ESSAYS ON APPLIED MICROECONOMICS

A Dissertation

by

BYEUNG-KUK OH

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Jonathan Meer
Committee Members,	Andrew Barr
	Kalena E. Cortes
	Mark Hoekstra
Head of Department,	Steve Puller

May 2022

Major Subject: Economics

ABSTRACT

This dissertation investigates how peers affect students' educational and non-educational outcomes or how pension eligibility has an effect on retirement and healthcare utilization. For the causal identification, I rely on as-good-as random or quasi-experimental variation in the treatment.

First, I examine the persistent effects of ordinal rank by using idiosyncratic variation in the test score distribution across classes. To avoid the self–selection problem, I exploit longitudinal data from middle schools in South Korea, where students are randomly assigned to classrooms. I find a consistently positive impact of students' ordinal rankings in 7^{th} grade on their test scores even five years later. I also find that a higher math rank in 7^{th} grade will increase the likelihood of choosing the math–science track, taking higher-level math courses, and being interested in obtaining a STEM degree. The results of postsecondary outcomes indicate that ordinal rank in 7^{th} grade is positively associated with the probability of attending a university. The results also show that a higher math rank in 7^{th} grade increases the likelihood of applying to a STEM major as well as majoring in a STEM field while attending a university. These effects are likely driven by changes in students' self-confidence and effort provisions, parental investment, and students' attendance in a preferred school.

In addition, I study the spillover effects of misbehaving boys on others' behavioral outcomes, leveraging the random assignment of students into classrooms in South Korean middle schools. Peers with single parents are more likely to misbehave in class, providing an instrument for peers' misbehavior to overcome the reflection problem. Misbehaving boys lead to an increase in the intensity of other students' misbehaviors in the classroom, whereas misbehaving girls do not have statistically significant effect on others. The heterogeneous analysis suggests that these effects are more salient for male pupils and interactions between male students who live in low-income households.

Lastly, using the exogenous rule for the public pension benefit and a dataset for the elderly population from South Korea – one of the developed Asian countries, I provide new evidence of how pension eligibility affects retirement and healthcare utilization. To overcome selection bias, I rely on a regression discontinuity design (RDD) to compare the outcomes of those barely above and below eligibility age thresholds. The findings indicate that crossing pensionable ages increases the probability of being in retirement and inpatient care utilization, while doing so has no significant effect on outpatient care utilization. The local average treatment effect (LATE) of retirement on healthcare utilization is consistent with previous studies in developing country or some high-income countries finding that retirement has a positive impact on inpatient care utilization. However, an additional analysis relying on the Intent-to-Treat (ITT) parameters for covariates-based subgroups suggests that retirement might not be a strong or credible driver of the changes in inpatient care utilization across all subgroups.

DEDICATION

To my family, professors, and friends.

ACKNOWLEDGMENTS

I could never have finished this dissertation without the help and support of many people. I am deeply grateful to my main advisor, Jonathan Meer, for his considerable and continuous guidance, support, advice, encouragement, and most of all patience throughout my doctoral degree. I also would like to thank the other members of my dissertation committee: Andrew Barr, Kalena E. Cortes, and Mark Hoekstra for their helpful comments. Also, I thank all the members of the applied microeconomics group for their comments and feedback on this research.

In addition, I would like to thank the Korean people who I met in this peaceful town of College Station. I gratefully acknowledge Hongil Lim and Hyekyung Lim for their prayers for the success, health, and faith of me and my family. I also want to thank Korean senior or junior from Texas A&M University, Eunseong Ma, Seungyeon Cho, Seyoon Lee, and Sunju Hwang, for their caring, encouragement, and helps.

Finally, I want to express my big thanks to my beloved family. Whenever I was exhausted, my parents, Sunpyo Oh and Hyosook Cho made me comfortable and stay healthy. My mother-in-law, Jungmi Joo, also encouraged my progress. Special thanks go to my wife, Saerom Park for her patience and love. Without her support and advice, I would have lost my track and could not have done this. I want to extend a special thanks to my kids, Habin Oh and Chanbin Oh. Playing and spending time with them were a lot of fun and refreshed me. I forever thank and love my family.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professors Jonathan Meer (Advisor), Andrew Barr, and Mark Hoekstra of the Department of Economics, and Professor Kalena E. Cortes of the Department of Public Service and Administration.

