldsymbol{\theta_{sgk}}$ are random block or school-grade-subject fixed effects, and $\boldsymbol{\varepsilon_{isgk}}$ is standard error. I cluster standard errors at the level of randomization blocks, which is equivalent to school-grade-subject.

Because teachers are randomly assigned to classrooms within blocks, student characteristics are orthogonal to teacher race and race-match status in expectation. The inclusion of student controls serves primarily to improve precision rather than to identify the same-race effect. For this reason, I include these characteristics additively rather than interacting them with race-match indicators.

Teacher effectiveness measures are also included additively. This specification allows average teacher effectiveness to differ by race but restricts the productivity of specific teacher attributes, such as value-added or instructional practices, to be constant across student–teacher racial pairings. While interacting these measures with race-match indicators would allow for richer forms of heterogeneity, doing so would substantially increase the complexity of the model and reduce its statistical power, particularly given the noise inherent in some effectiveness measures. Reassuringly, the estimated same-race teacher coefficient is largely unchanged when these controls are included, suggesting that differential sorting on observable teacher effectiveness is unlikely to explain the results.

The main parameter of interest is α_1, which measures the average outcome gains for Black students from being taught by Black teachers compared to Black students taught by White teachers. The parameter related to other combinations of racial interactions, for instance, α_2, which stands for the effect of a Black teacher on the outcomes of White students, allows me to shed more light on whether the effect of a same-race teacher is/not confounded by better general teacher ability to communicate. A positive impact of Black teachers on communication with non-matched/other-race students would mean that Black teachers are more effective communicators. If Black teachers are on average better at communication with students of all racial groups, then the effect of a same-race teacher will be overstated. To estimate the effect of a same-race teacher on test scores, I use a value-added specification that controls for prior test scores on the right-hand side. Although the value-added model specification may be highly sensitive to endogeneity bias when relevant inputs are omitted (Todd & Wolpin, 2003), it is commonly used by previous literature.

To explore potential mechanisms underlying the estimated same-race teacher effect, I augment the baseline specification with a measure of teacher–student communication effectiveness:

$$\boldsymbol{Y_{isgk} = \alpha_0 + \alpha_1 BS \times BT_i + \alpha_2 W\,S \times BT_i + \alpha_3 W\,S \times W\,T_i + \alpha_4 X_i + \beta Communication_{isgk} + \theta_{sgk} + \varepsilon_{isgk}} \quad \textit{(1.1)}$$

where i, *s, g*, and k index students, school, grade, and subject, respectively. $\boldsymbol{Y_{isgk}}$ denotes the set of student standardized test scores. $\boldsymbol{Communication_{isgk}}$ captures teacher–student communication effectiveness as reported in the student survey. Importantly, communication effectiveness is not randomly assigned and may be jointly determined by unobserved student ability, teacher effort, or classroom dynamics. Conditioning on communication may therefore introduce post-treatment or omitted-variable bias, and the resulting estimates should not be interpreted as causal mediation effects in the sense of Imai et al. (2010)[12].

The mediation exercise is intended to be a descriptive decomposition rather than a causal mediation analysis. Specifically, the comparison $\boldsymbol{\alpha_1}$, $\boldsymbol{\alpha_2}$, and $\boldsymbol{\alpha_3}$ across specifications with and without $\boldsymbol{Communication_{isgk}}$ assesses how much of the reduced-form same-race teacher effect is mechanically attenuated when communication is included as an additional control. Under strong assumptions, this attenuation can be interpreted as an upper bound on the extent to which communication accounts for the same-race teacher effect. Given the potential endogeneity and

[12] Formally, a causal interpretation of the mediation effect would require a version of sequential ignorability (Imai et al., 2010): conditional on observed covariates, (i) teacher race match is independent of potential outcomes and potential communication measures, and (ii) communication is independent of potential outcomes given teacher race match and controls. While the experimental assignment of teachers supports the first condition, the second condition is unlikely to hold in this context, as communication is plausibly influenced by unobserved factors that also affect student outcomes.

measurement error in the communication measure, the results are interpreted cautiously and serve primarily to shed light on plausible mechanisms rather than to establish causal mediation.

1.5 Results

This section presents the main results on the effects of a same-race teacher on student performance and robustness checks. Subsection 1.5.1 demonstrates the effects of a same-race teacher on student test scores and teacher-student communication effectiveness. Subsection 1.5.2 documents the heterogeneity of the effects of a same-race teacher and possible underlying mechanisms. Subsection 1.5.3 describes the robustness checks.

1.5.1 The Effects of a Same-Race Teacher on Test Scores and Communication Effectivenes

In this subsection, I demonstrate the findings on the impacts of same-race teachers on standardized test scores and student-reported communication effectiveness. The first two columns of Table 1.1 present the results of a value-added specification, in which the Math test score in the 2010-2011 academic year is the outcome and a prior test score is the control, while the next two columns show the estimated results of a more restricted specification, so-called test score gains, where the outcome is the difference in Math test scores across two adjacent grades. The comparison group is Black students assigned to White teachers in the same school, grade, and subject. The effects of a same-race teacher on Math test scores[13] are positive and significant in all specifications and vary from 0.12 to 0.20 of SD.

Table 1.1 The Effects of a Same-Race Teacher on Math Test Scores

Standartized Test Scores	Value-Added		Test-Score Gain	
	(1)	(2)	(3)	(4)
Black T× Black S	0.134**	0.122**	0.201***	0.194***
	(0.059)	(0.061)	(0.066)	(0.067)
Black T× White S	0.064	0.047	0.004	-0.007
	(0.073)	(0.074)	(0.081)	(0.081)

13

White T× White S	0.061	0.064	-0.011	-0.008
	(0.039)	(0.039)	(0.041)	(0.041)
Male Teacher		-0.040		0.032
		(0.056)		(0.061)
Prior Teacher Value-Added		0.237*		0.219
		(0.132)		(0.145)
Within-District		-0.001		-0.000
Teacher Experience		(0.003)		(0.003)
Prior Teaching Practices		-0.064		-0.007
FFT Communication		(0.079)		(0.067)
R-squared	0.715	0.716	0.166	0.168
Observations	1,637	1,637	1,637	1,637

Notes: The comparison group is Black students taught by White teachers. Models include controls for predetermined student characteristics, including prior test scores, student ELL status, SPED status, 'gifted' status, free and reduced-price lunch eligibility, gender, age; teacher gender, prior value-added, prior observed teaching practices in communication and randomization block fixed effects. Standard errors in parentheses are clustered at the level of randomization block. * p < .10, ** p < .05, *** p < .01

Columns 2 and 4 of Table 1.1 show that the effects of a same-race teacher are robust to inclusion of teacher characteristics including gender, experience within school district, prior value-added[14], prior observed teaching practices according to FFT, and prior average student perceptions of teaching practices. The fact that results are robust to inclusion of teacher characteristics may suggest that systematic differences in teacher effectiveness and other teacher characteristics between Black and White teachers do not drive the result.

The second row of Table 1.1 provides evidence that being taught by Black teachers has small and insignificant impacts on the performance of White students on Math test scores, which may suggest that the positive effects of same-race teachers on Black students is not driven by higher effectiveness of Black teachers compared to their White counterparts. However, this result should be viewed with caution, as only five percent of White students are taught by Black teachers in the sample. I do not find th evidence of the effects of same-race tecahers on English test scores

[14] The effects of a same-race teacher on Math test scores does not vary with prior test scores in the specification which allows for the interaction of a same-race teacher with prior student test scores.

(Table A1.7) The findings align with previous findings from the randomized STAR study (Dee, 2004)[15] and observational study by Egalite et al. (2015).

Beyond test scores, I also examine the effects of same-race teachers on teacher-student communication effectiveness[16]. The first two columns of Table 1.2 present the estimated effects of a same-race teacher on communication in English classes, while the second two columns demonstrate the estimates of same-race teachers in Math classes. The comparison group is Black students taught by White teachers. The results[17] in Table 1.2 indicate that being taught by a same-race teacher increases communication effectiveness with Black students by 0.33 of SD and 0.29 of SD, respectively, in English and Math. The results are robust to controlling for teacher quality measured by prior value-added and prior teaching practices according to FFT (odd columns). The magnitude of the estimated effects of a same-race teacher is comparably larger than the size of estimated effects of same-race doctors on communication with Black patients (Alsan et al., 2019), which may be due to differences in the duration of exposure and contexts.

Results in Rows 2, 4, and 5 of Table 1.2[18] show that Black teachers have no significant positive effect on communication with White, Hispanic, and other-race students, suggesting that there are no negative externalities for non-matched students.

Table 1.2 The Impacts of a Same-Race Teacher on Teacher-Student Communication Effectiveness

Outcome = Communication	English classes		Math classes	
Black T× Black S	0.329**	0.338**	0.348**	0.294**
	(0.166)	(0.156)	(0.141)	(0.144)
Black T× White S	0.102	0.098	0.047	0.011

[15] I cannot directly compare the magnitudes of estimates, as Dee (2004) used a percentile rank based on test scores from different math and reading tests, and did not control for prior test scores.

[16] I also study the effects of a same-race teacher on other non-cognitive skills, including grit, effort, and malleability of skills. The results in Table A1.8 show that there is a positive impact of a same-race teacher on grit of Black students, however, there are no effects on student effort and malleability of skills. The choice of non-test academic outcomes is defined by the data availability.

[17] Using the alternative measure of teacher-student communication effectiveness, I repeat the analysis on the impact of a same-race teacher. Table A8 show that results are similar in magnitude to results in Table 2 when I use the measure of teacher-student communication constructed from the underlying questions from the Student survey.

[18] These results should be viewed with caution for two reasons: first, the small sample of White students taught by Black teachers may lead to imprecise estimates; second, White students on average report a lower level of communication than minority students. Hence, the estimated effects may reflect the level difference in reporting from different racial groups of students.

	(0.174)	(0.155)	(0.161)	(0.153)
White T× White S	-0.003	-0.010	0.028	0.027
	(0.071)	(0.072)	(0.095)	(0.094)
Black T × Hispanic S	-0.073	-0.034	0.063	0.015
	(0.164)	(0.157)	(0.145)	(0.139)
Black T × Other-race S	0.060	0.081	-0.058	-0.101
	(0.144)	(0.125)	(0.218)	(0.212)
Teacher controls	No	Yes	No	Yes
R-squared	0.193	0.204	0.194	0.195
Observations	2,970	2,970	2,364	2,364

Notes: The comparison group is Black students taught by White teachers. Models include controls for predetermined student characteristics, including prior test scores, student ELL status, SPED status, 'gifted status', free and reduced-price lunch eligibility, teacher gender, prior value-added, prior observed teaching practices in communication and randomization block fixed effects. Standard errors in parentheses are clustered at the level of randomization block. * $p < .10$, ** $p < .05$, *** $p < .01$

I further explore the extent to which the effectiveness of teacher–student communication may account for the same-race teacher effect on student achievement. To this end, I augment the baseline specification in Equation (1.1) by including a measure of communication effectiveness as an additional control. The results in Table 1.3 indicate that inclusion of communication reduces the estimated same-race teacher effect on Black students' mathematics test scores by approximately 4 percent. This exercise should be interpreted as descriptive rather than causal, as communication is potentially endogenous and may itself be influenced by unobserved student and teacher characteristics. Moreover, conditioning on communication may introduce post-treatment bias if communication lies on the causal pathway from teacher race to student outcomes. Consistent with this interpretation, the estimated attenuation is modest and sensitive to measurement error in the communication measure. These findings align with prior evidence showing that teacher effects on non-test outcomes are only weakly correlated with teacher effects on test scores (Blazar and Kraft, 2017).

Table 1.3. Does Communication Explain the Effect of a Same-Race Teacher on Test Scores?

Outcome = Math test scores	(1)	(2)	(3)	(4)
Black T× Black S	0.229*	0.220*	0.219	0.213
	(0.132)	(0.129)	(0.136)	(0.133)
Black T× White S	0.064	0.065	0.050	0.050
	(0.114)	(0.111)	(0.113)	(0.110)
White T× White S	0.061	0.059	0.064	0.063
	(0.053)	(0.054)	(0.054)	(0.054)
Communication		0.044**		0.043**
		(0.017)		(0.017)
Teacher Characteristics	No	No	Yes	Yes
R-squared	0.742	0.744	0.744	0.746
Observations	1,241	1,241	1,241	1,241

Notes: The comparison group is Black students taught by White teachers. Models include the same set of controls as in Table 1. Standard errors in parentheses are clustered at the level of randomization block. * p < .10, ** p < .05, *** p < .01

1.5.2 Heterogeneity and Possible Explanation

Heterogeneity analysis of the effects of a same-race teacher on communication shows that Black girls and students who are not eligible for free and reduced-price lunch gain most from being taught by a same-race teacher (Panel A of Table A1.10). The explanation behind the larger effect of a same-race teacher for girls may be that girls not only share a race with their teacher but also a gender since the majority of teachers are female. I do not find evidence that students with a lower prior performance report better communication. In Panel B of Table A1.10, I test whether the effect of a same-race teacher varies with teacher characteristics and do not find evidence that the effect differs for teachers with higher prior value-added, teaching practices, and more years of within-district experience.

The important question remains which mechanisms can explain the positive effect of same-race teacher on communication effectiveness. Using the rich survey information provided in the MET Student Perception Survey, I present evidence on three possible mechanisms, including higher general communication ability of same-race teachers, more attention towards same-race students, and better understating of same-race teacher's explanations. First, the evidence on the lack of positive effects of Black teachers on other-race students (Table 1.2) suggests that a better general ability to communicate does not drive the effect of a same-race teacher on communication.

Table 1.4. Heterogeneity of the Effect of a Same-Race Teacher on Communication: by Racial Composition of Class

Outcome = Communication	All classes	English classes	Math classes
Black T× Black S	0.364**	0.367*	0.356**
	(0.142)	(0.206)	(0.172)
Black T × Black S ×	-0.045	-0.073	-0.025
Predominantly Black Classes	(0.177)	(0.234)	(0.206)
Black T × White S	0.104	0.115	0.050
	(0.136)	(0.181)	(0.163)
White T × White S	0.017	0.006	0.029
	(0.057)	(0.071)	(0.094)
Predominantly Black Classes	0.316	0.305	0.423*
	(0.645)	(0.660)	(0.216)
Observations	5,372	2,970	2,364
R-squared	0.186	0.194	0.194

Notes: The comparison group is Black students taught by White teachers. Models include controls for predetermined student characteristics, including prior test scores, student ELL status, SPED status, 'gifted' status, free and reduced-price lunch eligibility, teacher gender, prior value-added, prior observed teaching practices in communication and randomization block fixed effects. I define classes with predominantly Black students as those in which more than two-thirds of the students are Black. The first column additionally controls for subject fixed effects. Standard errors in parentheses are clustered at the level of randomization block. *p < .10, **p < .05, ***p < .01

The second potential explanation for the positive effect of a Black teacher on communication effectiveness may be that same-race teachers pay more attention to matched students. As teacher attention towards a particular student is unobserved, I verify whether the effect of a same-race teacher varies with the fraction of matched students in the class, following Penney (2017b). A teacher in classes with a large fraction of same-race students may give less attention to same-race students than in classes with a small fraction of same-race students due to time constraints. The results in the second row of Table 1.4 show that the interaction effect of a same-race teacher with dummy for the classes with predominantly Black students is small and insignificant. These results suggest that there is no evidence that Black teachers allocate more attention towards same-race students at the expense of non-matched students.

Table 1.5. Impact of a Same-Race Teacher on Components of Teacher- Student Communication

Components of Communication	Black T× Black S (1)	Black T× White S (2)	White T× White S (3)	R^2 (4)
Teacher Explanation	0.254**	-0.012	0.012	0.155
	(0.108)	(0.124)	(0.063)	
Teacher Explanation: Several ways	0.347***	0.130	-0.042	0.148
	(0.113)	(0.143)	(0.060)	
Clear Explanation	0.302***	0.153	0.012	0.169
	(0.109)	(0.120)	(0.054)	
Class Understanding	0.285***	0.104	-0.017	0.137
	(0.092)	(0.131)	(0.063)	
Clarifying Questions	0.175**	0.095	0.032	0.131
	(0.070)	(0.100)	(0.058)	
Checking Understanding	0.139	0.035	-0.005	0.176
	(0.092)	(0.102)	(0.054)	
Thoughts Sharing	0.268*	0.131	-0.001	0.169
	(0.151)	(0.112)	(0.069)	
Students Speak Up	0.316**	0.159	-0.023	0.129
	(0.133)	(0.139)	(0.072)	
Student Explanation	0.270***	0.187*	0.102*	0.121
	(0.102)	(0.112)	(0.059)	
Time to Explain	0.211*	0.088	0.024	0.161
	(0.117)	(0.137)	(0.066)	
Teacher Summarizing	0.195*	-0.041	-0.061	0.160
	(0.117)	(0.137)	(0.079)	
Correcting mistakes	0.173*	0.048	-0.111**	0.167
	(0.096)	(0.104)	(0.048)	
Care	0.371***	0.149	0.142*	0.208
	(0.137)	(0.156)	(0.079)	
Understanding of feelings	0.176	0.088	0.041	0.121
	(0.174)	(0.164)	(0.084)	

Notes: The comparison group is Black students taught by White teachers. Models include controls for predetermined student characteristics, including prior test scores, ELL status, SPED status, 'gifted' status, free and reduced-price lunch eligibility, gender, age, teacher gender, prior value-added, prior observed teaching practices in communication, randomization block and subject fixed effects. Sample consists of 4726 observations and include both Math and English classes. Standard errors in parentheses are clustered at the level of randomization block. * p < .10, ** p < .05, *** p < .01

Third, I examine the effects of a same-race teacher on separate underlying questions related to communication effectiveness (list of questions in Table A1.4). Column 1 of Table 1.5 suggests that more effective communication between Black students and teachers (e.g., better understanding of same-race teachers' explanations) explains the positive effect of a same-race teacher on teacher-student communication effectiveness. These results align with the literature on culturally relevant pedagogy (Irvine, 1989; Ladson-Billings, 1995; Dee and Penner, 2017), according to which Black teachers are more effective at instructing same-race students due to shared cultural background.

1.5.3 Discussion and Robustness Checks

In this subsection, I discuss the possible sources of bias and show that results are robust to various robustness checks.

In-group bias: One concern is related to the in-group bias. Specifically, Black students may report better communication with same-race teachers as they belong to the same race group but do not indeed have more effective communication. To test this issue, I examine whether Black students also report a higher level of happiness and interest when same-race teachers teach them. The results in Table A1.11 show that Black students do not report a higher level of happiness and do not like classes taught by same-race teachers more than classes taught by other-race teachers, suggesting that in-group bias towards same-race teachers is not likely to drive the results.

Are the results affected by average higher effectiveness of same-race teachers? I further test whether the positive effect of same-race teachers stems from same-race teachers being on average more effective at communication than other-race counterparts, I replace the teacher-student communication based on student perceptions with a fixed effect for each teacher to further analyze the teacher's ability to communicate. Afterward, I explore which teacher characteristics correlate with fixed effects estimates obtained from the regression with communication as an outcome variable (Table A1.12). Teacher race explains approximately 60 % of the cross-sectional variation, but the effect of teacher race is small and insignificant. This implies that there is no significant difference in time-invariant teacher ability to communicate between Black and White teachers. In the next four columns of Table A1.12, I add dummies for whether a teacher taught in classes with predominantly Black classes, a content knowledge test, a principal survey rating (PSVY), and teacher experience within the district. The correlation between time-invariant teacher communication ability and those teacher characteristics is positive and insignificant, except for teacher experience. These teacher characteristics do not explain much about the variation in teacher fixed effects. These results provide suggestive evidence that there is no significant difference in time-invariant teacher ability to communicate between White and Black teachers.

Are the results affected by the exposure to a same-race teacher in previous grade? The estimated effects of a same-race teacher on test scores and communication effectiveness may be biased by exposure to a same-race teacher in previous grades. Penney (2017a) finds that the effect of having a second same-race teacher is relatively small; however, earlier exposure to same-race teachers is more beneficial than in later grades. If the effect of a second same-race teacher is decreasing, the estimated effects in Table 1 may be understated and reflect a lower bound of the true estimates.

1.6 Conclusion

This paper investigates the effects of same-race teacher-student matches on both test scores and non-test academic outcomes, focusing in particular on the effectiveness of teacher-student communication. Using the random assignment of teachers within the MET project, I show that Black students assigned to Black teachers perform better on math tests and report significantly higher communication effectiveness. These communication improvements do not appear to stem from general communication ability or preferential attention by Black teachers.

This study makes three main contributions. First, it adds to the growing evidence that teacher-student demographic congruence improves not only test-based outcomes, but also key relational dimensions of classroom life. Second, it provides direct empirical support for the hypothesis that shared cultural background between teachers and students enhances instructional clarity and understanding, in line with theories of culturally relevant pedagogy. Third, it connects these findings to broader educational goals by suggesting that non-test outcomes like communication may help explain the long-run effects of same-race teachers documented in prior research, including improved high school graduation and college enrolment rates (Gershenson et al., 2022).

Although the communication improvements do not fully mediate test score gains in the same class, they may have more subtle and lasting effects. Prior research shows that positive teacher-student interactions—such as better communication, feedback, and trust—are predictive of long-term outcomes, including student motivation, college-going, and labour market success

(Jackson, 2018; Hamre & Pianta, 2001). In this light, communication effectiveness can be understood as a developmental input: it strengthens student engagement and understanding, even if its full payoff is realized years later.

The policy implications are twofold. First, these findings reinforce the importance of increasing teacher diversity, especially in schools serving large numbers of minority students. Second, they suggest that training non-minority teachers in culturally responsive pedagogy may help replicate some of the communication benefits seen in same-race matches. While hiring more Black teachers remains a long-term structural challenge, culturally relevant instructional practices may offer a complementary and more immediate strategy for improving minority student outcomes.

1.A Appendix

Table A1.1 Descriptive Statistics

	Mean (1)	SD (2)	Min (3)	Max (4)
Panel A: Student Characteristics				
Black	0.21	0.41	0.00	1.00
Hispanic	0.40	0.49	0.00	1.00
White	0.28	0.45	0.00	1.00
Other race	0.11	0.31	0.00	1.00
Male	0.48	0.50	0.00	1.00
ELL	0.14	0.34	0.00	1.00
Gifted Status	0.10	0.30	0.00	1.00
Special Education Status	0.07	0.25	0.00	1.00
FRL	0.60	0.49	0.00	1.00
Age	10.48	1.51	7.62	14.56
Prior Math Test Scores	0.25	0.91	-3.00	3.17
Prior English Test Scores	0.25	0.94	-2.93	2.87
Panel B: Teacher Characteristics				
Black	0.26	0.44	0.00	1.00
White	0.74	0.44	0.00	1.00
Male	0.17	0.37	0.00	1.00
Experience within district	8.26	7.34	0.00	41.00
Master degree	0.27	0.45	0.00	1.00
Prior Teaching Practices	2.62	0.35	1.59	3.50
FFT: Communication				
Prior Value-Added	0.09	0.216	-1.06	0.67
Panel C: Outcomes				
Communication	0.06	0.91	-4.49	1.77
Clarify	0.04	0.59	-3.31	1.43
Confer	0.02	0.68	-3.62	1.36
Care	0.06	0.79	-3.27	1.44
Consolidate	0.05	0.79	-2.54	1.20
Captivate	0.03	0.79	-2.51	1.24
Challenge	0.03	0.68	-3.5	0.95
Control	0.05	0.70	-2.49	1.37

Notes: The sample comprises data on the 2010-2011 school year in which teachers were randomly assigned to classes within randomization blocks.

Table A1.2 Teacher Characteristics and Quality

	Years of Experience (1)	Master Degree (2)	Prior FFT Math (3)	Prior FFT English (4)	Prior VA Math (5)	Prior VA English (6)
Black mean	6.84	0.49	2.58	2.61	0.035	-0.003
White mean	8.73	0.21	2.65	2.70	0.009	0.003
P values	0.110	0.000	0.460	0.198	0.217	0.441

Notes: Table reports means of teacher characteristics by race. P-values are taken from the Kolmogorov - Smirnov test.

Table A1.3 "7C" Student Perception Correlations and Factor Loadings

	Clarify (1)	Care (2)	Confer (3)	Consolidate (4)	Captivate (5)	Challenge (6)	Control (7)	Communication Loadings (8)	Strictness Loadings (9)
Clarify	1							0.4247	-0.0061
Care	0.6868	1						0.4200	0.0209
Confer	0.6705	0.6788	1					0.4170	0.0321
Consolidate	0.6442	0.5371	0.5836	1				0.4273	-0.0677
Captivate	0.5947	0.5022	0.4713	0.4124	1			0.3590	0.1168
Challenge	0.5867	0.3860	0.4551	0.3171	0.2435	1		0.3974	-0.0818
Control	0.3933	0.4478	0.4089	0.3246	0.3409	0.2551	1	-0.0091	0.9867
P values	0.110	0.000	0.460	0.198	0.217	0.441			

Notes: The first seven columns show correlations between "7C" student perception components of teaching practices. The last two columns present factor loadings from exploratory factor analysis after performing an oblique rotation and keeping the first two factors. The first factor explains 62 % of the variance in the data, and the second explains another 11 % of variance.

Table A1.4 Underlying Questions Related to the Teacher-Student Communication Effectiveness

#	Question from the Survey	
1.	If you don't understand something, my teacher explains it another way.	Teacher Explanation
2.	My teacher has several good ways to explain each topic that we cover in this class.	Teacher Explanation: several ways
3.	My teacher explains difficult things clearly.	Clear Explanation
4.	My teacher knows when the class understands, and when we do not.	Class Understanding
5.	My teacher asks questions to be sure we are following along when he/she is teaching.	Clarifying Questions
6.	My teacher checks to make sure we understand what he/she is teaching us.	Checking Understanding
7.	My teacher wants us to share our thoughts.	Thoughts Sharing
8.	Students speak up and share their ideas about class work.	Students Speak Up
9.	My teacher wants me to explain my answers –why I think what I think.	Student Explanation
10.	My teacher gives us time to explain our ideas.	Time to Explain
11.	My teacher takes the time to summarize, what we learn each day.	Teacher Summarizing
12.	In this class, we learn to correct our mistakes.	Correcting mistakes
13.	My teacher in this class makes me feel that she/he really cares about me.	Care
14.	My teacher seems to know if something is bothering me.	Understanding of feelings

Table A1.5 Balance Tests

	Black students		White students	
Outcome = Same-race teacher	English (1)	Math (2)	English (3)	Math (4)
Prior test score	-0.024	-0.031	0.008	0.006
	(0.016)	(0.021)	(0.09)	(0.012)
ELL status	-0.107	-0.067	0.005	-0.046
	(0.144)	(0.099)	(0.09)	(0.039)
FRL eligibility	-0.021	0.001	0.003	-0.008
	(0.026)	(0.027)	(0.024)	(0.015)
'Gifted' status	0.038	0.070	0.000	0.031
	(0.063)	(0.063)	(0.031)	(0.026)
Male Student	-0.021	0.013	-0.014	0.003
	(0.020)	(0.016)	(0.018)	(0.018)
SPED Student	0.001	-0.041	0.014	-0.006
	(0.045)	(0.056)	(0.031)	(0.019)
Observations	1,050	875	1,032	807
R-squared	0.6465	0.6975	0.5621	0.7348
Joint test F-statistics	0.75	1.24	0.22	0.76
[p-value]	0.6125	0.2929	0.9697	0.6064

Notes: The dependent variable is an indicator for being taught by a same-race teacher, regressed on student characteristics, controlling for randomization block or school-grade-subject fixed effects. Standard errors in parentheses are clustered at the level of randomization block.

Table A1.6 Non-Compliance of Students to Classes Taught by Randomly Assigned Teachers

Outcome = Non-complier	Math Classes	English Classes
	(1)	(2)
Black Teacher	0.025	0.016
	(0.029)	(0.023)
Male Teacher	0.012	0.083
	(0.021)	(0.064)
Prior Value-Added	-0.026	0.050
	(0.051)	(0.058)
Teacher Experience	0.003*	-0.001
	(0.002)	(0.001)
Prior Classroom	0.013	0.007
Average Test Score	(0.038)	(0.039)
Fraction of Black students	0.031	0.066
	(0.181)	(0.198)
Fraction of Hispanic students	-0.079	0.003
	(0.117)	(0.202)
Fraction of Other-race students	-0.056	0.328
	(0.101)	(0.304)
Fraction of ELL students	-0.099	-0.121
	(0.101)	(0.130)
Fraction of 'Gifted' students	-0.119	-0.054
	(0.211)	(0.106)
Fraction of Male students	-0.008	-0.020
	(0.121)	(0.198)
Fraction of FRL students	0.117	-0.029
	(0.158)	(0.090)
Fraction of SPED students	0.048	0.138
	(0.155)	(0.158)
Observations	5,156	5,861
R-squared	0.7212	0.6426
Joint test F-statistic	0.73	0.42
P-value	0.7434	0.9656

Notes: Each column reports the results from the one regression in which the outcome variable is non-compliance status of students, which equals one if student is a non-complier and zero otherwise. Non-compliers are students who were initially assigned to a class with randomly assigned teachers but specifically opted out for another class or school. Students who were initially assigned to Math classes taught by teachers with more experience are more likely to be non-compliers. The impact of these variables is jointly insignificant (F-statistics= 0.73, p-value is 0.7434).

Table A1.7 The Effect of a Same-Race Teacher on English Test Scores

Specifications	Value-added		Test score gain	
	(1)	(2)	(3)	(4)
Black T × Black S	-0.001	-0.026	0.068	0.053
	(0.051)	(0.053)	(0.054)	(0.057)
Black T × White S	-0.101	-0.133*	-0.162**	-0.181**
	(0.072)	(0.073)	(0.008)	(0.081)
White T× White S	0.004	0.005	-0.055	-0.056
	(0.043)	(0.043)	(0.047)	(0.047)
Male Teacher		-0.059		-0.049
		(0.058)		(0.063)
Prior Teacher Value-Added		0.064		0.014
		(0.138)		(0.160)
Teacher Experience Within School District		0.006**		0.003
		(0.003)		(0.002)
Prior Teaching Practices FFT: Communicate		-0.095		-0.069
		(0.072)		(0.073)
R-squared	0.683	0.684	0.163	0.164
Observations	2,052	2,052	2,052	2,052

Notes: The comparison group is Black students taught by White teachers. Models include controls for student predetermined characteristics, including student prior test score, ELL status, SPED status, „gifted" status, gender, free and reduced-price lunch eligibility, and randomization block fixed effects. Standard errors in parentheses are clustered at the level of randomization block. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$ (two-tailed tests)

Table A1.8 The Impact of a Same-Race Teacher on Other Non-Test Academic Outcomes

Math-specific outcomes	Grit		Effort		Skills malleability	
	(1)	(2)	(3)	(4)	(5)	(6)
Black T× Black S	0.263*	0.260*	0.130	0.128	-0.042	-0.030
	(0.151)	(0.150)	(0.130)	(0.129)	(0.139)	(0.136)
Black T× White S	0.035	0.046	-0.041	-0.033	-0.236	-0.229
	(0.150)	(0.150)	(0.137)	(0.142)	(0.165)	(0.166)
White T× White S	-0.099	-0.104	-0.133*	-0.139*	-0.132	-0.135*
	(0.097)	(0.099)	(0.071)	(0.071)	(0.081)	(0.081)
Black T× Hispanic S	-0.151	-0.150	-0.183	-0.175	-0.028	-0.009
	(0.141)	(0.138)	(0.114)	(0.111)	(0.119)	(0.118)
Black T× Other-Race S	0.081	0.078	-0.031	-0.041	-0.303	-0.299
	(0.194)	(0.195)	(0.180)	(0.186)	(0.211)	(0.208)
Teacher Characteristics	No	Yes	No	Yes	No	Yes
R-squared	0.168	0.169	0.113	0.117	0.172	0.174
Observations	2,036	2,036	2,284	2,284	2,229	2,229

Notes: The comparison group is Black students taught by White teachers. Models include controls for predetermined student characteristics, including prior test scores, student ELL status, SPED status, 'gifted status', free and reduced-price lunch eligibility, teacher gender, prior value-added, prior observed teaching practices and randomization block fixed effects. Standard errors in parentheses are clustered at the level of randomization block. *p < .10, **p < .05, ***p < .01

Table A1.9 The Impact of a Same-Race Teacher on Teacher-Student Communication: Alternative Measures

Outcome	English classes		Math classes	
Communication	(1)	(2)	(3)	(4)
Black T × Black S	0.311**	0.319**	0.337**	0.293**
	(0.154)	(0.137)	(0.131)	(0.136)
Black T × White S	0.156	0.164	-0.014	-0.044
	(0.152)	(0.128)	(0.173)	(0.171)
White T × White S	0.044	0.024	-0.019	0.016
	(0.069)	(0.067)	(0.089)	(0.088)
Black T × Hispanic S	-0.057	-0.005	0.039	0.001
	(0.159)	(0.145)	(0.143)	(0.142)
Black T × Other-race S	0.117	0.138	0.099	0.065
	(0.122)	(0.131)	(0.165)	(0.161)
Teacher Controls	No	Yes	No	Yes
Observations	2,630	2,630	2,067	2,067
R-squared	0.227	0.244	0.239	0.242

Notes: The comparison group is Black students taught by White teachers. Models include controls for predetermined student characteristics, including prior test scores, student ELL status, SPED status, 'gifted status',free and reduced-price lunch eligibility, teacher gender, prior value-added, prior observed teaching practices in communication and randomization block fixed effects. Standard errors in parentheses are clustered at the level of randomization block. *p < .10, **p < .05, ***p < .01

Table A1.10 Heterogeneity of the Effect of a Same-Race Teacher on Communication with Black Students: by Student and Teacher Characteristics

Panel A: Student Characteristics			
X=	Prior Test Score	Male	Low-Income Family
	(1)	(2)	(3)
Same-race teacher × X	-0.038	-0.128*	-0.185*
	(0.053)	(0.069)	(0.096)
Same-race teacher	0.327***	0.395***	0.460***
	(0.120)	(0.123)	(0.127)
X	0.033	-0.075**	0.025
	(0.023)	(0.030)	(0.035)
R-squared	0.193	0.193	0.193
Panel B: Teacher Characteristics			
X=	Prior Value-Added	Prior Teaching Practices	Within-District Experience
	(1)	(2)	(3)
Same-race teacher × X	-0.057	-0.087	-0.001
	(0.297)	(0.079)	(0.010)
Same-race teacher	0.338***	0.334***	0.340***
	(0.119)	(0.117)	(0.124)
X	0.510	0.016	-0.014**
	(0.350)	(0.049)	(0.006)
R-squared	0.193	0.193	0.193
Observations	5,349	5,349	5,349

Notes: The comparison group is Black students taught by White teachers. Models include controls for predetermined student characteristics, including prior test scores, student ELL status, SPED status, 'gifted status', free and reduced-price lunch eligibility, teacher gender, prior value-added, prior observed teaching practices in communication and randomization block fixed effects. Standard errors in parentheses are clustered at the level of randomization block.
*p < .10, **p < .05, ***p < .01

Table A1.11 Robustness Check: In-group Bias

Math-specific outcomes	Happiness		Like Classes	
	(1)	(2)	(3)	(4)
Black T× Black S	0.018	-0.049	0.027	-0.008
	(0.162)	(0.166)	(0.110)	(0.115)
Black T× White S	-0.139	-0.186	-0.089	-0.115
	(0.153)	(0.148)	(0.147)	(0.139)
White T× White S	0.089	0.092	-0.120*	-0.121*
	(0.089)	(0.089)	(0.071)	(0.072)
Teacher controls	No	Yes	No	Yes
R-squared	0.151	0.157	0.191	0.196
Observations	2,333	2,333	2,364	2,364

Notes: The comparison group is Black students taught by White teachers. Models include controls for predetermined student characteristics, including prior test scores, student ELL status, SPED status, 'gifted' status, free and reduced-price eligibility, gender, age; teacher gender, prior value-added, prior observed teaching practices and randomization block fixed effects. Standard errors in parentheses are clustered at the level of randomization block. *p < .10, **p < .05, ***p < .01

Table A1.12 Correlations between Teacher FE and Teacher Characteristics

	Teacher FE				
	(1)	(2)	(3)	(4)	(5)
Black Teacher	0.032	0.043	0.094	0.118	0.232
	(0.175)	(0.182)	(0.281)	(0.187)	(0.373)
Taught in predominantly black classes		0.104			
		(0.329)			
Content Knowledge Test			0.002		
			(0.013)		
Principal Survey Rating (PSVY)				0.103	
				(0.064)	
Within-District Experience					-0.075
					(0.074)
Observations	111	111	99	102	55
R-squared	0.6060	0.6064	0.6075	0.6227	0.7035

Notes: The subsample includes randomized teachers. Teacher fixed effects are calculated from the regression of communication on teacher fixed effects controlling for student characteristics.

2 Gender Differences in STEM Choice Under Exit Options[19]

2.1 Introduction

Gender disparities in STEM fields have persisted for decades, with women remaining significantly underrepresented in disciplines such as engineering, computer science, and physics (Encinas-Martín & Cherian, 2023). This imbalance is particularly concerning given that STEM degrees often lead to higher post-graduation earnings (Arcidiacono, 2004) and are critical drivers of innovation and economic growth (Peri, Shih, & Sparber, 2015). Understanding how students choose their fields, and why these choices differ by gender, is essential for promoting equitable educational outcomes and addressing broader economic challenges.

In an increasingly globalized world, educational choices are shaped not only by domestic labour market conditions but also by international opportunities. A growing body of research shows that the prospect of migration can incentivize greater investment in education in origin countries, particularly in skills that are transferable across borders (Beine, Docquier, & Rapoport, 2008; Docquier & Rapoport, 2012; Shrestha, 2017; Mota Aquino, 2023; Abarcar & Theoharides, 2024; Khanna & Morales, 2017). However, little is known about whether men and women respond differently to these migration incentives, especially in high-value fields such as STEM.

This paper examines how expanded migration opportunities affect gender-specific preferences for STEM fields, using the 2017 visa liberalization policy between the European Union and Ukraine as a natural experiment. The policy, which removed visa requirements for Ukrainian citizens holding biometric passports, led to a substantial increase in migration flows. Ukraine's centralized university admissions system, which requires applicants to rank their preferred degree programs and allocates seats based primarily on standardized test scores, offers a unique setting to study these effects. This institutional structure allows us to compare male and female applicants with similar academic profiles, thereby isolating the impact of the policy on gender-specific educational choices.

We employ two complementary empirical strategies. First, we use an event study framework that compares STEM preferences before and after the policy change, incorporating an interaction term between gender and the post-policy period. Second, we exploit regional variation

[19] Co-authored with Davit Adunts (Institute for Employment Research).

in pre-policy emigration rates through a difference-in-differences (DiD) design, comparing responses of applicants from high-emigration regions to those from lower-emigration areas. This dual approach strengthens our identification of the causal impact of migration opportunities on field-of-study choices.

Our results show that the 2017 visa liberalization significantly widened the gender gap in STEM field choices among Ukrainian university applicants. Before the reform, female applicants were 27.3 percentage points less likely than male applicants to list a STEM field as their first choice. Following the policy change, this gap widened by approximately 12.2 percent, driven primarily by a stronger shift toward STEM among male applicants—particularly those with low academic achievement—while female preferences remained largely unchanged. Importantly, we find no evidence that the policy discouraged high-achieving female students from pursuing STEM fields. We further show that this gendered response is most pronounced in regions with higher pre-policy emigration rates and in regions where male migration intentions were particularly strong. Our findings are consistent across both stated preferences and actual enrollment outcomes, suggesting that expanded international migration opportunities can amplify existing gender disparities in educational choices.