All work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a fellowship from Texas A&M University

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

This dissertation that is consisted of three articles examines the effects of school peers on a student's educational or behavioral outcomes or the effects of crossing pensionable ages on retirement and healthcare utilization. To obtain the causal relationship, I use as good as random variation in class rank, quasi-experimental variation in peer quality, and an exogenous rule for pension benefits.

1.1 Persistent Effects of Ordinal Rank: Evidence from Middle Schools in South Korea

Peer quality is one of the important school inputs to affect students' outcomes in education (1). Recently, studies addressing that a higher rank (or having lower-achieving peers) within a reference group could play an important role in improving student outcomes have been growing. From a policy perspective, these studies tell policy makers about the negative consequence for students with high achievement but a low rank in class. However, there is limited evidence on the serial persistence of its effects both on secondary and postsecondary outcomes and potential mechanisms behind these effects. I address this gap in the literature by providing comprehensive evidence that a student's class rank during middle school has an impact on their outcomes later in life.

The ideal setting for causal identification is to randomly assign rank to students and compare their outcomes. This is unrealistic: thus, I rely on idiosyncratic variation in the test score distribution across classes that have the same class inputs, which leads students who have the same test scores to have different ranks in their classes. Also, the situation that students self-select their classes according to their expected rank confounds "true" rank effects. To circumvent this possibility, I focus on randomly assigned classrooms in South Korean middle schools. For my analysis, I define classroom-subject cells as classes and compute local rank in these classes by following the previously outlined method (2).

My results demonstrate a consistently positive impact of class rankings in 7^{th} grade on test scores present even five years later. The heterogeneous analyses suggest that these effects are more pronounced for female students and students who live in the low-income household. I also find that a higher math rank in 7^{th} grade will increase the likelihood that students choose the math–science track, take higher-level math courses, and show interest in obtaining a STEM degree. Regarding postsecondary outcomes, I find that a higher class rank in 7^{th} grade increases the likelihood of attending a university. A student's math rank in 7^{th} grade has a positive effect on both applying to a STEM major and majoring in a STEM field while attending a university. Exploring potential mechanisms behind these results suggests that rank effects on students' later life outcomes are likely driven by changes in students' self-confidence and effort provisions, parental investment, and students' attendance in a preferred school.

1.2 Watch Out for Misbehaving Boys! Contagion Effects in Randomly Assigned Classrooms

In education, there is a well-documented gender gap in noncognitive skills, especially with boys performing worse than girls on behavioral outcomes (3). Behavioral difficulties among boys in school, including the inability to pay attention in class, to work with other pupils, to follow homework or class materials and to seek help from other classmates can help explain boys' disadvantage for academic and labor market success. However, little is known about how this gap differentially spills over other students' behavioral problems while overcoming threats to causal identification.

I present new evidence that misbehaving boys have spillover effects on other students' behavioral outcomes. Disentangling "true" peer effects from other confounding factors can be a significant challenge in peer effects studies. In this regard, addressing non-random sorting is important for identification. Hence, I rely on an unusual setting that students in South Korean middle schools are randomly assigned to a physical homeroom classroom in which they interact with the same classmates throughout the entire school day. In addition, I circumvent the reflection problem – in which the direction of peer effects is opaque – by linking the reasons for classroom peers' misbehavior to family trouble like a single-parent household and exploiting peers with single–parent households as the instrumental variable (IV). For the straightforward analysis, I use the method of

(4) to compute a single misbehavior composite index (MCI).

By employing a linear-in-means model, I find that misbehaving male students statistically significantly increase the intensity of others' misbehaviors in the classroom, whereas misbehaving female students do not. Specifically, the increase of one single misbehaving male student among 30 students in the classroom leads to an increase of other students' MCI by around 3.2% of standard deviations. The heterogeneous analyses reveal that those effects are even stronger for male students and interactions between male pupils who live in low-income households.