This paper contributes to two main strands of literature. First, it extends research on the determinants of gender differences in STEM field choice. While prior studies have highlighted the roles of academic preparedness (Card & Payne, 2021; Delaney & Devereux, 2019; Jiang, 2021), preferences and lifestyle expectations (Wiswall & Zafar, 2018), gender stereotypes (Favara, 2012), self-perceived ability (Saltiel, 2023), cultural norms (Lipman-Blumen, 1972; Blickenstaff, 2005; Kanny et al., 2014), and peer composition (Park et al., 2018; Brenøe & Zölitz, 2020), few studies have examined the role of migration opportunities. A further contribution lies in our use of complete application data, which allows us to analyze preferences across the full pool of applicants, not just those who are ultimately admitted—a limitation in much of the existing literature (with the exception of Delaney & Devereux, 2019).

Second, our study relates to research on how migration opportunities shape human capital formation in origin countries. While much of this literature focuses on educational attainment broadly (Batista et al., 2012; De Brauw & Giles, 2017; Dinkelman & Mariotti, 2016; Saad & Fallah, 2020; Shrestha, 2017; Theoharides, 2018; Chand & Clemens, 2023), recent work shows that migration prospects can also influence field-of-study decisions, particularly where skill-specific returns abroad are high (Abarcar & Theoharides, 2024; Khanna & Morales, 2017). Our study provides a broader perspective by examining how a nationwide policy shift affects gender-specific interest in STEM fields, rather than focusing narrowly on a single occupation or sector.

The findings have important implications for policymakers seeking to address the global shortage of skilled STEM workers. This shortage is becoming increasingly acute and is compounded by the persistent underrepresentation of women in STEM, which limits the available talent pool. As STEM skills are vital for long-term innovation and economic resilience (Peri, Shih, & Sparber, 2015), understanding how migration policies affect gendered educational choices can help design more effective strategies to build and retain human capital.

The remainder of the paper proceeds as follows. Section 2 provides the institutional background and describes the data. Section 3 presents the empirical strategy. Section 4 discusses the main empirical findings on gender differences in STEM preferences and their response to expanded migration opportunities. Section 5 concludes.

2.2 Institutional Setting and Data

2.2.1 University Admission System

University admissions in Ukraine are centralized, meaning that applicants do not need to apply separately to individual higher education institutions. Instead, all applications are processed through a unified national platform, enabling a standardized and transparent admissions process across the country. Admission to tertiary education programs is based primarily on scores from the Independent External Test (IET)—Ukraine's national university entrance examination—along with high school grade point averages (GPAs). Students take the IET during their final year of high school. Introduced as a pilot between 2004 and 2007 and fully implemented in 2008, the IET was designed to promote fairness and transparency in university admissions.

Unlike standardized tests such as the SAT or ACT in the United States, the IET consists of multiple subject-specific examinations. All applicants are required to take a test in Ukrainian Language and Literature, regardless of their intended field of study. Additional subject requirements vary by university program: for example, STEM programs typically require exams in Mathematics and a science subject (e.g., Physics or Chemistry). Examinations are administered between May and June, with results released approximately three weeks later. Each IET exam is scored on a standardized 200-point scale, with 100 points as the minimum passing mark. Students receive a separate certificate for each subject passed. During the admissions period, applicants submit the relevant certificates corresponding to the subject requirements of their preferred programs. Most programs require three subject scores, including Ukrainian Language and Literature, although some fields—such as Arts or Architecture—may require fewer IET scores combined with a practical or creative entrance examination.

Admission to a specific university program is contingent upon meeting a minimum score threshold, which varies annually depending on the number of available seats and applicant demand. Applicants may rank up to five fields of study and up to seven specific university programs. If an applicant meets or exceeds the cutoff score for their highest-ranked program, they are admitted to it. Otherwise, they are considered for their next-highest ranked program for which their score meets the threshold. Although students may decline the assigned offer and pursue a different program, doing so typically forfeits eligibility for state-funded (tuition-free) study.

2.2.2 Visa Liberalization Policy and Emigration from Ukraine

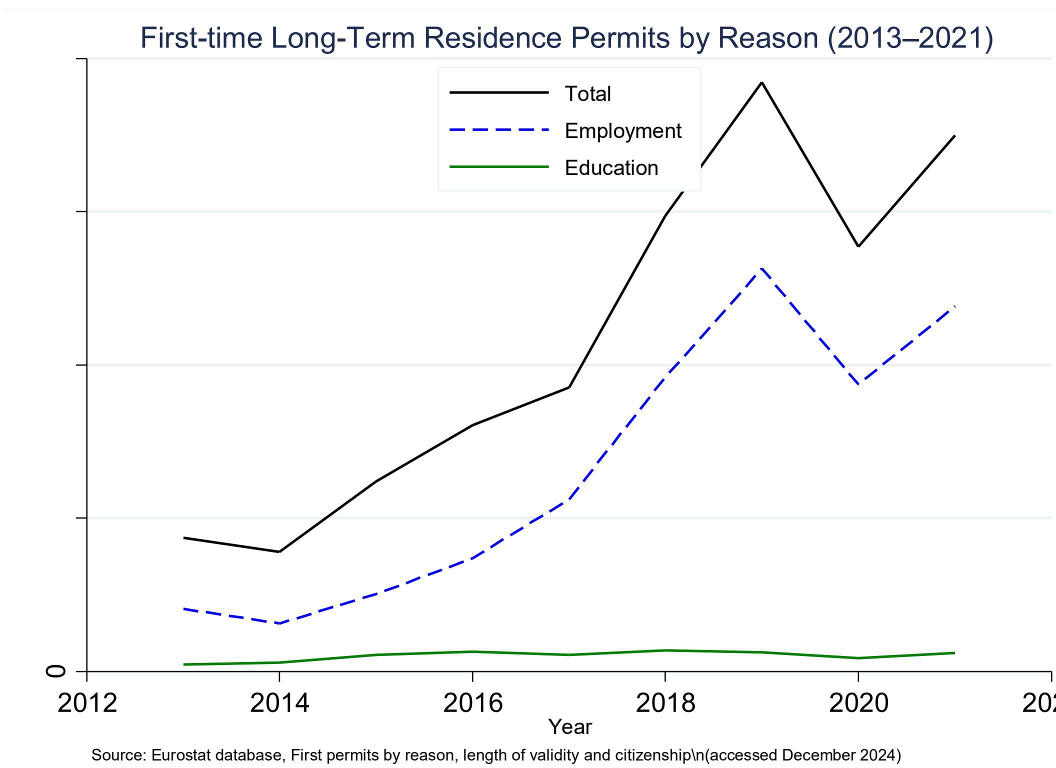
On June 11, 2017, a visa liberalization agreement between the European Union and Ukraine came into effect, allowing Ukrainian citizens holding biometric passports to enter the Schengen Area without a visa for short stays of up to 90 days within any 180-day period. Although the policy did not grant the right to work in Schengen countries without first obtaining a valid work permit, it significantly facilitated international mobility by making it easier for Ukrainians to travel abroad to search for job opportunities. In doing so, it likely reduced job search frictions and signaling costs. Additionally, the policy may have lowered screening costs for foreign employers by enabling face-to-face interactions and preliminary assessments of candidates' qualifications.

While the formal implementation occurred in mid-2017, the policy had been publicly discussed and negotiated for several years. As a result, students and families may have anticipated the reform, at least to some extent. However, the precise timing, scope, and confirmation of the policy introduced a discrete and salient shift in perceived migration opportunities, particularly for individuals nearing the end of secondary education. Importantly, high school students in Ukraine typically begin preparing for their standardized university entrance exams—commonly linked to their intended field of study—at least a year in advance. This makes it unlikely that students could retroactively change their subject choices in response to the sudden confirmation of the reform in 2017.

Following the introduction of the visa-free regime, emigration from Ukraine increased markedly. The total number of first-time long-term (valid for 12 months or more) residence permits issues to the citizens of Ukraine rose from 80,331 in 2016 to 192,196 in 2019 - an increase of approximately 139 percent (Eurostat, 2024). Similarly, the number of long-term residence permits issued for employment purposes rose sharply after the policy's implementation (Figure 2.1). In contrast, the number of first-time residence permits issued for educational purposes remains relatively stable before and after the liberalization, suggesting that the policy primarily affected employment-related migration channels rather than education-driven mobility.

At the same time, Figure 2.1 reveals a pronounced upward trend in permit issuance already between 2014 and 2016, prior to the visa liberalization, which raises the possibility that part of the observed increase reflects confounding factors rather than the policy itself.

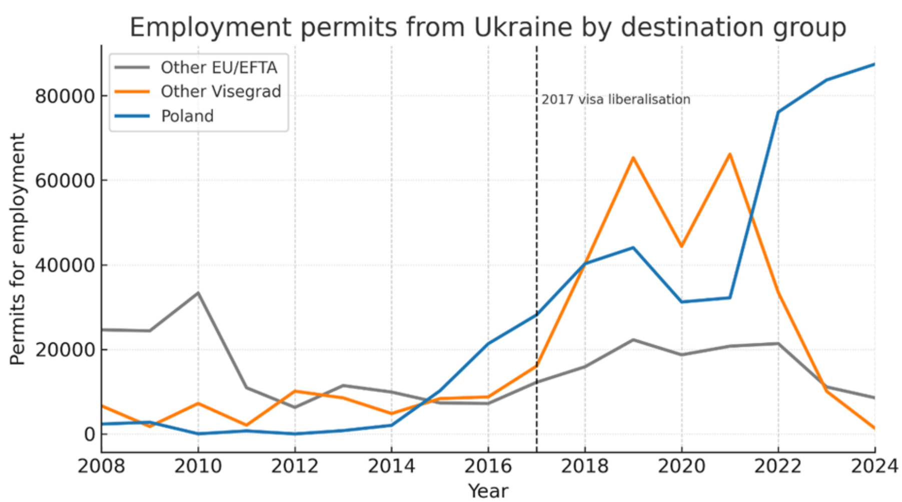
Figure 2.1 First-Time Long-Term Residence Permits Issued by EU-27 Member States to Citizens of Ukraine



Two main forces likely account for this pre-liberalization trend. First, the 2014 annexation of Crimea and the onset of armed conflict in Eastern Ukraine generated a substantial conflict-induced emigration wave, affecting both men and women. This broad push factor increased migration pressure independently of EU policy changes and contributed to rising residency permit numbers during the pre-2017 period. While these conflict dynamics were not uniform across regions, robustness checks excluding regions bordering Russia or located close to conflict areas yield results that are qualitatively unchanged, suggesting that the main findings are not driven by region-specific shocks.

Second, migration policies in neighbouring Visegrad countries, particular Poland, became considerably more permissive for Ukrainian workers beginning in 2014–2015, including the introduction of simplified work-permit procedures. A destination-country decomposition of residence permits confirms that the bulk of the pre-2017 increase is attributable to migration to Poland (Figure 2.2). In contrast, permits issued by other EU and EFTA countries remained relatively flat 2014–2016. After the visa liberalization, however, the composition of destinations changes markedly: residence permits to other Visegrad countries as well as to Western and Northern EU destinations begin to rise sharply, and the overall destination mix becomes substantially more diversified.

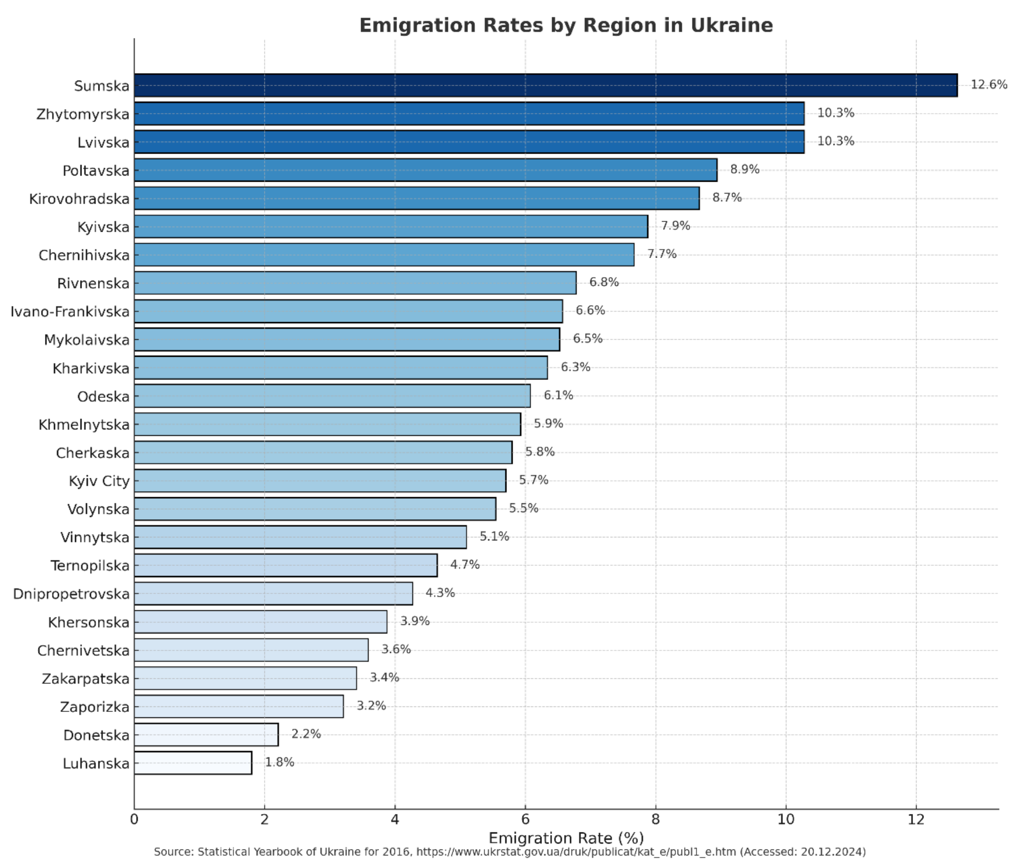
Figure 2.2 Employment Permits Issued by to Citizens of Ukraine by Destination Group



This shift in the geographic composition of migration helps to distinguish pre-existing trends from the effects of the visa reform. While the pre-2017 increase largely reflects conflict-related push factors and country-specific policy changes in Poland, the post-2017 period is characterized by expansion of migration opportunities beyond Poland and into a broader set of EU destinations.

Regions in Ukraine exhibited substantial variation in migration rates prior to the EU visa liberalization. Figure 2.3 displays regional emigration rates before the policy change. We exploit these pre-policy differences by applying a difference-in-differences (DiD) design that compares the responses of applicants from high-emigration regions to those from regions with lower emigration rates. Specifically, regions with emigration rates above the 75th percentile in 2016 serve as the "treated" group, while the remaining regions form the control group. This approach allows us to test whether the impact of the visa liberalization policy on STEM preferences is stronger in areas with greater pre-policy migration exposure, thereby providing more direct evidence that expanded migration opportunities influence educational choices.

Figure 2.3 Emigration Rates by Region in Ukraine, 2016



A growing body of research highlights that STEM degrees, particularly in engineering, computer science, and mathematics, are more transferable across borders than fields such as education, healthcare, or language studies. This is largely due to the global demand for STEM skills, the standardized nature of technical curricula, and the absence of host-country licensing barriers that typically constrain professions like teaching and nursing (Abarcar & Theoharides, 2024; Khanna & Morales, 2017; Rabben, 2013). As a result, students with migration aspirations are more likely to prioritize STEM fields, which offer greater international labour market access and portability of qualifications.

Additionally, we use regional data from the Gallup World Poll—an internationally recognized survey widely used in migration research (e.g., Clemens and Mendola, 2024; Docquier, Peri, and Ruyssen, 2014; Tjaden, Auer, and Laczko, 2019; Guriev, Melnikov, and Zhuravskaya, 2021)—to capture variation in emigration intentions. We restrict the sample to individuals aged 16-25, using survey waves from 2010 to 2017. Emigration intentions are measured based on responses to the question: *"Ideally, if you had the opportunity, would you like to move permanently to another country, or would you prefer to continue living in this country?"* Respondents who answer affirmatively are then asked to specify their preferred destination. This measure captures latent migration preferences, even among individuals without immediate migration prospects.

For each region, we compute the share of young individuals expressing a desire to migrate, disaggregated by gender and by destination (EU or worldwide). We classify a region as *male-dominated* if more than 75 percent of prospective emigrants are male. The maps in Appendix Figure A2.1 show that both overall migration intensity and female-dominated migration intensity are geographically dispersed across regions of Ukraine. High- and low-migration regions are mixed across the west, center, and east, rather than being concentrated in any single part of the country. This spatial pattern supports our research design by demonstrating that the treatment variation does not merely capture broad geographic differences, but reflects genuine cross-sectional variation in pre-policy migration patterns. This classification enables us to test whether migration-driven incentives to pursue STEM fields vary systematically with the gender composition of regional migration intentions. If heightened male migration interest is a key driver of the observed gender divergence in STEM preferences, we would expect the policy's effect to be stronger in male-dominated regions.

2.2.3 Data and Descriptive Statistics

In this study, we use individual-level data on applicants to state-funded bachelor's programs at Ukrainian universities, alongside publicly available information on high school graduates from the Ukrainian Center for Evaluation of the Quality of Education. The data are available through the Center’s website (https://zno.testportal.com.ua/opendata). This dataset includes individual records of scores from the nationwide centralized Independent External Test (IET) and the subjects selected for examinations administered between 2017 and 2020, covering 2,562 schools across Ukraine. Additionally, the dataset contains demographic and institutional information such as gender, date of birth, year of high school graduation, school address, and school type (e.g., vocational or academic). For our analysis, we exclude records of high school graduates from previous years (approximately 5 percent of the sample) who retook the IET to improve their chances of university admission.

Information on university applicants is available through the Abit-poisk website (https://abit-poisk.org.ua/), where users can search for application details by name and admission year. Although the data are publicly accessible, they are fragmented across university programs and years, dispersed over thousands of webpages. To streamline data collection, we developed a web scraping algorithm to automate the process of navigating these webpages and extracting the relevant information. This dataset provides detailed records on applicants' rankings of university programs, IET test results, subject choices, and final admission outcomes.

We construct a unified dataset by merging these two sources based on test results and subject selections, achieving a 25 percent match rate in identifying unique applicants. Because the linkage relies on combinations of non-unique identifiers, a natural concern is that the matched subsample might differ systematically from the full applicant population, potentially affecting the external validity of the analysis. To assess this possibility, we conduct several robustness checks. First, we compare the distribution of IET test scores for matched and unmatched applicants (Figure A2.2). Although the matched group has a slightly higher mean score (approximately one point), the overall distributions are highly similar: medians, interquartile ranges, and the prevalence of unusually high or low scores closely overlap. This suggests that matching success is not driven by extreme performance. Second, we examine differences in subject choices between the two groups (Figure A2.3). The probability of selecting each subject is nearly identical for matched and unmatched applicants, with no systematic pattern indicating that applicants with rare or unusual subject combinations are more likely to be matched. Taken together, these diagnostics provide reassurance that the matched subsample is broadly representative of the overall applicant pool.

We classify fields of study into STEM and non-STEM categories according to the International Standard Classification of Education (ISCED). STEM fields include Natural Sciences, Mathematics, and Statistics (ISCED-05); Information and Communication Technologies (ISCED-06); and Engineering, Manufacturing, and Construction (ISCED-07). However, our classification deviates slightly from ISCED by excluding certain fields, such as Wildlife (0522), Food Processing (0721), and Materials (0722).

While some studies include Dentistry (0911), Medicine (0912), Pharmacy (0916), and Veterinary Science (0841) as part of STEM, we exclude these fields from our analysis. In Ukraine, these programs are typically offered only at the master's level, whereas our study focuses on bachelor's programs. We also exclude Nursing, as it is not classified as tertiary education in Ukraine and does not require standardized IET scores for admission. Prior research (Delaney & Devereux, 2019) has shown that including Nursing tends to reduce the observed gender gap in STEM preferences. Consequently, the gender gap in our analysis may appear larger compared to studies that include Nursing as a STEM field. Figure 2.4 presents the distribution of first-choice preferences across the main STEM fields for both genders combined.

Figure 2.4 Main STEM Fields Indicated as the 1st Preference

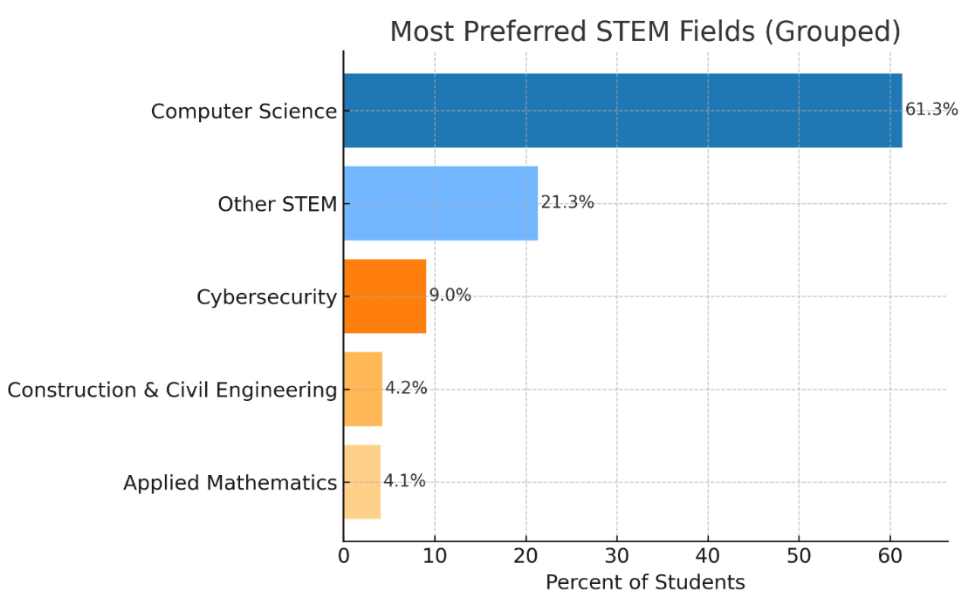


Table 2.1 presents descriptive statistics for the primary variables in our matched sample. Approximately 25 percent of applicants select STEM fields as their first choice, with 15 percent specifically choosing Computer Science. Female applicants constitute 54 percent of the sample. The mean age of applicants is 17.3 years. On average, applicants score 163 out of 200 points on the IET, with an average score of 163 in Ukrainian Language and 155 in Mathematics. Our analysis further shows that 37 percent of male applicants who prioritize STEM fields ultimately enroll in a STEM program, compared to 42 percent of female applicants with the same preference. Among those who do not list a STEM field as their first choice, fewer than 0.5 percent of female applicants and about 2.2 percent of male applicants eventually enroll in a STEM program. Overall, both male and female applicants are unlikely to pursue STEM programs if they do not rank them as their first choice.

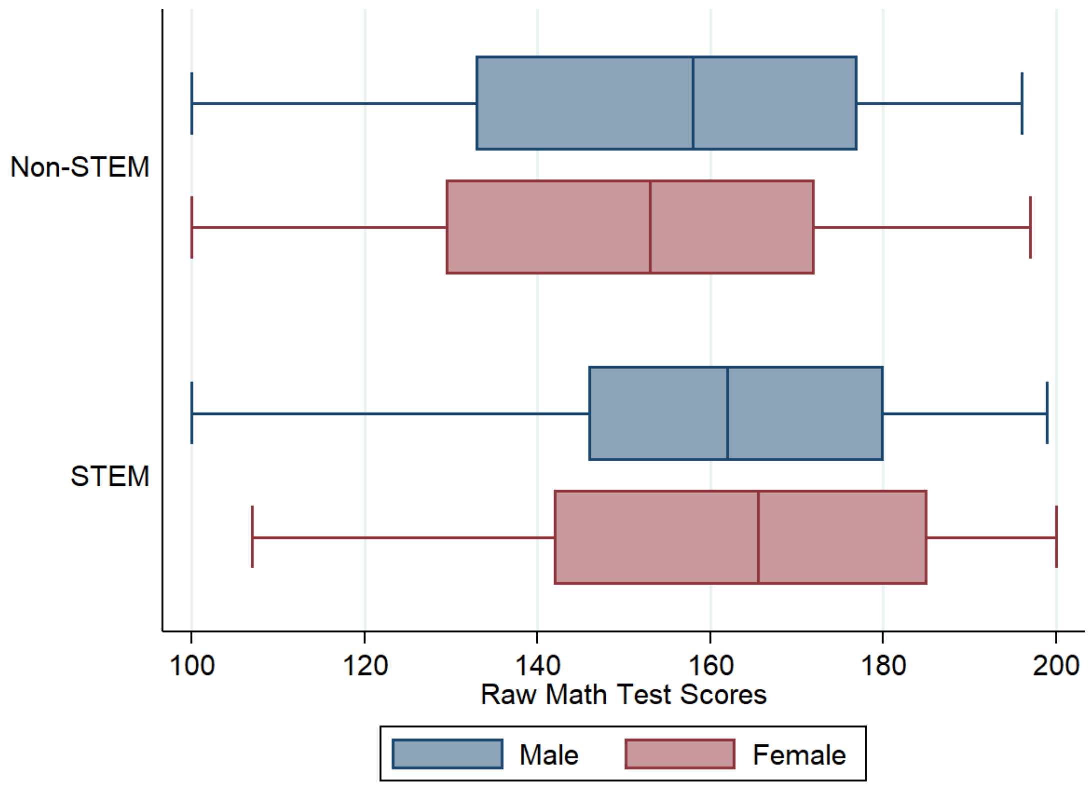
Table 2.1 Descriptive Statistics

Variables	Mean	SD	Min	Max
STEM field ranked as 1st preference	0.252	0.434	0	1
Computer Science ranked as 1st preference	0.171	0.377	0	1
Female	0.542	0.498	0	1
Age	17.29	0.458	17	18
Average student IET score	162.97	19.7	100.062	200
Math test score	155.76	24.79	100	200
Ukrainian test score	163.93	22.12	100	200
Foreign language	159.98	24.73	100	200
Year			2017	2020
N of schools		2,240		
N of regions		23		
N		54,182		

Notes: This table shows the summary statistics of the main variables of our sample. The sample excludes applicants from Crimea, Lugansk and Donetsk regions. The number of observations is lower for math test scores (31,152) and foreign language test scores (29,907) as tests from these subjects are not mandatory. Similarly, the number of observations for computer science preferences (48,910) and other STEM fields preferences (45,254) are lower, as we excluded other STEM fields and computer science when constructing these variables.

Previous research suggests that part of the gender gap in STEM choices may result from differences in performance in STEM-related subjects, particularly mathematics (Aucejo & James, 2016; Speer, 2017). To explore this further, we examine whether there is a gender gap in average math test scores. As shown in Figure 2.5, the median math scores of female applicants who select non-STEM fields as their top preference are slightly lower than those of their male counterparts, suggesting that female applicants, on average, perform similarly to male applicants on math tests. Moreover, among applicants who choose STEM fields, the median math scores of females are slightly higher than those of males. These findings indicate that differences in math test performance do not fully explain the observed disparities in STEM field choices. Consequently, additional factors beyond test performance are likely influencing prospective students' preferences in the context of STEM education.

Figure 2.5 Distribution of Math Test Scores by Fields of Study and Gender



Source: Own calculation based on the data from Abit-poisk website (https://abit-poisk.org.ua/)

Following the implementation of the visa liberalization policy, there was a significant increase in the proportion of male applicants prioritizing STEM fields as their first choice, while the share of female applicants remained relatively stable (see Figure A2.4 in the Appendix). In our matched sample, male applicants were approximately three times more likely than female applicants to select STEM fields as their top preference. This gender gap is primarily driven by choices in Computer Science. As shown in Figure A2.5, fewer than 5 percent of female applicants selected Computer Science as their most preferred field, compared to over 25 percent of male applicants—a gap that widened further following the visa liberalization policy. These findings provide initial evidence of a gendered response to expanded migration opportunities, reflected in applicants' field-of-study preferences.

2.3 Identification Strategy

This section outlines the identification strategy used to estimate the impact of expanded migration opportunities on gender differences in field-of-study choices. We begin with an event study design that compares outcomes for male and female applicants before and after the introduction of the visa liberalization policy. Specifically, we estimate the following linear equation:

$$\boldsymbol{STEM_{ist} = \beta_0 + \beta_1 Female_i + \beta_2 Post_t + \beta_3 Female_i \times Post_t + } \\ \boldsymbol{+ X_{ist} + \theta_s + \varepsilon_{it}} \qquad \textbf{(2.1)}$$

where i indexes students, s denotes schools, and t indexes years. The dependent variable is an indicator equal to one if a student lists a STEM field as their first preference, and zero otherwise. The primary variable of interest is the interaction between the female applicant indicator, denoted as $Female_i$, and the post-policy period indicator, denoted as $Post_t$. The pre-policy period is defined as the 2016-2017 academic year, and the post-policy period spans 2018-2019.

The vector X_{ist} includes individual-level controls, specifically age and average IET performance. We also include school fixed effects, θ_s, to account for unobserved, time-invariant differences across schools that may be correlated with the gender gap in STEM field choice. The coefficient β_3 on the interaction term is the main parameter of interest, capturing the differential effect of the policy change on female applicants relative to male applicants. Standard errors are bootstrapped and clustered at the school level.

While the event study analysis in Equation (2.1) examines whether the gender gap in STEM field choices widened after the policy change, it does not directly attribute this effect to expanded migration opportunities. Other simultaneous policy or economic changes could also explain the observed widening. To address this limitation, we implement a difference-in-differences (DiD) strategy that leverages pre-existing differences in emigration rates across regions. Specifically, we compare applicants from regions with emigration rates above the 75th percentile (high-emigration regions) to those from regions with lower emigration rates. Our DiD specification is given by the following model:

$$\boldsymbol{STEM_{ist} = \delta_0 + \delta_1 Post_t + \delta_2 (Post_t \times Treated\, regions_i) +}$$
$$\boldsymbol{+ X_{ist} + \theta_r + \varepsilon_{it}} \qquad \textbf{(2.2)}$$

In this specification, $Treated\ regions_i$ is a binary indicator equal to one if a student resides in a high-emigration region, and zero otherwise. As before, $Post_t$ identifies the post-policy period (2018–2019), and X_{ist} includes controls for age and IET scores. We also include region fixed effects θ_r to control for time-invariant regional characteristics, and cluster standard errors at the regional level to reflect the level of treatment variation. The coefficient δ_2 captures the differential effect of the policy on applicants from high- versus low-emigration regions. This design allows us to examine whether migration incentives had a stronger effect on students in areas with a higher propensity to migrate, thereby providing more direct evidence on the mechanism behind the observed widening of the gender gap.

The validity of Equation (2.2) relies on the parallel trends assumption: in the absence of the policy change, STEM preferences in treated and control regions would have evolved similarly over time. Although data limitations restrict the analysis to a single pre-policy year, limiting the scope for formal pre-trend tests, the credibility of the design can still be assessed by examining whether treated and control regions are comparable along observable dimensions prior to the reform. To this end, I analyze differences in socioeconomic and demographic characteristics across high- and low-migration regions using individual-level data from the (2010-2016) Gallup World Poll for Ukraine. Respondents are linked to their region of residence and merged with the pre-policy regional emigration classification, allowing comparison of income, age, tertiary education attainment, parental status, and partnership status.

Appendix Figure A2.6 reports difference-in-means estimates for these characteristics, together with 95 percent confidence intervals. Across all outcomes considered, the estimated differences between high- and low-migration regions are small in magnitude and statistically indistinguishable from zero. These findings indicate that, prior to the reform, regions with different emigration intensities were broadly similar in terms of observable socioeconomic and demographic composition. While unobserved factors may still differ across regions, the absence of systematic pre-policy differences in income, education, or family structure helps to mitigate concerns that subsequent divergence in STEM field choices reflects pre-existing regional disparities rather than changes in migration opportunities induced by the reform.

A potential source of bias in both specifications is the 2014 Russian invasion of Ukraine, which may have disproportionately affected male applicants by increasing university enrollment as a means of avoiding military conscription. If these applicants were also more likely to prefer STEM fields, our estimates could be upwardly biased. However, this scenario is unlikely. For such bias to arise, conscription-avoidant males would need both the motivation and the academic ability to enroll in highly competitive STEM programs. Given that students motivated primarily by conscription concerns are more likely to come from the lower end of the academic performance distribution, they are less likely to meet the admission requirements for STEM fields. Instead, they may have gravitated toward non-STEM programs with lower thresholds. Therefore, any bias from this channel is likely attenuated toward zero, suggesting that our estimates may reflect a lower bound of the true effect. To further address this concern, we exclude from the analysis regions directly affected by Russian aggression. The results remain robust to this exclusion.

2.4 Results

2.4.1 Main Results

Table 2.2 presents the event study and difference-in-differences (DiD) estimates of the impact of expanded migration opportunities—introduced through the EU visa liberalization policy—on applicants' preferences for STEM fields.

Column 1 reports estimates from Equation (2.1) using applicant-level data from 2017 to 2020[20]. The coefficient on the female indicator shows that, prior to the policy reform, female applicants were 27.0 percentage points less likely than male applicants to list a STEM field as their first-choice preference. The After coefficient indicates that, following the policy reform, male applicants (the reference group) increased their likelihood of choosing a STEM field by 1.4 percentage points, a statistically significant but modest effect. The Female × After interaction term is negative and statistically significant (−0.033), indicating that the gender gap in STEM

[20] We re-estimate the analysis excluding the 2019-2020 academic year, during which applicants' preferences may have been affected by the COVID-19 pandemic. The results are available upon request.

preferences widened after the policy by an additional 3.3 percentage points. This implies that the relative increase in STEM preferences was smaller for female applicants. When expressed as a percentage of the pre-policy gender gap (0.033 / 0.270 ≈ 12.2%), this suggests that the gap widened substantially. These results indicate that the visa liberalization disproportionately increased male applicants' interest in STEM fields, suggesting a gendered behavioral response to expanded migration opportunities.

Table 2.2 The Effect of Migration Opportunities on Gender Gap in STEM Fields

Outcome = STEM field ranked 1st preference	(1) All applicants	(2) All applicants	(3) Male Applicants	(4) Female Applicants
Female	-0.270***	-0.309***		
	(0.007)	(0.010)		
After	0.014**	-0.011**	-0.005	-0.014***
	(0.007)	(0.004)	(0.007)	(0.004)
Female × After	-0.033***			
	(0.008)			
After × Treated regions		0.020***	0.032**	0.007
		(0.008)	(0.016)	(0.009)
Student average test score	0.002***	0.002***	0.003***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
School FE	X			
Region FE		X	X	X
Observations	54,182	54,182	24,793	29,389
R^2	0.112	0.127	0.020	0.007

Notes: This table presents the results from estimating Equation 2.1 (Column 1) and Equation 2.2 (Columns 2-4) using DID regression. The analysis excludes applicants to the undergraduate programs in Lugansk and Donetsk regions. Outcome variable is a dummy variable equal to one if the student ranked STEM field as the first priority and zero otherwise. Treated regions are those with emigration rates above the 75th percentile in 2016. Models control for applicants' age, average test score on IET centralized exams, school (Column 1) and region fixed effects (Columns 2-4). Bootstrapped (1000reps.) standard errors in parentheses are clustered at the school (Column 1) and region (Columns 2-4) level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Columns 2–4 present DiD estimates from Equation (2.2), which incorporates regional heterogeneity by interacting the post-policy period with an indicator for high-emigration regions, i.e., those with emigration rates above the 75th percentile in 2016. The After coefficients indicate a general decline in stated interest in STEM fields following the reform across the applicant pool, with the decrease driven primarily by female applicants (Column 4: −1.4 percentage points, statistically significant). This aggregate decline should not be interpreted as a reduction in the perceived value of STEM education per se. Instead, it is consistent with the presence of capacity constraints in Ukrainian university STEM programs. Even though visa liberalization increased the expected returns to STEM education, particularly through improved labor-market opportunities abroad, the number of available places in STEM programs remained limited.

Against this backdrop, the positive and statistically significant interaction term (After × Treated regions) in Columns 2 and 3 indicates that applicants from high-emigration regions were more likely to shift toward STEM fields following the reform. This response is especially pronounced among male applicants (Column 3: +3.2 percentage points), while the corresponding effect for female applicants is small and statistically insignificant (Column 4: +0.7 percentage points). One plausible interpretation is that the reform disproportionately raised the expected returns to STEM education for men, leading male applicants, particularly in high-emigration regions, to respond more strongly by prioritizing STEM fields. Given fixed program capacity, this intensified male demand likely crowded out some female applicants at the margin, generating the observed decline in female STEM interest despite unchanged or increasing incentives. Overall, these results suggest that migration opportunities can reshape educational preferences in gender-specific ways, mediated by institutional constraints in higher education supply.

We next examine whether the observed increase in the gender gap in STEM field preferences following the visa liberalization policy can be attributed to a stronger interest in international migration among male applicants relative to their female peers. Specifically, we test whether the policy's impact on STEM preferences is larger in regions where prospective emigrants are disproportionately male, compared to regions with more gender-balanced migration intentions. If heightened male migration interest is a key driver, we would expect the policy's effect to be more pronounced in male-dominated regions.

The results of this analysis, presented in Table 2.3, show that the impact of the visa liberalization policy on the gender gap in STEM field preferences is approximately 34 percent larger in regions where prospective emigrants are predominantly male compared to more gender-balanced regions (Columns 1 and 2; calculated as (0.039/0.029 – 1)×100)). In Columns 3 and 4, we repeat the analysis focusing specifically on migration intentions toward EU countries only. The results remain broadly consistent, reinforcing the conclusion that expanded migration opportunities had a stronger effect in regions with higher male migration interest.

Table 2.3. The Effect of Migration Opportunities on Gender Gap in STEM Fields by Migration Wishes

	Migration Intentions		**Migration Intentions to EU Countries**	
	Male-dominated	Female-dominated	Male-dominated	Female-dominated
	(1)	(2)	(3)	(4)
Female	-0.269***	-0.258***	-0.268***	-0.259***
	(0.012)	(0.008)	(0.012)	(0.008)
After	0.037***	0.014*	0.037***	0.015*
	(0.012)	(0.008)	(0.013)	(0.008)
Female × After	-0.039***	-0.029***	-0.040***	-0.029***
	(0.014)	(0.010)	(0.015)	(0.009)
Observations	17373	37669	15739	39303
R^2	0.118	0.103	0.118	0.103

Notes: This table presents the results from estimating Equation 2.1. The analysis excludes applicants to the undergraduate programs in Lugansk and Donetsk regions. In columns 1 and 2, we include applicants from male- and female-dominated regions in terms of migration intentions to any destination, respectively. Male-dominated regions are regions where the share of male population who intend to migrate is above the 75 percentile. In columns 3 and 4, we define male-dominated regions as those where the share of male population who intend to migrate to the EU is above the 75 percentile. In all specifications, the outcome variable is an indicator for ranking a STEM field as the first priority and zero otherwise. Models control for applicants' age, average test scores, and school fixed effects. Standard errors in parentheses are clustered at the school level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Overall, our findings consistently show that expanded international migration opportunities, introduced through the 2017 EU visa liberalization policy, widened the gender gap in STEM field preferences among Ukrainian university applicants. The increase in the gap is primarily driven by a stronger shift toward STEM among male students, while female preferences experienced modest decline. The effects are more pronounced in regions with higher historical emigration rates and in areas where migration intentions are disproportionately male, providing further evidence that gender-specific migration incentives shape educational choices.

2.4.2 Heterogeneity Analysis and Robustness Checks

We further examine which specific STEM fields are driving the overall effects by disaggregating the analysis into computer science and other STEM disciplines. Computer science is the most popular STEM field in our sample, accounting for approximately 61 percent of all first-choice STEM preferences. As shown in Table 2.4, the gender gap in selecting computer science as a first-choice field is nearly twice as large as the gap observed for other STEM fields. Following the visa liberalization policy, the gender gap in ranking computer science as the top preference widened by approximately 2.4 percentage points (Panel A, Column 1), representing a 10.2 percent increase relative to the pre-policy gap (0.024/0.235). In comparison, the gender gap for other STEM fields widened by 1.8 percentage points, or 18.0 percent relative to the pre-policy gap (0.018/0.100).