1.3 Retirement and Healthcare Utilization: Evidence from Pension Eligibility Ages in South Korea

In most countries, the eligibility for the public pension benefit is determined by a specific cutoff age in one's declining years. A reform in this cutoff generates the changes in labor activities of elderly workers as well as social and economic effects for them (5). For instance, an increase in the eligibility age for pension benefits would directly incentivize elderly workers to retire later. It may also alter their healthcare utilization as the unintended second-order effect. Combining these effects also helps to understand the correlation between retirement and healthcare utilization, which is ultimately important for an optimal design of retirement policies, including either increasing or decreasing in pension eligibility ages. While recent studies have examined the treatment effects of retirement on healthcare utilization, empirical findings of these studies are conflicting regarding the contexts that researchers are considered. Compared to the direction and size of such treatment effects, there is little discussion on the role of retirement as the force underlying the changes in healthcare utilization.

This paper assesses how pension eligibility ages affect the changes in retirement and healthcare utilization outcomes in the context of South Korea – one of the developed Asian countries. To do so, I use longitudinal data that surveyed individuals aged over 45, which includes information on demographic characteristics, labor market activities, and outcomes regarding inpatient care, outpatient care, and dental care utilization. To address selection bias, I rely on a discontinuity in the normal pension age eligible for old-age pension benefits, providing identifying variation for an

RD design.

I find that crossing pensionable ages increases the probability of being in retirement, and inpatient and dental service utilization, while doing so decreases outpatient service utilization. The local average treatment effect (LATE) of retirement on healthcare utilization shows that retirement increases inpatient care utilization, but decreases outpatient care utilization. The relationship by using the pair of estimated coefficients of the Intent-to-Treatment (ITT) impact on these outcomes, obtained from a separate RD design for subgroups based on covariates, shows that retirement might not be a strong or credible driver for the changes in healthcare utilization across these subgroups.

2. PERSISTENT EFFECTS OF ORDINAL RANK: EVIDENCE FROM MIDDLE SCHOOLS IN SOUTH KOREA

2.1 Introduction

The large and growing amount of literature documents how school peers shape student outcomes, suggesting that the academic and behavioral outcomes of those students are driven by high-achieving peers ((6); (7)), composition in peer gender ((8); (9); (10)), and disruptive peers ((11); (12); (1)). In particular, many previous studies find that selection into schools with higher–performing peers does not improve students' academic performance.[1] One of the possible reasons for this evidence is that the potential benefit of attending such schools may be attenuated by the negative consequence of students with high ability but a low rank in a high-performing class.

While recent studies have suggested that a student's academic rank within a reference group could play an important role in generating a gap in their educational outcomes ((2); (21)), there is little evidence on the serial persistence of its effects on secondary and postsecondary outcomes or what mechanisms contribute to these effects.[2] The impact of school-based inputs and interventions may generally fade over time (22); thus, evaluating the persistence of peer effects is crucial for the assessment of education policies that affect peer composition (1).[3] In addition, the presence of the serial persistence of peer effects provides insight into the role of selection into schools and peer composition as potential determinants of a continuous gap in educational outcomes among students.

In this paper, I present comprehensive evidence that a student's class rank during middle school has a persistent impact on their outcomes later in life. Based on longitudinal data, I estimate the

[1] In the existing literature, the findings on the returns to selection into high–quality schools are mixed. Some studies document the null effect of attending schools with higher-performing peers on students' academic performance ((13); (14); (15); (16)). In contrast, others find a positive effect of attending more selective schools on students' academic achievement ((17); (18); (19) (20)).

[2] One exception is (21) who shed light on the effects of students' class rank both on their secondary and postsecondary outcomes in the US. However, these authors do not focus on potential mechanisms behind rank effects.

[3] Specifically, if peer effects shrink over time and do not affect longer-term outcomes, then concerns over how educational policies such as tracking, within-school random assignment, school choices, or school vouchers affect peer composition may be exaggerated.

effects of a student's class rank in 7^{th} grade on standardized test scores from 8^{th} through 12^{th} grade as well as self-reported STEM outcomes in 11^{th} grade. I also examine how class rank in 7^{th} grade will affect their postsecondary outcomes, including college enrollments, application to a STEM major, being accepted in a STEM major, and actually majoring a STEM field. The variety of information collected on students, teachers, and parents from 8^{th} through 12^{th} grade allows me to investigate the potential driving forces behind the results of the main analysis.