Table 2.4 The Effect of Migration Opportunities on Gender Gap in Computer Science and Other STEM Fields

Panel A Computer Science	(1) All applicants	(2) All applicants	(3) Male Applicants	(4) Female Applicants
Female	-0.235^{***}	-0.264^{***}		
	(0.006)	(0.010)		
After	0.013^{*}	-0.007^{*}	-0.005	-0.008^{**}
	(0.007)	(0.004)	(0.006)	(0.004)
Female × After	-0.024^{***}			
	(0.008)			
After × Treated regions		0.016^{***}	0.034^{**}	0.004
		(0.006)	(0.015)	(0.005)
School FE	X			
Region FE		X	X	X
Observations	48,910	48,910	21,179	27,731
R^2	0.113	0.130	0.037	0.015

Panel B Other STEM fields	(1) All applicants	(2) All applicants	(3) Male Applicants	(4) Female Applicants
Female	-0.100^{***}	-0.120^{***}		
	(0.006)	(0.007)		
After	0.006	-0.009^{**}	-0.003	-0.013^{***}
	(0.006)	(0.004)	(0.007)	(0.005)
Female × After	-0.018^{**}			
	(0.007)			
After × Treated regions		0.011	0.016	0.009
		(0.010)	(0.015)	(0.009)
School FE	X			
Region FE		X	X	X

Observations	45,254	45,254	17,677	27,577
R^2	0.033	0.036	0.001	0.001

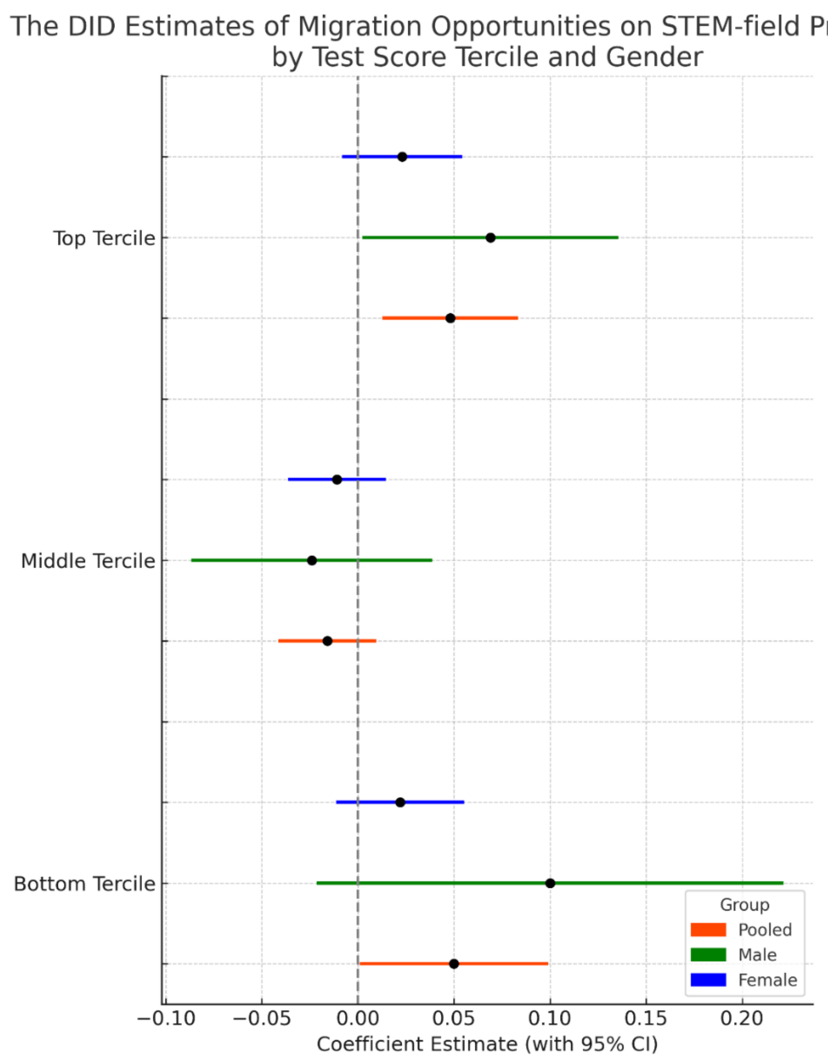
Notes: This table presents the results from estimating Equation 2.1 (Column 1) and Equation 2.2 (Columns 2-4) using DID regression. The analysis excludes applicants to the undergraduate programs in Lugansk and Donetsk regions. Outcome variable is a dummy variable equal to one if the student ranked computer science (Panel A) and other STEM fields (Panel B) as the first priority and zero otherwise. The number of observations differs from Table 2, as we exclude other non-STEM fields from Panel A and computer science from panel B. Treated regions are regions where the share of population who migrated in 2016 is above the 75 percentile. Models control for applicants' age, average test score on IET centralized exams, and school fixed effects. Bootstrapped (1000reps.) standard errors in parentheses are clustered at the school (Column 1) and region (Columns 2-4) level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Columns 2–4 of Table 2.4 report Difference-in-Differences estimates based on Equation (2.2), distinguishing between preferences for computer science (Panel A) and other STEM fields (Panel B). In Panel A, Column 3, the interaction between the post-reform period and high-emigration regions is positive and statistically significant. This indicates that male applicants from areas with higher pre-reform emigration rates became more likely to prioritize computer science following the policy change. By contrast, the corresponding interaction in Column 4 for female applicants is small and not statistically significant, suggesting no comparable shift in preferences.

For other STEM fields (Panel B), the interaction term in the full sample (Column 2) is positive but lacks statistical significance, indicating no clear difference between applicants from high- and low-emigration regions. A similar pattern holds in Columns 3 and 4, where the interaction terms for male and female subsamples are again small and estimated with considerable imprecision. On average, applicants show a modest decline in interest in other STEM fields relative to non-STEM options following the reform. This overall reduction appears to be largely driven by female applicants, who became 1.3 percentage points less likely to select these fields as their first-choice preference.

Finally, we examine whether the impact of migration opportunities on gender differences in STEM field preferences varies across the distribution of academic performance. Figure 2.4 presents DID estimates disaggregated by terciles of IET test scores. We find that the visa liberalization policy had the largest effect on applicants from regions with higher pre-policy emigration rates in the bottom and top terciles, with increases of 4.8 and 5.0 percentage points. Specifically, low- and high-performing male applicants from the regions with higher pre-policy emigration rates drive the results and are 6.9 and 10.0 pp more likely to rank STEM fields as the most preferred. Female applicants from the regions with higher pre-policy emigration rates are not responsive to the expansion of migration opportunities. Middle-performing students—regardless of gender—appear unaffected in terms of their STEM preferences. The findings suggest that expanded migration opportunities can influence educational preferences, particularly nudging lower-performing and high-performing male students toward more economically promising STEM fields.

Figure 2.4. The Effect of Migration Opportunities on Gender Differences in Ranking STEM First by Student Test Score Tercile



Notes: This figure presents the estimated effects of the 2017 EU visa liberalization on the probability of ranking a STEM field as the top study preference, using a difference-in-differences framework. Estimates are shown separately for students in the top, middle, and bottom test score terciles, and by gender (male, female, pooled). The results indicate that the largest effect is observed among low-achieving male students, while the response among female students is smaller and statistically insignificant across all terciles.

2.5 Conclusion

Using a natural experiment provided by the 2017 EU visa liberalization policy for Ukrainian citizens, this study examines the impact of expanded international migration opportunities on gender disparities in STEM field choices. Our analysis of comprehensive administrative data on university and field-of-study application preferences reveals that, prior to the policy change, female applicants were significantly less likely than their male counterparts to select STEM fields as their top choice. Following the reform, the gender gap in STEM fields widened by approximately 12.2 percentage points in stated preferences. These shifts are driven primarily by a stronger increase in STEM preferences among male applicants—especially those with low academic achievement—while female applicants' preferences remain largely unchanged. Importantly, we find no evidence that the policy discouraged high-achieving female students from pursuing STEM.

Moreover, our investigation demonstrates that the effect of expanded migration opportunities is not uniform across regions or academic performance levels. In regions with higher pre-policy emigration and among students in the lower terciles of the test score distribution, the policy's impact on the gender gap in STEM preferences is more pronounced. This suggests that migration-driven incentives may be altering the composition of the applicant pool by attracting more male students from specific performance segments, rather than causing a universal shift in the behavior of all male applicants.

These findings carry significant policy implications. As countries continue to face shortages in skilled STEM workers, it is important to recognize that migration policies—while effective in increasing overall interest in STEM—may unintentionally widen existing gender disparities. In particular, our results suggest that male students are more responsive to international migration incentives than their female peers, potentially reinforcing gender gaps in STEM participation. Policymakers aiming to expand the STEM talent pool should therefore consider targeted interventions—such as mentorship programs, scholarships, or information campaigns specifically designed to support and encourage female students—to ensure that policies attracting internationally mobile talent do not deepen structural inequalities in educational choices.

2.A Appendix

Figure A2.1 Migration Intention by Region

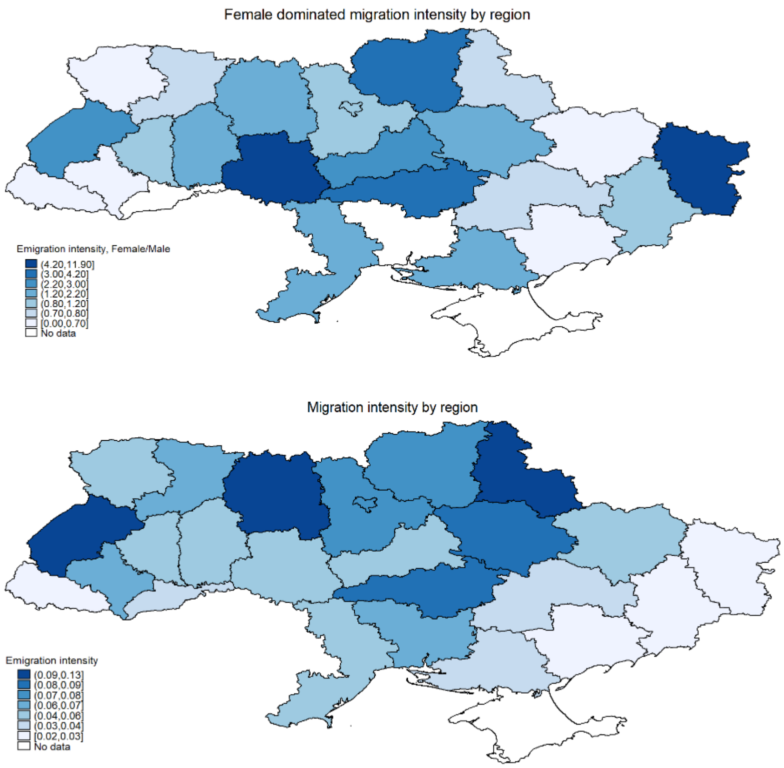
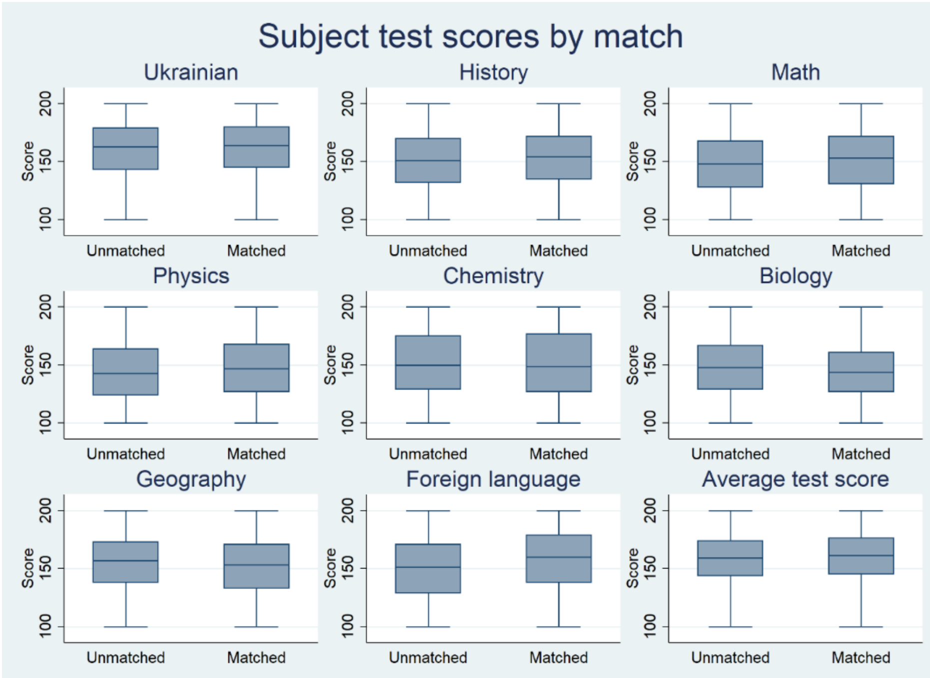


Figure A2.2 Distribution of Test Scores by Subject and Match



Notes: The figure compares mean test scores between matched and unmatched applicants. While the matched group has slightly higher scores (by about one point), the small difference suggests that the matched sample is broadly representative of the full applicant pool.

Figure A2.3 Probability of Test Subject by Match

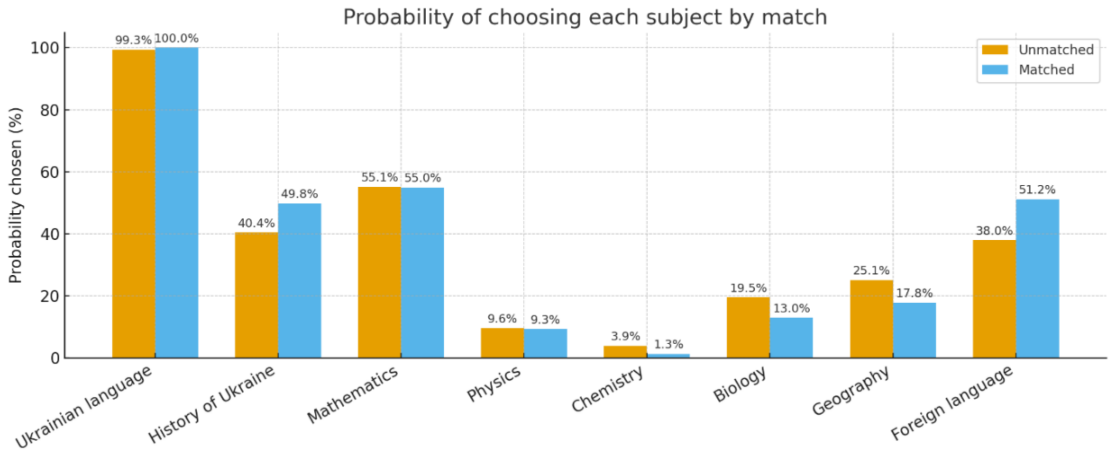
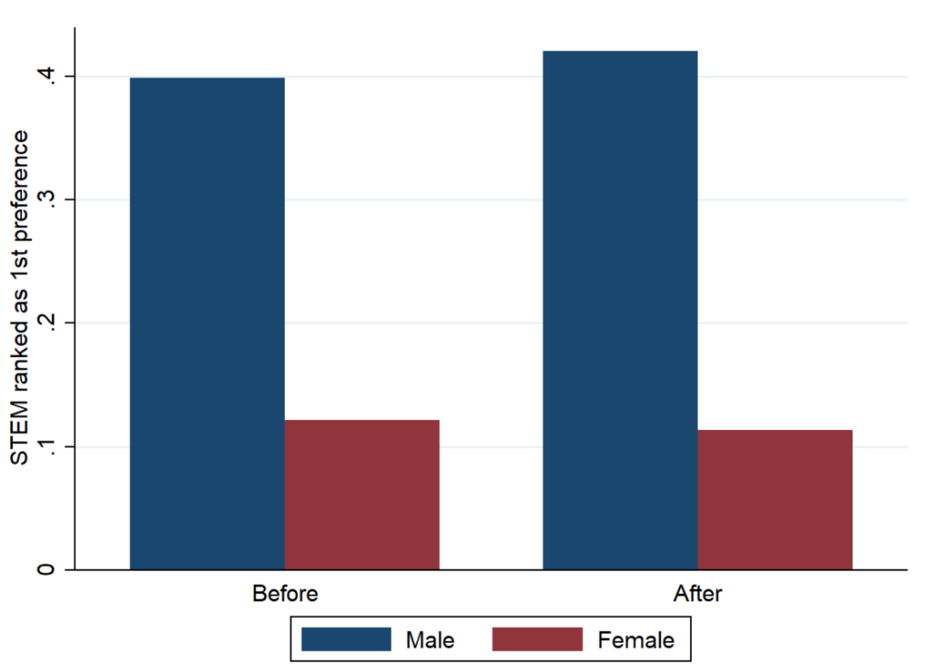
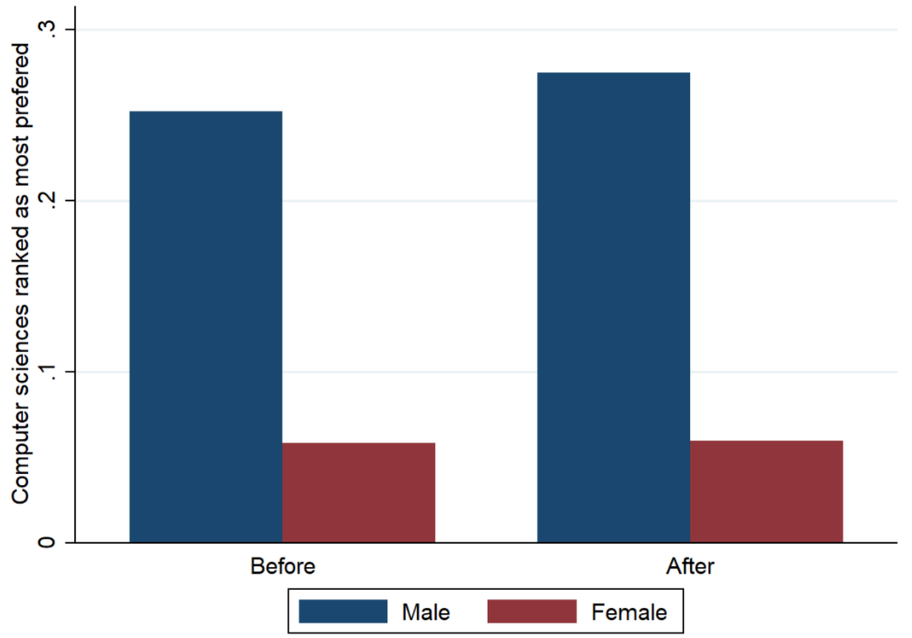


Figure A2.4 STEM Field of Study by Gender



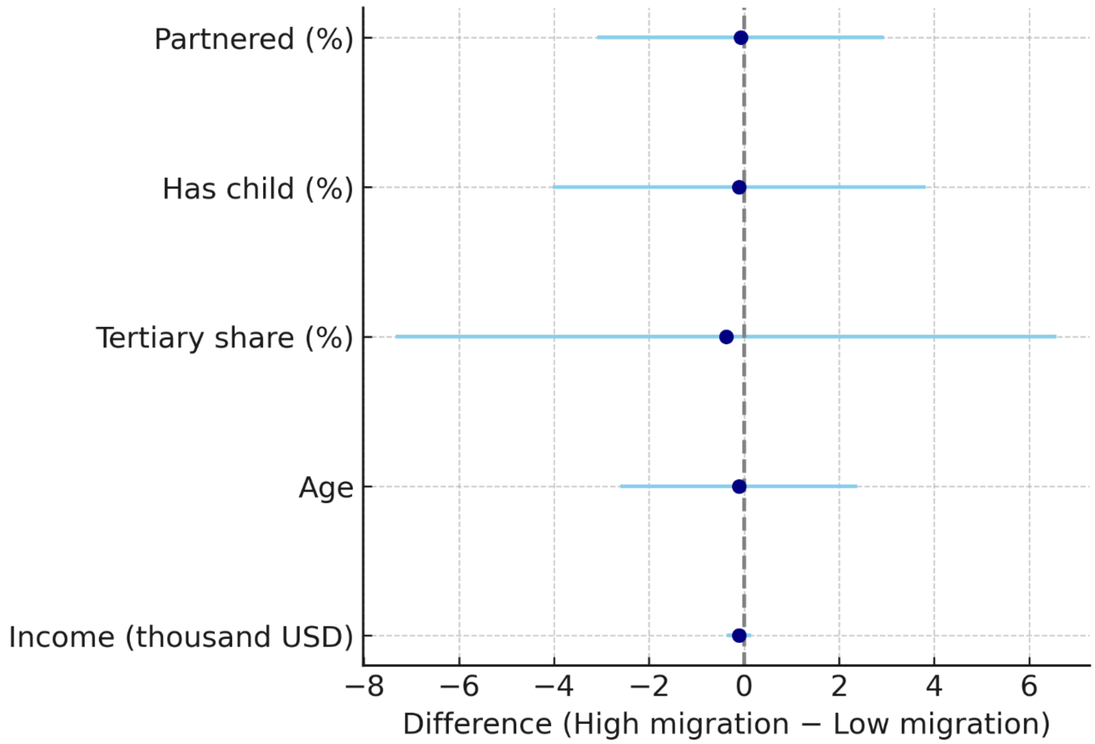
Notes: The figure shows the share of male and female applicants ranking a STEM field as their first preference, before and after the 2017 EU visa liberalization for Ukrainian citizens. The gender gap widens after the policy change, driven by an increase in STEM interest among male applicants, while female preferences remain largely unchanged.