To identify the causal effect of a student's class rank, I follow previously outlined methods that exploit idiosyncratic variation in the test score distribution across classes. This variation may lead students with the same test scores to have different ranks according to which classes they attend. By focusing on classroom–subject cells (classes) to define the local rank measure, my identification strategy leverages the idiosyncratic variation in the distribution of test scores across classrooms and subjects. However, raw test scores are not comparable across classes. If classes with better resources generate higher test scores but do not affect class rank (which can pick up true ability), then the rank effects would be biased upward. I account for this by including classroom by subject fixed effects; thus, I compare students in different classroom–subject cells who have the same test scores demeaned by their own cell mean, but different class rankings due to the test score distribution of their own cells.

The additional challenge plaguing the identification is the self-selection problem. If students can self-select into a peer group conditional on their expectation about within-classroom rank, then the estimated effects of class rank will mostly reflect potential confounds due to sorting. To avoid this problem, I use the random assignment of students into a homeroom classroom where they remain with the same classmates throughout most of the day for an entire year of middle school in South Korea. Hence, students' homeroom classmates are their own reference group members necessary to measure class rankings.

My results demonstrate that class rankings in 7^{th} grade consistently positively impact test scores even five years later. Being at the top of the class in 7^{th} grade compared to the bottom, conditional on achievement in 7^{th} grade and class fixed effects, would obtain a gain of 21.8%, 41.9%, 48.8%, 68.8%, and 37.8% of a standard deviation of test scores from 8^{th} to 12^{th} grade. These effects are statistically distinguishable from zero. The heterogeneous analyses suggest that the positive effects of class rank in 7^{th} grade on test scores are more pronounced for female students and for students who live in the low-income household from 9^{th} to 11^{th} grade. I also find that math rank in 7^{th} grade has a positive impact on the 11^{th} grade reported likelihood that students choose the math–science track, take higher-level math courses, and show interest in obtaining a STEM degree in college. Regarding postsecondary outcomes, I find that class rank in 7^{th} grade positively affects the likelihood of attending a university. A student's math rank in 7^{th} grade has a positive effect on both applying to a STEM major as well as majoring in a STEM field while attending a university.

To explore potential mechanisms behind these results, I test several channels of students' self–confidence, self–expectation, class engagement, and study effort provisions, teacher behavior, parental investment, and high school choices during secondary school. While I find no statistically significant effect of class rank in 7^{th} grade on teacher behavior, I find evidence that class rank positively impacts on students' self–confidence and effort provisions, parental investment, and students' attendance in a preferred high school later. These results indicate that the persistent effects of class rank in 7^{th} grade on students' later life outcomes are likely driven by changes in student– and parent–side mechanisms.

As the potential determinants of class rank in 7^{th} grade, I find that students' class rank is positively associated with parental income rank in class, and female students matched with female teachers are more likely to have a higher rank in class than male students matched with female teachers.

In addition, my results from the test for student effort allocation provide suggestive evidence that students' increased confidence due to a higher rank in their classes could improve their test scores via learning about their ability through class rank.[4] These effects are salient for male students.

[4] To obtain this finding, I replicate the work of (2).

This paper speaks to the growing collection of literature on the future effects of a student' having a higher rank. (21) and (2) exploit idiosyncratic variation in the test score distribution across schools, grades, and subjects to study the long-term impact of a student's rank in US or UK primary schools. They find evidence of positive effects on subsequent test scores, high school graduation, college enrollment, and earnings. (23) and (24) use within-school variation in the test score distribution to show that a higher ability rank of a US high school student positively affects high school completion and college attendance. They also document that it leads to less engagement in risky behaviors. (25) find that within–high school rank in English and mathematics differentially affects college major choices in Ireland. (26) exploits the rank measure determined by idiosyncratic variation in the distribution of classroom peers' cognitive abilities and documents that rank has a positive effect on students' subsequent mid-term test scores in Chinese middle schools. (27) find a positive rank effect on conscientiousness among personality traits conditional on the distribution of peer academic achievement across classes within schools, school types and cohorts in Italy.

This paper is also related to the broader literature on rank. (28) conduct a randomized experiment to inform workers of information on the salaries of their coworkers and find that this information treatment has a negative effect on the job satisfaction of lower–ranked workers. (29) rank children and parents based on their incomes relative to other children and parents in the birth cohort and document that parents' higher income rank is correlated with their children's higher income rank and hence increases their children's college attendance and teenage birth rate. (30) estimate the causal effects of neighborhoods and address that longer experience in better neighborhood quality increases children's percentile rank in the adulthood income distribution.

My study contributes to the literature in several ways. The first is to provide new evidence on a persistent impact of academic rank in a different context. Previous studies use idiosyncratic variation in the test score distributions across elementary school classes in the UK or US and estimate rank effects on test scores ((2); (21)). Compared with them, my approach that exploits idiosyncratic variation in the test score distributions across middle school classes in South Korea suggests more sizable effects on test scores. In addition, using test scores through 12^{th} grade, I find that the positive impacts of class rank in 7^{th} grade on test scores continuously exist until that grade. This result is novel in the literature and provides a suggestive channel for the rank effect on later educational attainment. Second, I find that math rank in 7^{th} grade has a positive and significant impact on preparation for postsecondary STEM study in high school and a change in its subsequent realization while attending a university. These findings reinforce existing evidence that documents the significance of a higher math rank in increasing the probability of a student declaring a STEM major during postsecondary education (21). Finally, I examine the persistence of potential mechanisms during secondary school. The previous study found that a higher rank in elementary school increases a student's confidence in studying three years later (2). My paper complements this evidence by suggesting that the positive effects of class rank on self–confidence, study effort, and parental investment during middle school persist until high school. These findings also enable a comprehensive understanding of the role of such mechanisms in driving the long-lasting impact of class rank on educational outcomes.

2.2 Background and Data

2.2.1 Institutional Background

The South Korean education system provides an unusual setting to circumvent nonrandom sorting in middle schools. Elementary school students who graduate after finishing the 6^{th} grade are assigned by lottery to a local middle school where they attend 7^{th} to 9^{th} grade. These students are randomly assigned to a homeroom classroom where subject teachers can rotate.[5] This practice

[5]Middle school students spend most of the time with classmates in their own homeroom classroom where the majority of classes of subjects are provided. Homeroom teachers can let them know their ordinal ranks in exams compared with their classmates. Thus, I assume that homeroom classroom peers are the reference group of students to obtain their ordinal ranks in a given subject, thereby considering classroom-subject cells as classes. The increased use of ability tracking requires some students to move to other classrooms with their ability group for a given subject within their school. In my data, I observe the common use of ability grouping for math and English, but infrequent for Korean. According to the main identification strategy, I include classroom by subject fixed effects when estimating main specifications later. However, I additionally examine whether the rank effects (in Panel A of Table 6) are still robust against ability grouping by instead including school by subject by ability group fixed effects. Controlling for these fixed effects enables me to capture the mean differences across school–subject–ability groups. In Panel A of Appendix Table D4, results indicate that most of the rank parameters do not change very much. This evidence suggests that ability grouping would not make a significant differences in the effects of ordinal rank computed within students' own homeroom classrooms.

follows social norms and government policies that seek to produce a homogeneous homeroom classroom with respect to academic ability (31). One of the most common practice is to order students by an entrance exam or academic performance from the previous year. The first-ranked student is assigned to classroom X, the second-ranked student is assigned to classroom Y, the third-ranked student is assigned to classroom Z, and so on. The rank is not known to either the student or the parents before the classroom assignment is completed.

After middle school students graduate, they go on to high school, spanning 10^{th} to 12^{th} grade. There are two types of admission processes to high schools (32). In the first process, students can apply to a magnet school, private autonomous high school, art high school, athletic high school, or vocational high school. These schools determine admission by judging academic performance and recommendations from middle school principals and teachers. In the second process, the admission process is conducted by lottery, where students can apply to autonomous public high schools, science-focused high schools, art-focused high schools, and general academic high schools. For example, middle school graduates in Seoul can apply to two or three schools in all the school attendance districts (1st stage), as well as within their own school attendance districts (2nd stage) linked to their residence. At each stage, admission is done by lottery. If students are not selected by those schools at either stage, then they are randomly assigned to a local high school within their own school attendance district.