Figure A2.5 Computer Sciences as the Most Preferred Field of Study by Gender



Notes: The figure displays the share of male and female applicants who ranked computer science as their most preferred field of study, before and after the 2017 EU visa liberalization. Male interest in computer science increased following the reform, while female preferences remained stable, contributing to a widening gender gap in this specific STEM field.

Figure A2.6 Balance test: Socioeconomic Differences across High- and Low- Emigration Regions of Ukraine



3 The Hidden Bias in Class Rank: Simulating Measurement Error in STAR Data

3.1 Introduction

A growing body of research emphasizes that students' ordinal rank in the classroom, i.e., how they perform relative to their peers, can have important and lasting effects on their educational and labour market trajectories (Delaney & Devereux, 2022). Even among students with similar levels of absolute achievement, those who are ranked higher relative to their classmates tend to exhibit greater confidence, motivation, and ambition, which in turn shape the goals they pursue and the effort they exert. These so-called "rank effects" are believed to operate through multiple channels, including changes in self-concept, teacher feedback, and parental expectations (Elsner & Isphording, 2017; Gill et al., 2019). Prior studies suggest that being ranked highly among one's peers increases the likelihood of persisting in school, selecting more demanding academic tracks, and aiming for more competitive postsecondary pathways (Murphy & Weinhardt, 2020; Denning, Murphy, & Weinhardt, 2023; Elsner, Isphording, & Zölitz, 2021).

Despite growing interest in this topic, studying rank effects empirically poses significant challenges. Classroom rank is a relative construct: a student's position depends not only on their own score but also on the distribution of scores among their peers. In many administrative or survey-based datasets, peer achievement information is only partially observed, making it difficult to calculate rank accurately. This introduces a form of non-classical measurement error that may bias estimated rank effects and obscure key forms of heterogeneity. While some work has addressed bias due to incomplete peer data in the context of average peer effects (Ammermueller & Pischke, 2009; Micklewright, Schnepf, & Silva, 2012; Sojourner, 2013), the implications for the growing literature on ordinal rank remain underexplored.

This paper addresses these limitations by studying the long-term academic effects of ordinal rank in kindergarten, with particular attention to the consequences of measuring rank using incomplete peer data. I use data from the Tennessee Student/Teacher Achievement Ratio (Project STAR) experiment, a large-scale randomized controlled trial in which students and teachers were randomly assigned to classrooms from kindergarten through third grade. The STAR dataset is well-suited to this analysis for several reasons. First, random peer assignment enables causal identification of rank effects conditional on test scores. Second, baseline test scores are measured at the start of kindergarten—before students accumulate meaningful schooling exposure—allowing for early rank measures that are unlikely to reflect prior tracking or self-selection. Third, these data are linked to long-run outcomes, including high school GPA, graduation, college entrance exam participation, and ACT scores. Finally, approximately 88 percent of STAR kindergarten classrooms have complete peer test score data, enabling me to calculate “true” classroom rank and simulate rank mismeasurement under varying levels of peer observability.

The analysis proceeds in two parts. First, I estimate the effects of kindergarten ordinal rank—separately for math and reading—on students’ later academic outcomes. The results reveal a strongly nonlinear relationship: in the subsample of classes with complete peer test score data, students ranked near the top of their kindergarten classrooms experience substantial gains in GPA, test participation, and ACT scores, with no detectable impact on high school graduation. These gains are concentrated in the top three deciles of the classroom rank distribution and are not evident for students below the median. The finding is consistent with evidence from other settings: for instance, a recent study by Dadgar (2021) in Sweden found that only students at the extremes of the rank distribution experienced long-run effects (positive for those near the top and adverse for those near the bottom), with no measurable effect for the middle ranks.

Second, I examine how incomplete peer data affect the estimation of these effects by simulating varying levels of peer test score observability under a missing completely at random (MCAR) assumption. As expected, the reduction in peer observability attenuates the estimated effects. Crucially, this attenuation is not uniform: the bias is most severe at the top of the rank distribution—precisely where the true effects are strongest. This alignment between bias severity and effect heterogeneity suggests that studies using incomplete peer data may systematically understate the benefits of being top-ranked. In this sense, measurement error is not merely a statistical issue but has substantive implications for understanding how relative position in the classroom shapes students’ long-term outcomes.

This paper contributes to the literature in two important ways. First, it provides new evidence on the long-run consequences of ordinal rank measured at the very start of formal schooling. Whereas most prior work on rank effects has focused on middle or high school students (Denning, Murphy, & Weinhardt, 2023; Elsner & Isphording, 2017; Murphy & Weinhardt, 2020), only a few recent studies (Carneiro et al., 2025; Rury, 2025) examine the effect of ordinal rank in early-childhood settings. This study demonstrates that rank effects emerge much earlier and persist into high school. Moreover, these effects are highly nonlinear and concentrated among the highest-ranked students. While earlier studies, such as Rury (2025), focus primarily on average treatment effects, the results here reveal substantial heterogeneity, showing that the gains from high rank are concentrated at the top of the distribution. Second, the paper provides new methodological insights by quantifying how incomplete peer data generate bias in rank-based estimates, particularly in the part of the distribution where effects are largest. By linking measurement error to the underlying shape of treatment effects, this study highlights a critical limitation of rank-based analyses using partial administrative or survey data.

The remainder of this chapter is organized as follows. Section 3.2 introduces the STAR dataset and describes the construction of key variables, including the measure of ordinal rank. Section 3.3 outlines the empirical strategy, including the identification approach and simulation framework. Section 3.4 presents the main results under full and partial peer observability. Section 3.5 concludes.

3.2 Data Description

I use data from the Tennessee Student/Teacher Achievement Ratio (Project STAR), a pioneering randomized controlled trial launched during the 1985–86 academic year to investigate how class size influences student achievement. The study randomly assigned 6,325 kindergarten students across 79 public schools in Tennessee to one of three classroom types: a small class with 13 to 17 students, a regular-sized class with 22 to 25 students, and a regular-sized class with the addition of a full-time teacher's aide. Participating schools implemented all three conditions and retained students in their assigned classrooms through third grade, after which all students returned to standard class settings.

Project STAR has since become a foundational dataset in education research, widely used to examine questions ranging from peer to teacher effects (Bietenbeck, 2020; Dee, 2004). It contains detailed student-level information, including standardized test scores, demographic characteristics, classroom and school identifiers, and a set of long-term outcomes obtained via administrative record linkages, including high school GPA, graduation status, participation in college entrance exams (SAT/ACT), and postsecondary enrolment (Krueger & Whitmore, 2001; Chetty et al., 2011).

Project STAR is uniquely suited for studying the causal impact of ordinal rank for several reasons. First, students were randomly assigned to classrooms, eliminating concerns about endogenous sorting of students or teachers into specific peer environments. This random assignment creates exogenous variation in peer composition—crucial for identification of rank effects conditional on ability. Second, the dataset contains standardized test scores in kindergarten, allowing us to measure classroom rank at a very early stage in a student's academic trajectory. This is a major advantage over studies that must rely on later achievement as a proxy for prior ability, which risks conflating the effects of rank with prior exposure to rank itself. Finally, STAR follows students over a long-time horizon, making it possible to link early classroom experiences to later-life academic outcomes, including high school graduation, high school GPA, SAT or ACT participation, and ACT test score[21].

Table 3.1 presents summary statistics for three linked analytical samples drawn from the Project STAR dataset: students with observed test scores in kindergarten, those with available data in primary school, and students for whom high school outcomes are recorded. The kindergarten sample is slightly more male (48.7%), but the gender balance shifts over time, with females comprising 55.2% of the primary school sample and 53.7% of the high school sample. The racial composition of the sample also changes across stages: while approximately 68% of students are white and 32% are Black in kindergarten, these proportions shift to 78.4% white and 21.5% Black by high school, suggesting differential attrition or follow-up. Socioeconomic composition changes in a similar fashion. Nearly half of the kindergarten sample (48.3%) qualifies for free lunch, but this rate declines to 39.5% in primary school and 34.2% in high school, indicating gradual upward selection in observed samples over time.

Table 3.1 Summary Statistics – Project STAR Sample

Variable	Kindergarten Sample	Primary School Sample	High School Sample
Math Score (KG)	485.4	498.7	495.9
	(47.7)	(44.75)	(45.58)
Percentile Rank (KG)	0.499	0.581	0.556

[21] The attrition rate in the STAR data is high; we can observe the high school outcomes for only 39% of the kindergarten sample. Rury (2025) argues that the attrition does not drive the results of relative rank on long-term outcomes.

	(0.301)	(0.281)	(0.293)
Gender (Female=1)	0.487	0.552	0.537
	(0.50)	(0.497)	(0.498)
White	0.676	0.718	0.784
	(0.468)	(0.449)	(0.411)
Black	0.324	0.281	0.215
	(0.468)	(0.449)	(0.411)
Free-Lunch Status	0.483	0.395	0.342
	(0.499)	(0.489)	(0.474)
Observations	5,822	3,234	2,274

Notes: The three columns in this table correspond to distinct analytical samples: column (1) includes students with observed test scores in kindergarten; column (2) includes those with observed test scores in primary school; and column (3) includes students for whom high school outcomes are available. Standard deviations are reported in parentheses. Each column displays the mean of the corresponding variable within that sample. These variables comprise the key covariates used in the main estimation for each outcome set. While the number of observations may not exactly match the total shown in the bottom row of each sample, the figures provide a close approximation of the data used in the empirical analysis.

Mean standardized math scores increase slightly between kindergarten (485.4) and primary school (498.7), before stabilizing at 495.9 in the high school sample. Percentile rank also rises from an initial average of 0.499 in kindergarten to 0.581 in primary school, suggesting that students retained in the later outcome datasets tend to be relatively better performing. On average, students in the sample had a GPA of 3.34 and a graduation rate of 83.8%. Approximately one-third of students in the sample took either the SAT or ACT, with an average composite score of 19 on the ACT's 35-point scale.

Table 3.2 presents the results of balance tests examining whether observable student characteristics are systematically related to their kindergarten classroom rank in reading and math. Each column reports estimates from a regression of classroom rank on students' demographic characteristics, classroom fixed effects, and kindergarten test scores, as well as interactions between individual scores and both the mean and variance of classroom-level ability.

Table 3.2. Balance Test

Outcome = Kindergarten Rank	(1) Math	(2) Reading
Female	0.001	0.032***
	(0.002)	(0.006)
Black	-0.005	-0.006
	(0.004)	(0.014)
Free-lunch Status	-0.001	-0.070***
	(0.002)	(0.008)
N	5,822	5,737

Notes: Each column reports estimates from a separate regression assessing the balance of baseline covariates with respect to kindergarten classroom rank. Column (1) regresses the listed control variables on students' classroom-specific math rank in kindergarten, while Column (2) does the same for reading rank. All models include controls for student gender (female), race (Black), free or reduced-price lunch status, classroom fixed effects, and kindergarten test scores. To account for classroom composition, the models also include interactions between each student's test score and both the mean and variance of test scores within their classroom. Robust standard errors are reported in parentheses.

Column 1 Table 3.2 indicates that students' socio-demographic characteristics—including gender, race, and free or reduced lunch status—are not systematically associated with their relative rank in kindergarten math, suggesting a good balance on observables. In contrast, Column 2 reveals that female students tend to have higher relative ranks in reading, while students eligible for free lunch are more likely to be ranked lower. To account for these modest associations and to rule out confounding effects, all subsequent analyses include controls for student gender and other individual- and classroom-level baseline characteristics.

3.3 Identification Strategy

3.3.1 Rank Variation and Identification

To identify the long-run effects of a student's ordinal position in the classroom, I build on the framework developed by Denning, Murphy, and Weinhardt (2023), who emphasize that differences in rank conditional on test scores arise due to random variation in the composition of peer groups. Because rank is inherently relative, two students with the same test score may end up in very different positions depending on the ability distribution of their classmates. A high-scoring student surrounded by equally strong peers may be ranked lower than a similarly performing student placed in a less competitive classroom.

I estimate the effect of percentile rank on long-run academic outcomes following Denning, Murphy, and Weinhardt (2023):

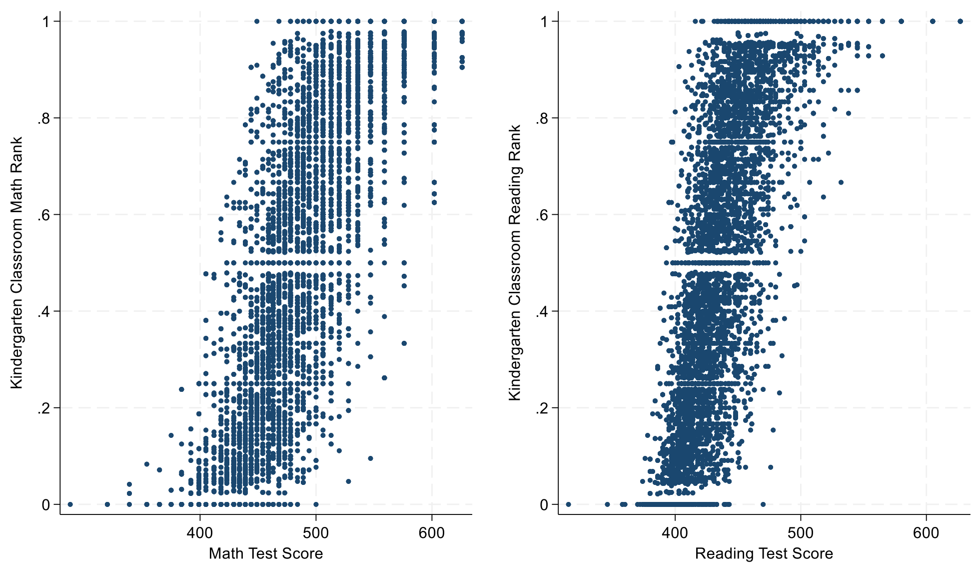
$$\boldsymbol{Y_{isc}} = \sum_{r=1,r\neq 5}^{10} \boldsymbol{I_n(R_{isc} = r)\,\beta_r} + \sum_{d=1}^{16} \boldsymbol{I_d(D_s = D)} \sum_{t=1,t\neq 10}^{10} \boldsymbol{I_t(T_{isc} = t)\delta_{dt}} + \\ + \boldsymbol{\theta_{sc}} + \boldsymbol{X_{isc}} + \boldsymbol{\varepsilon_{isc}} \qquad \textbf{(3.1)}$$

Here, Y_{isc} denotes the outcome of interest for student i in school s, classroom c (e.g., SAT-taking), and $I_r(R_{isc} = r)$ is an indicator for whether the student falls into percentile rank decile r, with the fifth decile omitted as the reference category. A key variable in this model is the student's percentile rank, which I calculate based on their position in the classroom test score distribution:

$$\boldsymbol{R_{isc}} = \frac{\boldsymbol{p_{isc}-1}}{\boldsymbol{N_{isc}-1}} \qquad \textbf{(3.2)}$$

where $\boldsymbol{p_{isc}}$ is the student's ordinal position based on test scores within their classroom, and $\boldsymbol{N_{isc}}$ is the total number of students in that classroom. The percentile rank measure ranges from 0 (lowest rank) to 1 (highest rank). In cases of tied scores, I assign the average rank among the tied students. Figure 1 depicts how this classroom rank varies in kindergarten, separately for reading and math, conditional on students' test performance. A key feature shown in these figures is that students with comparable test scores often end up with different ranks, depending on the distribution of peer scores within their classroom. This within-score variation in classroom rank independent of absolute achievement provides the central source of identifying variation for estimating the causal impact of classroom rank on later academic outcomes.

Figure 3.1 Kindergarten Rank and Test Scores



Notes: This figure plots students' kindergarten raw math (left) and reading (right) scores against where they rank in terms of math and reading scores in their kindergarten classroom.

To control for baseline academic ability, I follow Denning, Murphy, and Weinhardt (2023) and use deciles of kindergarten test scores $\boldsymbol{T_{isc}}$, interacted with indicators for classroom types $\boldsymbol{D_s}$, which are defined by the quartiles of the classroom-level mean and variance in test scores (yielding 16 types). This interaction allows the relationship between prior test scores and outcomes to vary flexibly across classroom contexts. I also control for classroom fixed effects $(\boldsymbol{\theta_{sc}})$ and student-level characteristics $(\boldsymbol{X_{isc}})$, including gender, race/ethnicity, age, and free lunch eligibility. Standard errors are clustered at the school level.

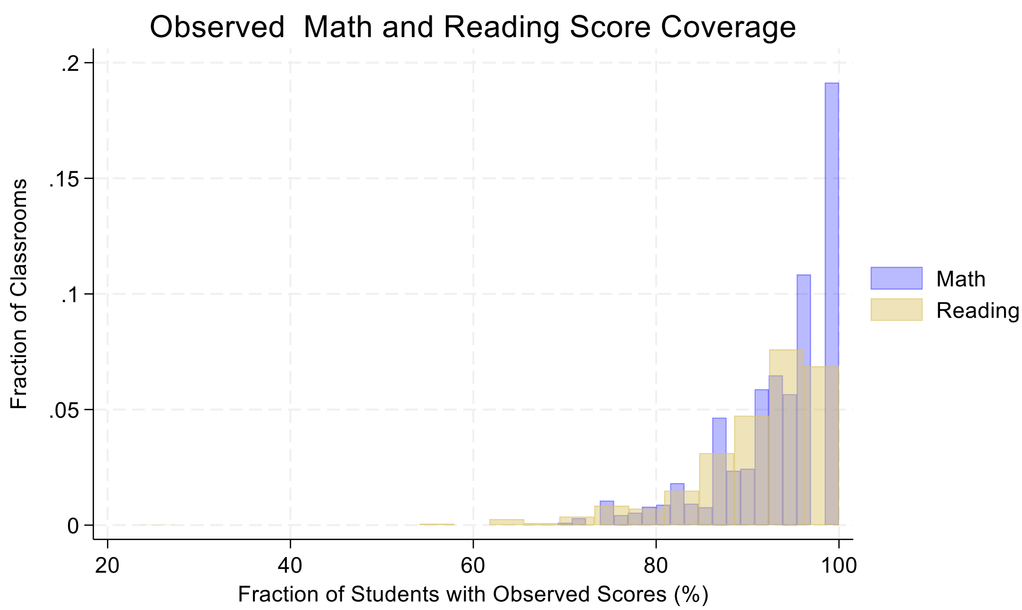
A potential concern is that classroom rank may reflect peer composition rather than being a purely a purely ordinal signal. Conditional on a student's own achievement, higher rank mechanically corresponds to lower average peer achievement, which could independently affect subsequent outcomes. This concern is mitigated by the inclusion of classroom fixed effects in equation (3.1). Conditioning on classroom fixed effect restricts identification to within-classroom variation in rank, thereby purging the estimates of all classroom-level factors that affect students symmetrically. In particular, the classroom fixed effect absorbs standard linear-in-means peer effects, the presence of disruptive peers, overall peer ability, and the dispersion of achievement within the classroom.