2.2.2 Dataset and Ordinal Rank Measure

My dataset is the Seoul Education Longitudinal Study of 2010 (SELS2010, hereafter). This dataset surveyed 7^{th} grade students in Seoul, along with their parents and subject teachers assigned to their own subjects. The longitudinal structure of this data enables me to capture these students' longer-term trajectories. This includes their standardized test scores – which are obtained after the first semester of the school year – for mathematics, English, and Korean from 7^{th} through 12^{th} grade. A follow-up study shortly after high school graduation includes information about college applications and enrollment, major choices, and labor market outcomes at age 18 or 19. Data sampling was performed in two steps: first, 74 middle schools were randomly selected from a population of about 370 public or private middle schools; second, two classrooms were randomly drawn within each school. The number of observations began with 4,544 in 2010. However, 4,347, 4,162, 3,541, 3,394, and 3,305 samples remained from 2011 through 2015 due to attrition. In the follow-up study (taking place within two months after the high school graduation), 2,195 students reported a variety of postsecondary outcomes.

To obtain a comparable rank measure across classes of different size, I construct a student's local (class) percentile rank ($Rank_{ijcs}$) (2):

$$Rank_{ijcs} = \frac{R_{ijcs} - 1}{N_{cs} - 1} \tag{2.1}$$

where N_{cs} is the size of classroom c of school s. A student i's ordinal rank position within classroom (c)–subject (j) cells is R_{ijcs} with the minimum (1) and the maximum numbers (N_{cs}). $Rank_{ijcs}$ is the school–classroom–subject-adjusted rank of students that I use to capture the class rank effects in the estimation. This measure is uniformly distributed from 0 (the lowest) to 1 (the highest).

Based on Equation (1), 7^{th} grade students have different percentile ranks across three subjects: mathematics, English, and Korean. Using three subjects allows me to obtain three student data sets with the same number of observations (4,544). I stack these datasets into a pooled dataset with the observations (13,632). Of 13,632 observations in 7^{th} grade, I drop 88 observations not linked to percentile ranks due to missed test scores. To rule out concern about this sample dropping, I compare the remaining 13,544 observations with those (88 observations) that are dropped. I find no statistically significant differences between them with respect to students' characteristics.

For my empirical work, I standardize test scores for each subject to have a mean of zero and a standard deviation of one, which enables me to interpret the estimated coefficients in an easier way. I include student and subject teacher characteristics, STEM outcomes in 11^{th} grade, and postsecondary outcomes into the dataset.

Table 2.1: Summary Statistics

	Mean	Std. Dev	Min	Max	N
A. Student Test Scores					
Math score in $8^{th} grade$	15.81	7.73	0.00	30.00	4,331
in $9^{th} grade$	15.07	8.36	1.00	30.00	4,120
in $10^{th} grade$	11.08	6.92	0.00	30.00	3,424
in $11^{th} grade$	9.18	5.51	0.00	30.00	3,377
in $12^{th} grade$	8.99	6.26	0.00	30.00	3,276
Eng score in 8th grade	20.74	8.90	0.00	35.00	4,331
in $9^{th} grade$	21.38	9.51	1.00	35.00	4,130
in $10^{th} grade$	16.94	8.50	0.00	35.00	3,439
in $11^{th} grade$	13.59	7.94	0.00	35.00	3,377
in $12^{th} grade$	13.38	8.43	0.00	35.00	3,277
Kor score in $8^{th} grade$	20.16	7.68	2.00	34.00	4,332
in $9^{th} grade$	18.17	7.61	1.00	35.00	4,135
in $10^{th} grade$	15.86	7.34	0.00	33.00	3,436
in $11^{th} grade$	15.35	7.66	0.00	34.00	3,378
in $12^{th} grade$	15.53	7.17	0.00	34.00	3,230
B. Student Rank					
Math rank in 7^{th} grade	0.50	0.30	0.00	1.00	4,515
Eng rank in 7^{th} grade	0.50	0.30	0.00	1.00	4,507
Kor rank in 7^{th} grade	0.50	0.30	0.00	1.00	4,522
C. Student Characteristics					
Female	0.46	0.50	0.00	1.00	4,544
Single parent	0.11	0.31	0.00	1.00	4,544
Number of siblings	0.84	0.74	0.00	6.00	4,439
Parents w/ BA or Higher Degree	0.69	0.46	0.00	1.00	4,185
Free lunch	0.09	0.29	0.00	1.00	4,543
D. Teacher Characteristics					
Female teacher	0.82	0.38	0.00	1.00	570
Teacher Age over 40	0.54	0.50	0.00	1.00	562
Teacher's college	0.62	0.48	0.00	1.00	567
Teacher's post graduate	0.33	0.47	0.00	1.00	567
Administrative teacher	0.16	0.37	0.00	1.00	565

Table 2.1 lists the summary for sample statistics. Panel A reports baseline statistics of raw test scores across subject from 8^{th} through 12^{th} grade. Means of test scores in math and English generally decrease over time. Panel B describes the sample statistics of 7^{th} grade rank in each subject, which is calculated by Equation (1). Standard deviations in all subjects are 0.30, which means that at a mean class size of 30 students class rank would vary on average by 9 (0.30×30=9) rank positions. Panels C and D report baseline statistics of student and teacher characteristics,

respectively. Female students are on average 46% of sample students, while female teachers are on average 82% in my samples. In addition, students who receive free school meals are on average 9%.

2.3 Identification Strategy

2.3.1 Empirical Specification

I examine the short-term effects of class rank in 7^{th} grade on test scores in 8^{th} grade by estimating the following equation:[6]

$$y_{ijs}^{8th} = \beta_0 + \beta_1 Rank_{ijcs}^{7th} + f(score^{7th}) + X_i\gamma_1 + T_j\gamma_2 + \theta_{jc} + \epsilon_{ijs} \tag{2.2}$$

where y_{ijs}^{8th} is a standardized test score of student i for subject j in 8^{th} grade at school s. $Rank_{ijcs}$ is a class percentile rank of student i for subject j in 7^{th} grade at school s. X_i is a vector of student characteristics, including indicators for a female student, for a single-parent household, the number of siblings, and indicators for having at least one parent with a BA degree or higher and for having a free lunch. T_j is a vector of subject teacher characteristics, including indicators for teacher gender, for teachers' having less than five years of teaching experience, for teachers' graduation from teachers college, for teachers having a Master's degree or higher, and for administrative teachers. I control for a third-order polynomial of a student's 7^{th} grade test scores, $f(score^{7th})$, to allow for nonlinearity in the relationship between the outcome and 7^{th} grade test scores.[7] I include classroom by subject fixed effects θ_{jc} to account for mean differences across classroom–subject cells.[8] Standard errors are clustered at the classroom level to allow for arbitrary correlations

[6]This specification follows a rank-augmented education production function, which assumes that rank, human capital and class effects are additively separable (2). This specification is also based upon the value-added model proposed by (33).

[7]Including the test score distribution for a student addresses my identification strategy comparing students who have different ranks but the same test score in the specification.

[8]Including classroom by subject fixed effects allows me to compare students in different classrooms who have the same test score relative to their classroom mean in a given subject, but different percentile ranks due to the test score distributions across classroom-subject cells. Also, these fixed effects can account for any kind of classroom-level shocks, including unobserved influences of peers and teachers in a classroom. However, note that these fixed effects do not completely subsume observable subject teacher characteristics in Equation (2.2) because students within the same homeroom classroom can have different subject teachers due to ability grouping in my data.

among students within the same classrooms. β_1 represents the short-term effect of ordinal rank on the test score.