Furthemore, the inclusion of classroom fixed effects is crucial: it ensures that the estimated rank coefficients capture ordinal rather than cardinal differences in performance. That is, I compare students who share a classroom and face the same mean, variance, and environment but differ only in their position relative to classmates. If students only responded to cardinal information, such as how far they are from the mean, then the rank coefficients should be insignificant.

3.3.2 Rank Mismeasurement and Simulation

A central empirical challenge in estimating rank-based effects stems from incomplete classroom test score data. Although our identification strategy exploits random variation in ordinal rank within kindergarten classrooms, constructing a student's percentile rank requires information on the actual test scores of all peers. As shown in Figure 3.2, approximately 88% of STAR classrooms report math test scores for over 90% of students, while coverage for reading scores is slightly lower at around 85%. When peer data are missing, the computed rank reflects only a partial view of the classroom, introducing non-classical measurement error into the key explanatory variable. This measurement issue is conceptually similar to the bias highlighted by Micklewright, Schnepf, and Silva (2012), who document bias in peer effect estimates arising from sampled data in the context of international assessments. While their focus is on average peer characteristics, the logic extends naturally to ordinal measures: when peer data are only partially observed, students' calculated ranks may diverge systematically from their true classroom positions, leading to biased inference in downstream analyses.

Figure 3.2 Distribution of Observed Math and Reading Test Scores per Kindergarten Class



Notes*:* This figure displays the distribution of classrooms by the fraction of students with observed math (blue) and reading (yellow) test scores. The x-axis reports the percentage of students in each classroom for whom test scores are observed, and the y-axis shows the fraction of all classrooms falling into each bin. While some classrooms have partial test score coverage (below 100%), a large share have full observability, particularly in math.

To examine how this mismeasurement affects our analysis, I conduct a simulation using classrooms where I observe full test score data. For each student in these fully observed classrooms, I compute two versions of rank: (i) the true percentile rank, based on all classmates, and (ii) a simulated observed rank, based on a random subset of peers. Let M be the full classroom size and $N < M$ the number of students with observed scores. Let m_i and n_i denote student i's rank position in the full and sampled groups, respectively. I define percentile rank in each case using the formula above, and the rank error as:

$$\boldsymbol{Error_i = r^{true} - r^{observed} = \frac{m_i - 1}{M - 1} - \frac{n_i - 1}{N - 1}} \quad \textbf{(3.3)}$$

I conduct simulation exercises in which peer scores are randomly masked at varying levels - from 0% to 50% of classmates. In the simulation, I assume that peer test score data are missing completely at random, meaning the likelihood of a peer's score being unobserved is unrelated to their actual performance or characteristics. This simplifying assumption provides a neutral baseline that isolates the mechanical impact of incomplete peer data on estimated rank effects, without conflating it with selection bias. While the true data-generating process may involve some non-random missingness—such as absences or administrative errors—these are often idiosyncratic and plausibly uncorrelated with rank, making MCAR a reasonable approximation. Moreover, the assumption offers a tractable and transparent framework for simulation, consistent with prior studies on peer effects using partial data (Micklewright et al., 2012; Sojourner, 2013).

I repeated each simulation 500 times, and for each iteration, I recalculate students' classroom ranks based on the incomplete peer data. I then compute the average estimated effects and confidence intervals by rank decile and level of data observability. This repeated-sampling approach enables me to quantify the extent to which rank mismeasurement arises under different levels of peer observability, and how such bias varies across the rank distribution. By explicitly modeling this mismeasurement, I provide a framework for assessing how partial data can affect empirical estimates of rank effects, offering clearer guidance for interpreting results in studies that rely on incomplete classroom information.

3.4 Results

This section examines how students' relative rank in kindergarten classrooms based on reading and math performance predicts their long-term academic outcomes. I focus on four outcomes: high school graduation, high school GPA, participation in college entrance exams (SAT or ACT), and ACT composite scores.

First, I replicate the analysis of Rury (2025) on the average effect of kindergarten reading-specific ordinal rank on long-term academic outcomes. Table 3.3 reports estimates for the full sample (odd columns) and for a subsample of classrooms with complete peer-score observability. Consistent with Rury's core findings, higher reading rank in kindergarten is positively associated with later academic performance and educational engagement. In the full sample, a one–decile increase in reading rank is associated with a 0.034 standard deviation increase in standardized high school GPA, closely aligning with the 0.038 standard deviation effect reported by Rury (2025)[22]. Higher rank is also associated with an increased likelihood of taking a college entrance exam, with an estimated effect of 0.012 standard deviations, comparable in magnitude to Rury's estimates. A one-decile increase in reading-specific kindergarten rank raises standardized ACT scores by 0.081

[22] The differences in the size of estimated coefficients stem from the model specifications, particularly, from the definition of whether the student is old for their grade onto each non-cognitive outcome.

standard deviations in the full sample and by 0.071 standard deviations in the full-observability subsample, with both estimates being highly statistically significant.

Table 3.3 The Long-Term Effects of Reading-Specific Kindergarten Rank

Panel A	(1)	(2)	(3)	(4)
Outcomes	**St. HS GPA**		**HS Grad**	
	Full sample	Full obs. subsample	Full sample	Full obs. subsample
Kindergarden Ordinal Rank in Reading (10 percentile)	0.034*	0.046***	0.003	-0.002
	(0.016)	(0.017)	(0.003)	(0.006)
N	2,230	640	2,832	800
R^2	0.362	0.378	0.218	0.269

Panel B	(1)	(2)	(3)	(4)
Outcomes	**SAT/ACT taking**		**St. ACT test scores**	
	Full sample	Full obs. subsample	Full sample	Full obs. subsample
Kindergarden Ordinal rank in Reading (10 percentile)	0.012*	0.025***	0.081***	0.071***
	(0.005)	(0.007)	(0.018)	(0.015)
N	5,739	1,508	2,238	639
R^2	0.266	0.292	0.429	0.402

Note: Each column represents the separate regression. The regression models control for gender, race, free lunch status, a third-degree polynomial in kindergarten test scores, relative age, classroom size, and classroom fixed effects. Standard errors are clustered at the school level and reported in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

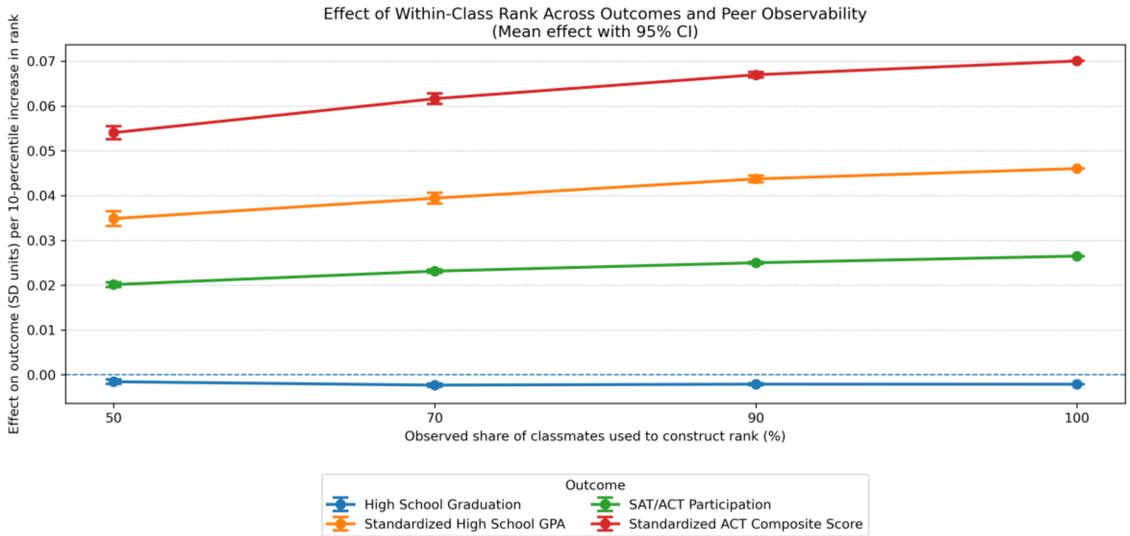
Table A3.1 presents the long-term effects of math-specific ordinal rank on student academic outcomes. The results mirror those obtained for reading-specific rank along several dimensions. In the full sample, a one-decile increase in math rank is associated with a 0.034 standard deviation increase in standardized high school GPA and a 0.012 standard deviation increase in the likelihood of taking the SAT or ACT. Math rank is also a strong predictor of standardized ACT composite scores, with an estimated effect of 0.081 standard deviations per decile. When I restrict the analysis to the full-observability subsample, the effects on GPA and test-taking become smaller and statistically insignificant, whereas the effect on ACT scores remains large and precisely estimated. Across both reading and math domains, there is little evidence that early rank affects high school graduation probabilities. These findings suggest that early relative standing in core academic subjects has persistent effects on test-based outcomes and college-oriented behaviors, while its influence on more extensive margins such as high school graduation is limited.

Comparing across samples, the results suggest that incomplete peer observability leads to attenuation in the estimated effects of early relative rank. When rank is constructed using complete classroom information, coefficients tend to be larger and more clearly distinguishable from zero, despite the loss of precision associated with smaller sample size. This pattern is consistent with measurement error in rank measures derived from incomplete peer data, which biases estimates toward zero. Taken together, the results replicate Rury's average findings, but add a methodological insight: in empirical settings where peer achievement is only partially observed, the long-run effects of early relative rank are likely understated.

To further examine the role of rank observability and to abstract from idiosyncrasies of the realized data, I complement the regression analysis with a simulation exercise. Figure 3.3 summarizes the simulated effects of reading-specific kindergarten rank on long-term outcomes under varying degrees of peer test-score observability[23]. The horizontal axis reports the share of classmates used to construct ranks (50, 70, 90, and 100 percent), while the vertical axis shows the estimated effect on each outcome, measured in standard-deviation units per 10-percentile increase in rank. Points represent mean coefficients across simulations, and vertical bars denote 95 percent confidence intervals.

[23] Figure A3.1 shows the simulated effects of math-specific kindergarten rank on long-term outcomes under varying degrees of peer test-score observability.

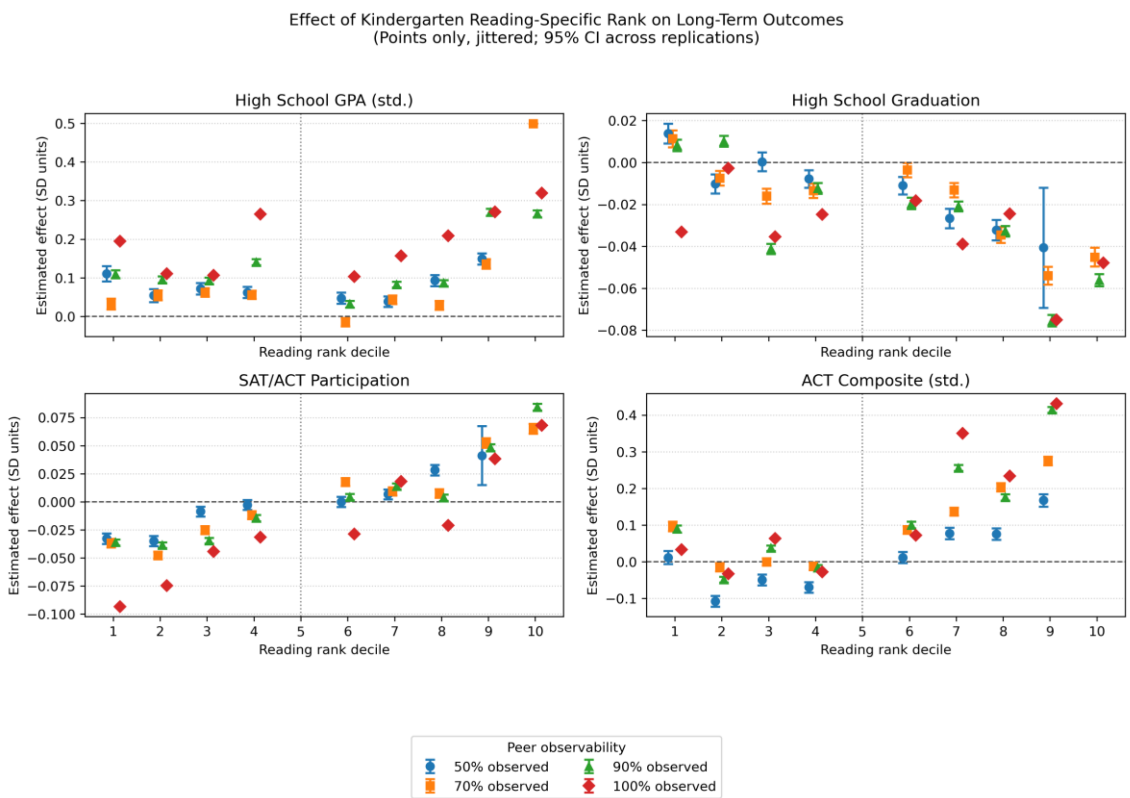
Figure 3.3 Simulated Effects of Kindergaten Reading-Specific Rank on Long-Term Outcomes



The simulation results reinforce the interpretation suggested by the empirical estimates. The estimated effect of rank increases monotonically with peer observability, consistent with attenuation from measurement error when rank is constructed from incomplete information. The strongest pattern is observed for standardized ACT composite scores, where the estimated effect rises steadily from just above 0.05 standard deviations at 50 percent observability to roughly 0.07 standard deviations under full observability. Effects on standardized high school GPA are smaller in magnitude, but follow a similar trajectory, increasing from approximately 0.035 to about 0.046 standard deviations as observability improves. SAT/ACT participation exhibits more modest effects overall, yet still displays a clear and systematic increase with peer coverage. The effect on high school graduation is close to zero at all observability levels and does not display a meaningful upward trend. Overall, the figure illustrates that incomplete peer information substantially understates the long-run impact of early relative standing, and that the strength of rank effects is revealed most clearly when classroom rank is measured using complete peer data.

Finally, I use the simulated rank assignments to explore heterogeneity in rank effects across the distribution of classroom rank. Figures 3.4 and 3.5 present results separately for reading-specific and math-specific rank, with outcomes shown in separate panels and coefficients estimated for each decile of the classroom rank distribution. This structure allows me to assess not only average rank effects, but also how these effects vary across students with different relative positions and how they depend on the salience of rank. The results reveal a pronounced convex pattern: gains are concentrated among students in the upper deciles of the rank distribution, while students in the lower half experience effects that are close to zero and statistically insignificant. In the case of high school GPA, students in the top three deciles experience gains of approximately 0.2 to 0.3 standard deviations, with the largest effects observed at the very top of the distribution. These findings extend Rury (2025) by showing that average rank effects mask substantial heterogeneity and that the long-run benefits of early relative standing accrue disproportionately to students at the top of the classroom hierarchy.

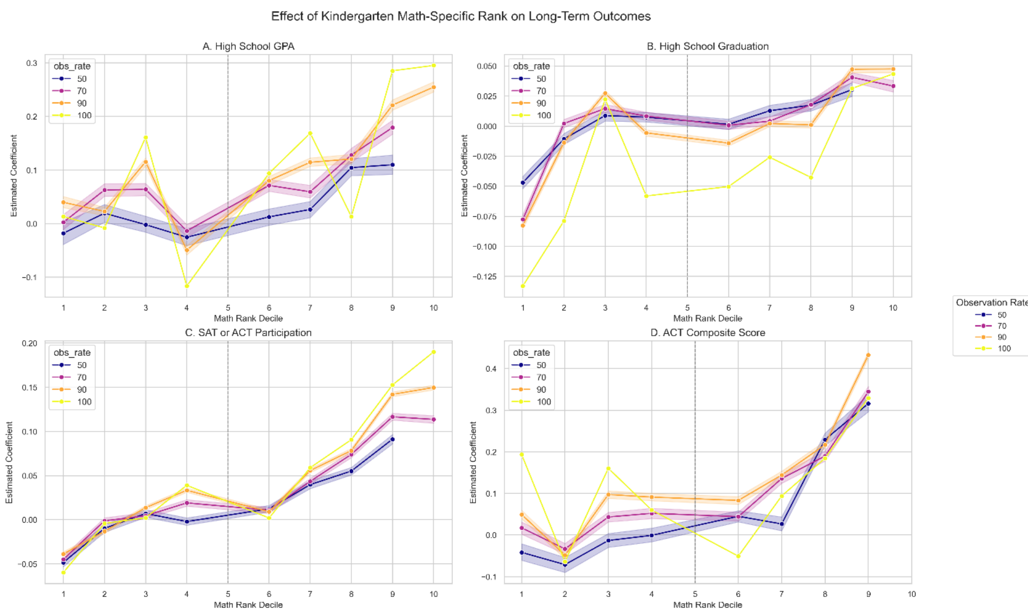
Figure 3.4 The Long-Term Effects of Kindergarten Reading-Specific Rank



Notes*:* Each panel plots the estimated coefficients of deciles for kindergarten reading-specific classroom rank with 95% confidence intervals calculated using the standard errors clustered at the school level. The fifth decile is the omitted reference category in the regression and is therefore not shown. Plotted points represent estimated effects for each non-omitted decile relative to the fifth decile. Estimates are obtained from regressions that control for gender, race, free lunch status, kindergarten test scores, interacted with mean and variance of classroom ability, and classroom fixed effects. Panels A–D correspond to separate outcomes: high school GPA, high school graduation, SAT/ACT participation, and ACT composite score, respectively. Results are presented separately by the proportion of observed peer scores used to calculate classroom rank (50%, 70%, 90%, and 100%), with lower observability inducing greater measurement error.

This nonlinear pattern is consistent with theories of cumulative advantage (DiPrete & Eirich, 2006), in which higher-ranked students may receive more encouragement, teacher attention, or reinforcement of academic identity. Importantly, the convexity I observe aligns closely with recent international findings. For example, Dadgar (2021) reports a similar nonlinear relationship in Sweden, where relative rank in grade 9 affects later academic and labour market outcomes in a pattern that favours students at the top of the classroom distribution. The replication of this pattern in a U.S. context further supports the generalizability of the rank effects across institutional settings.

Figure 3.5 The Long-Term Effects of Kindergarten Math-Specific Rank



Notes*:* Each panel plots the estimated coefficients of deciles for kindergarten math-specific classroom rank with 95% confidence intervals calculated using the standard errors clustered at the school level. The reference category is students in the 5th decile (i.e., 41st to 50th percentile) of the rank distribution. Estimates are obtained from regressions that control for gender, race, free lunch status, kindergarten test scores, interacted with mean and variance of classroom ability, and classroom fixed effects. Panels A–D correspond to separate outcomes: high school GPA, high school graduation, SAT/ACT participation, and ACT composite score, respectively. Results are presented separately by the proportion of observed peer scores used to calculate classroom rank (50%, 70%, 90%, and 100%), with lower observability inducing greater measurement error.

As expected, the magnitude and statistical precision of the GPA effects decline markedly when rank is constructed from incomplete peer data. Under 50% or 70% peer observability, the convex pattern in estimated effects persists but with substantially attenuated magnitudes, reflecting the bias introduced by measurement error in rank construction when peer test score data are incomplete.

In contrast to GPA, I find no evidence that either reading-specific or math-specific relative rank in kindergarten significantly affects the likelihood of high school graduation (Panels B). Estimated effects remain close to zero across the distribution and are statistically indistinguishable from zero, even under full observability. Although slight positive trends appear in the upper deciles of the math-specific rank distribution at high observation rates, these results are imprecise. This suggests that while relative standing may influence students' engagement and performance, it is less consequential for threshold-based outcomes like graduation, which may be governed more by structural or non-academic factors.

Panels C of Figures 3.3 and 3.4 document a clear positive association between kindergarten rank and students' likelihood of participating in the SAT or ACT. The effect is again most pronounced in the upper portion of the rank distribution. When peer data are fully observed, students in the top three deciles of reading rank are 0.07 to 0.09 percentage points more likely to take a college entrance exam than those in the middle deciles. For math, these effects reach 0.10 to 0.18 percentage points. These patterns suggest that high ranked students in kindergarten are more likely to pursue academically challenging pathways. As with GPA, partial observability again leads to substantial attenuation in both effect size and statistical precision, with the top of the distribution disproportionately affected.

Panels D show that both reading- and math-specific kindergarten rank are associated with higher ACT composite scores, particularly at the top of the distribution. The relationship is again strongly convex, with top-decile students achieving gains of 0.40 SD (reading) and 0.45 SD (math), compared to near-zero effects for students below the median rank. These results suggest that the benefits of a high rank not only shape academic engagement (as in SAT/ACT participation) but may also translate into measurable differences in skill development over time. The consistent attenuation of estimated effects under partial peer observability suggests that measurement error in the construction of relative rank—stemming from incomplete classroom data—biases the estimated coefficients toward zero. This attenuation underscores the importance of accurately measuring students' ordinal standing to identify the true causal impact of relative rank on long-run outcomes.

Overall, the findings suggest that where a student ranks in their kindergarten classroom matters well beyond the early years of schooling. Students who are ranked near the top of their class go on to achieve higher GPAs, are more likely to take college entrance exams, and score better on the ACT nearly a decade later. These benefits are not evenly distributed across the rank spectrum; they are concentrated among those in the top third of the classroom distribution, with little evidence of positive effects for students below the median. The simulations further show that missing peer data can mask these patterns. When information on classmates is incomplete, the estimated effects of high rank become smaller and less precise, particularly at the very top of the distribution, where the true effects are strongest. This alignment between effect heterogeneity and bias severity highlights the importance of having complete peer data when studying classroom dynamics. Taken together, the results point to the lasting influence of early relative standing and underscore the need for careful measurement when evaluating how classroom environments shape long-term academic outcomes.

3.5 Conclusion

This chapter shows that students ranked near the top of their kindergarten classroom enjoy lasting academic advantages—particularly in GPA, college entrance exam participation, and ACT scores—nearly a decade later. These effects are highly nonlinear and concentrated among top-ranked students, consistent with models of social comparison and cumulative advantage. A key methodological contribution of this study is the analysis of measurement error in classroom rank arising from incomplete peer data. Simulations based on mismeasured rank under a missing completely at random assumption reveal that partial observability attenuates estimated effects most sharply at the top of the rank distribution—precisely where the true impacts are largest. This alignment suggests that studies relying on incomplete peer data may systematically understate the benefits of a high rank. Accurate measurement of ordinal rank is therefore essential not only for methodological validity but also for understanding how early classroom environments shape long-term academic trajectories.

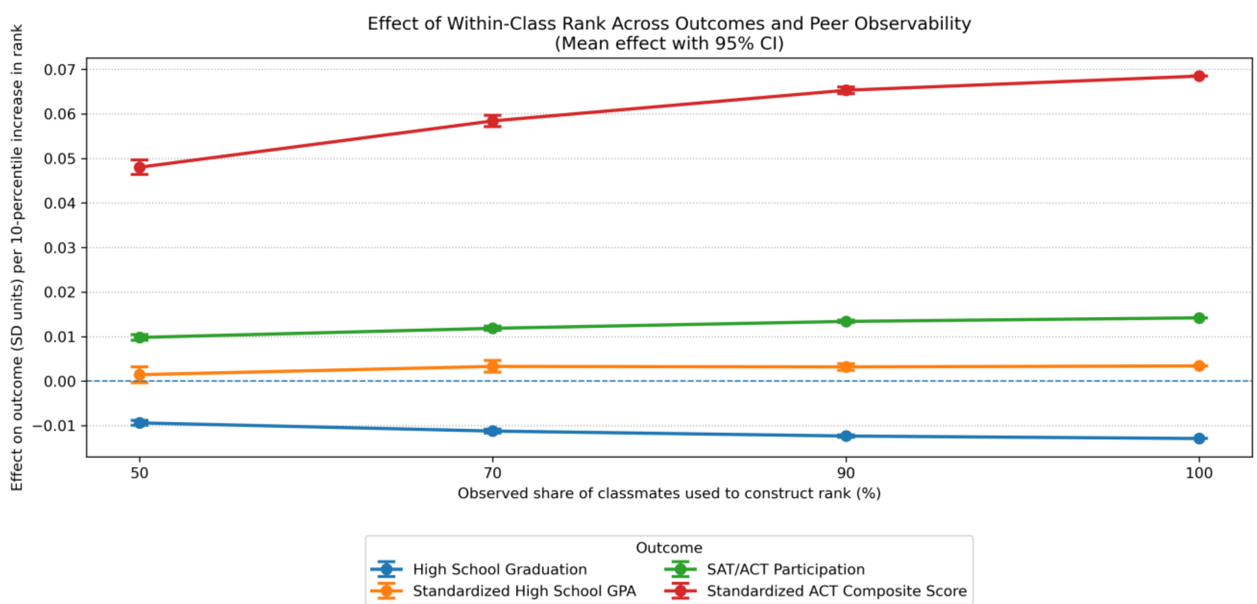
3.A Appendix

Table A3.1 The Long-Term Effects of Math-Specific Kindergarten Rank

Panel A	(1)	(2)	(3)	(4)
Outcomes	**St. HS GPA**		**HS Grad**	
	Full sample	Full obs. subsample	Full sample	Full obs. subsample
Math-Specific Kindergarden Rank (10 percentile)	0.034*	0.008	0.003	-0.012
	(0.016)	(0.016)	(0.003)	(0.006)
N	2,230	688	2,832	848
R^2	0.362	0.378	0.218	0.269
Panel B	(1)	(2)	(3)	(4)
Outcomes	**SAT/ACT taking**		**St. ACT test scores**	
	Full sample	Full obs. subsample	Full sample	Full obs. subsample
Math-Specific Kindergarden Rank (10 percentile)	0.012*	0.014**	0.081***	0.073***
	(0.005)	(0.008)	(0.018)	(0.018)
N	5,739	1,584	2,238	672
R^2	0.266	0.292	0.429	0.402

Note: Each column represents the separate regression. The regression models control for gender, race, free lunch status, a third-degree polynomial in kindergarten math test scores, relative age, classroom size, and classroom fixed effects. Standard errors are clustered at the school level and reported in parentheses $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

Figure A3.1 Simulated Effects of Kindergarten Reading-Specific Rank on Long-Term Outcomes



List of References

Abarcar, P., & Theoharides, C. (2024). Medical worker migration and origin-country human capital: Evidence from us visa policy. *Review of Economics and Statistics*, *106*(1), 20-35.

Alsan, M., Garrick, O., & Graziani, G. (2019). Does diversity matter for health? experimental evidence from oakland. *American Economic Review* , *109* (12), 4071–4111. *American Economic Review* , *95* (2), 158–165.

Ammermueller, A., & Pischke, J. S. (2009). Peer effects in European primary schools: Evidence from the progress in international reading literacy study. *Journal of Labor Economics*, *27*(3), 315-348.

Anastasiadou, A., Kim, J., Sanlitürk, A. E., de Valk, H., & Zagheni, E. (2023). *Sex-and gender-based differences in the migration process: a systematic literature review*. Max Planck Institute for Demographic Research.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of econometrics*, *121*(1-2), 343-375.

Aucejo, E. M., & James, J. (2016). *The Path to College Education: Are Verbal Skills More Important than Math Skills?* (No. 1602).

Aucejo, E., Coate, P., Fruehwirth, J. C., Kelly, S., & Mozenter, Z. (2022). Teacher effectiveness and classroom composition: Understanding match effects in the classroom. *The Economic Journal*, *132*(648), 3047-3064.

Batista, C., Lacuesta, A., & Vicente, P. C. (2012). Testing the 'brain gain' hypothesis: Micro evidence from Cape Verde. *Journal of Development Economics*, *97*(1), 32-45.

Beine, M., Docquier, F., & Rapoport, H. (2008). Brain drain and human capital formation in developing countries: winners and losers. *The Economic Journal*, *118*(528), 631-652.

Bietenbeck, J. (2020). The long-term impacts of low-achieving childhood peers: Evidence from project star. *Journal of the European Economic Association*, *18*(1), 392-426.

Blazar, D., & Kraft, M. A. (2017). Teacher and teaching effects on students' attitudes and behaviors. *Educational evaluation and policy analysis*, *39* (1), 146–170.

Bond, T. N., & Lang, K. (2018). The black–white education scaled test-score gap in grades k-7. *Journal of Human Resources*, *53* (4), 891–917.

Brenøe, A. A., & Zölitz, U. (2020). Exposure to more female peers widens the gender gap in STEM participation. *Journal of Labor Economics*, 38(4), 1009-1054.

Card, D., & Payne, A. A. (2021). High school choices and the gender gap in STEM. *Economic Inquiry*, *59*(1), 9-28.

Carneiro, P., Cruz-Aguayo, Y., Salvati, F., & Schady, N. (2025). The effect of classroom rank on learning throughout elementary school: experimental evidence from Ecuador. *Journal of Labor Economics*, *43*(2), 433-466.

Chand, S., & Clemens, M. A. (2023). Human capital investment under exit options: Evidence from a natural quasi-experiment. *Journal of Development Economics*, *163*, 103112.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *Ameri- can economic review* , *104* (9), 2633–79.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *The Quarterly journal of economics*, *126*(4), 1593-1660.

Clark Blickenstaff*, J. (2005). Women and science careers: leaky pipeline or gender filter?. *Gender and education*, *17*(4), 369-386.

Clemens, M. A., & Mendola, M. (2024). Migration from developing countries: Selection, income elasticity, and Simpson’s paradox. *Journal of Development Economics*, *171*, 103359.

Dadgar, I. (2022). The effect of ordinal rank in school on educational achievement and income in Sweden.

De Brauw, A., & Giles, J. (2017). Migrant opportunity and the educational attainment of youth in rural China. *Journal of Human Resources*, *52*(1), 272-311.

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter?

Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of economics and statistics*, *86* (1), 195–210.

Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of economics and statistics*, *86*(1), 195-210.

Dee, T. S., & Penner, E. K. (2017). The causal effects of cultural relevance: Evidence from an ethnic studies curriculum. *American Educational Research Journal* , *54* (1), 127–166.

Delaney, J. M., & Devereux, P. J. (2019). Understanding gender differences in STEM: Evidence from college applications☆. *Economics of Education Review*, *72*, 219-238.

Delaney, J. M., & Devereux, P. J. (2022). Rank effects in education: What do we know so far?. *Handbook of Labor, Human Resources and Population Economics*, 1-24.

Delgado, W. (2025). Disparate teacher effects, comparative advantage, and match quality. *Economics of Education Review*, *106*, 102648.

Denning, J. T., Murphy, R., & Weinhardt, F. (2023). Class rank and long-run outcomes. *Review of Economics and Statistics*, *105*(6), 1426-1441.

Dinkelman, T., & Mariotti, M. (2016). The long-run effects of labor migration on human capital formation in communities of origin. *American Economic Journal: Applied Economics*, *8*(4), 1-35.

DiPrete, T. A., & Eirich, G. M. (2006). Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments. *Annu. Rev. Sociol.*, *32*(1), 271-297.

Docquier, F., & Rapoport, H. (2012). Globalization, brain drain, and development. *Journal of economic literature*, *50*(3), 681-730.

Docquier, F., Peri, G., & Ruyssen, I. (2014). The cross-country determinants of potential and actual migration. *International Migration Review*, *48*, S37-S99.

Egalite, A. (2024). What We Know About Teacher Race and Student Outcomes: A review of the evidence to date. Education Next, *24(1), 42-49.*

Egalite, A. J., & Kisida, B. (2018). The effects of teacher match on students' aca- demic perceptions and attitudes. *Educational Evaluation and Policy Analysis*, *40* (1), 59–81.

Egalite, A. J., Kisida, B., & Winters, M. A. (2015). Representation in the class- room: The effect of own-race teachers on student achievement. *Economics of Education Review* , *45* , 44–52.

Ehrenberg, R. G., Goldhaber, D. D., & Brewer, D. J. (1995). Do teachers' race, gender, and ethnicity matter? Evidence from the National Educational Longitudinal Study of 1988. *ILR Review*, *48*(3), 547-561.

Elsner, B., & Isphording, I. E. (2017). A big fish in a small pond: Ability rank and human capital investment. *Journal of Labor Economics*, *35*(3), 787-828.

Elsner, B., Isphording, I. E., & Zölitz, U. (2021). Achievement rank affects performance and major choices in college. *The Economic Journal*, *131*(640), 3182-3206.

Encinas-Martín, M. & Cherian M. (2023), *Gender, Education and Skills: The Persistence of Gender Gaps in Education and Skills*, OECD Skills Studies, OECD Publishing, Paris, https://doi.org/10.1787/34680dd5-en.

Eurostat. (2024). *First permits by reason, length of validity and citizenship* [Data set]. Retrieved December 20, 2024, from https://ec.europa.eu/eurostat/databrowser/view/migr_resfirst/default/table?lang=en

Fairlie, R. W., Hoffmann, F., & Oreopoulos, P. (2014). A community college instructor like me: Race and ethnicity interactions in the classroom. *American Economic Review* , *104* (8), 2567–91.

Favara, M. (2012). The cost of acting 'girly': Gender stereotypes and educational choices.

Gershenson, S., Hart, C. M., Hyman, J., Lindsay, C. A., & Papageorge, N. W. (2022). The long-run impacts of same-race teachers. *American Economic Journal: Economic Policy*, 14(4), 300-342.

Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who believes in me? the effect of student–teacher demographic match on teacher expectations. *Economics of education review* , *52* , 209–224.

Gill, D., Kissová, Z., Lee, J., & Prowse, V. (2019). First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision. *Management Science*, *65*(2), 494-507.

Graham, B. S., Ridder, G., Thiemann, P. M., & Zamarro, G. (2020). *Teacher-to- classroom assignment and student achievement* (Tech. Rep.). National Bureau of Economic Research.

Guriev, S., Melnikov, N., & Zhuravskaya, E. (2021). 3g internet and confidence in government. *The Quarterly Journal of Economics*, *136*(4), 2533-2613.

Hamre, B. K., & Pianta, R. C. (2001). Early teacher–child relationships and the trajectory of children's school outcomes through eighth grade. *Child development*, *72*(2), 625-638.

Holt, S. B., & Gershenson, S. (2019). The impact of demographic representation on absences and suspensions. *Policy Studies Journal* , *47* (4), 1069–1099.

Imai, K., Keele, L., & Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects.

Irvine, J. J. (1989). Beyond role models: An examination of cultural influences on the pedagogical perspectives of black teachers. *Peabody Journal of Education*, *66* (4), 51–63.

Jackson, C. K. (2013). Match quality, worker productivity, and worker mobility: Direct evidence from teachers. *Review of Economics and Statistics*, *95* (4), 1096–1116.

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

Jiang, X. (2021). Women in STEM: Ability, preference, and value. *Labour Economics*, *70*, 101991.

Joshi, E., Doan, S., & Springer, M. G. (2018). Student-teacher race congruence: New evidence and insight from tennessee. *AERA Open*, *4* (4), 2332858418817528.

Kane, T. J., & Staiger, D. O. (2012). Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains. research paper. met project. *Bill & Melinda Gates Foundation*.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). Have we identified effective teachers? validating measures of effective teaching using random assignment. In *Research paper. met project. bill & melinda gates foundation.*

Kanny, M. A., Sax, L. J., & Riggers-Piehl, T. A. (2014). Investigating forty years of STEM research: How explanations for the gender gap have evolved over time. *Journal of Women and Minorities in Science and Engineering*, *20*(2).

Khanna, G., & Morales, N. (2017). The it boom and other unintended consequences of chasing the American dream. *Center for Global Development Working Paper*(460).

Krueger, A. B., & Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR. *The Economic Journal*, *111*(468), 1-28.

Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American educational research journal*, *32*(3), 465-491.

Lindsay, C. A., & Hart, C. M. (2017). Exposure to same-race teachers and stu- dent disciplinary outcomes for black students in north carolina. *Educational Evaluation and Policy Analysis*, *39* (3), 485–510.

Lipman-Blumen, J. (1972). How ideology shapes women's lives. *Scientific American*, *226*(1), 34-43.

Lusher, L., Campbell, D., & Carrell, S. (2018). Tas like me: Racial interactions between graduate teaching assistants and undergraduates. *Journal of Public Economics*, *159* , 203–224.

Micklewright, J., Schnepf, S. V., & Silva, P. N. (2012). Peer effects and measurement error: The impact of sampling variation in school survey data (evidence from PISA). *Economics of Education review*, *31*(6), 1136-1142.

Mota Aquino, J. M. (2023). *Essays on the economic causes and consequences of international migration* (Unpublished doctoral dissertation).

Murphy, R., & Weinhardt, F. (2020). Top of the class: The importance of ordinal rank. *The Review of Economic Studies*, *87*(6), 2777-2826.

Park, H., Behrman, J. R., & Choi, J. (2018). Do single-sex schools enhance students' STEM (science, technology, engineering, and mathematics) outcomes?. *Economics of Education Review*, *62*, 35-47.

Penney, J. (2017a). Racial interaction effects and student achievement. *Education Finance and Policy* , *12* (4), 447–467.

Penney, J. (2017b). An instructor like me: Uncovering the channels of racial inter- actions.

Penney, J. (2023). Same race teachers do not necessarily raise academic achievement. *Economics Letters*, *223*, 110993.

Peri, G., Shih, K., & Sparber, C. (2015). STEM workers, H-1B visas, and productivity in US cities. *Journal of Labor Economics*, *33*(S1), S225-S255.

Rabben, L. (2013). Credential recognition in the United States for foreign professionals. Washington, DC: Migration Policy Institute.

Redding, C. (2019). A teacher like me: A review of the effect of student–teacher racial/ethnic matching on teacher perceptions of students and student academic and behavioral outcomes. *Review of educational research*, 89(4), 499-535.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American economic review* , *94* (2), 247–252.

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Education finance and Policy* , *6* (1), 43–74.

Rury, D. (2025). Putting the K in rank: how kindergarten classrooms impact short-and long-run outcomes. *Applied Economics*, 1-15.

Saad, A. F., & Fallah, B. (2020). How educational choices respond to large labor market shocks: Evidence from a natural experiment. *Labour Economics*, *66*, 101901.

Saltiel, F. (2023). Multi-dimensional skills and gender differences in stem majors. *The Economic Journal*, *133*(651), 1217-1247.

Shrestha, S. A. (2017). No man left behind: Effects of emigration prospects on educational and labour outcomes of non-migrants. *The Economic Journal*, *127*(600), 495-521.

Sojourner, A. (2013). Identification of peer effects with missing peer data: Evidence from Project STAR. *The Economic Journal*, *123*(569), 574-605.

Speer, J. D. (2017). The gender gap in college major: Revisiting the role of pre-college factors. *Labour Economics*, 44, 69-88.

Theoharides, C. (2018). Manila to Malaysia, Quezon to Qatar: International migration and its effects on origin-country human capital. *Journal of Human Resources*, *53*(4), 1022-1049.

Tjaden, J., Auer, D., & Laczko, F. (2019). Linking migration intentions with flows: Evidence and potential use. *International Migration*, *57*(1), 36-57.

Todd, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the pro- duction function for cognitive achievement. *The Economic Journal* , *113* (485), F3–F33.

Todd, P. E., & Wolpin, K. I. (2007). The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human capital* , *1* (1), 91–136.

Wedenoja, L., Papay, J. P., & Kraft, M. A. (2020, July). *Second Time's the Charm? How repeat student–teacher matches build academic and behavioral skills*.

Wiswall, M., & Zafar, B. (2018). Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics*, *133*(1), 457-507.