To investigate the longer-term effects of ordinal rank in 7^{th} grade, I estimate the following equation:

$$y_{ijs}^{T} = \beta_0 + \beta_1 Rank_{ijcs}^{7th} + f(score^{7th}) + \theta_{jc} + \epsilon_{ijs} \tag{2.3}$$

where y_{ijs}^{T} is test scores in year T=2011 (8^{th} grade) through 2015 (12^{th} grade). I expect that the randomness of 7^{th} grade rank effects conditional on students' test scores and classroom by subject fixed effects will hold for these grades. Hence, Equation (2.3) takes the most parsimonious form to estimate the rank effects on academic performance over time. Despite the efficiency of this approach, I will obtain the rank parameters by including student and teacher characteristics in Equation (2.3) to show the robustness of the estimates on ordinal rank. This task will also show the validity for the random assignment of ordinal rank to students conditional on the test score distributions and classroom by subject fixed effects.

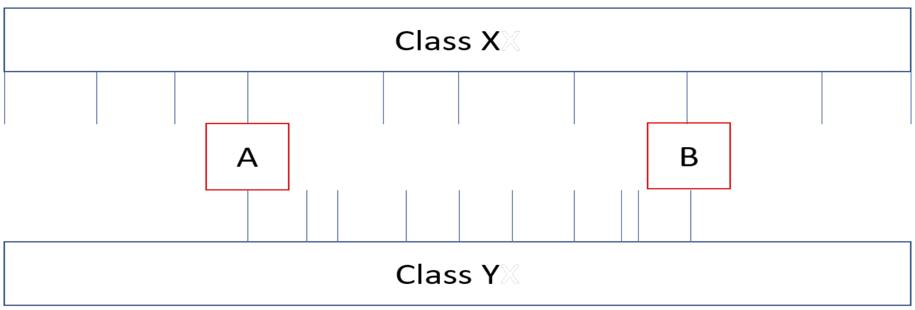
I rely on Equation (2.3) when estimating 7^{th} grade rank effects on STEM outcomes in 11^{th} grade and postsecondary outcomes. These outcomes consist of things that do not differ across subjects. Therefore, I additionally estimate the marginal effects of class rank in each subject by collapsing this specification to a single observation per student.

2.3.2 Variation in Class Rank

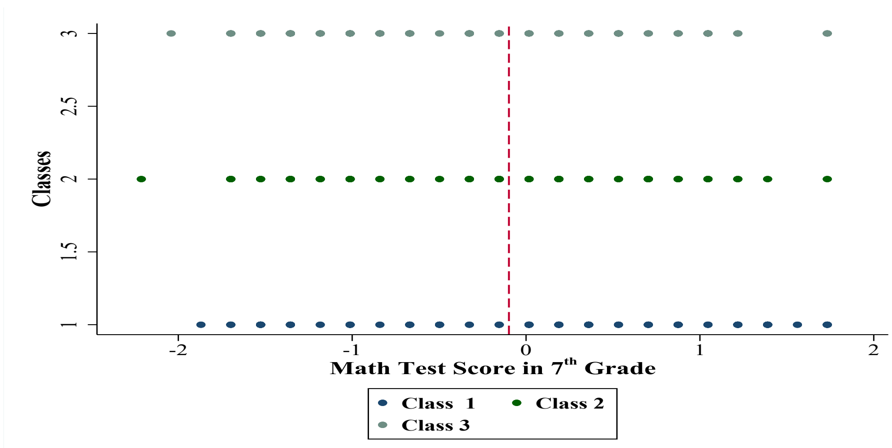
To identify class rank effects, I exploit idiosyncratic variation in the test score distributions across middle school classes. The class ranks of students can be determined upon this variation. For example, students with the same test score in 7^{th} grade may have different ranks according to their classmates' achievement scores. Panel A of Figure 1 supports this possibility by assuming that the average score in classes X and Y is the same, but the test score distributions are different in these classes. Thus, a student with test score B above the average score will have a higher rank in Class Y than Class X, whereas a student with test score A below the average score will have a higher rank in Class X than Class Y. Panel B of Figure 1 actualizes the example of Panel

A that students have different ranks across math classes, conditional on their own test score and the test score distributions of classes in my data. Here, there are three math classes that have the same mean, but different minimum and maximum student test scores. These three classes all have students at the standardized test score 0.19, but each of those students has different ranks of 0.56 (class 1), 0.63 (class 2), and 0.61 (class 3).

Figure 2.1: Relation Between Rank and Test Score Distribution



(a) Panel A: Ordinal Rank Dependency on Test Score Distributions

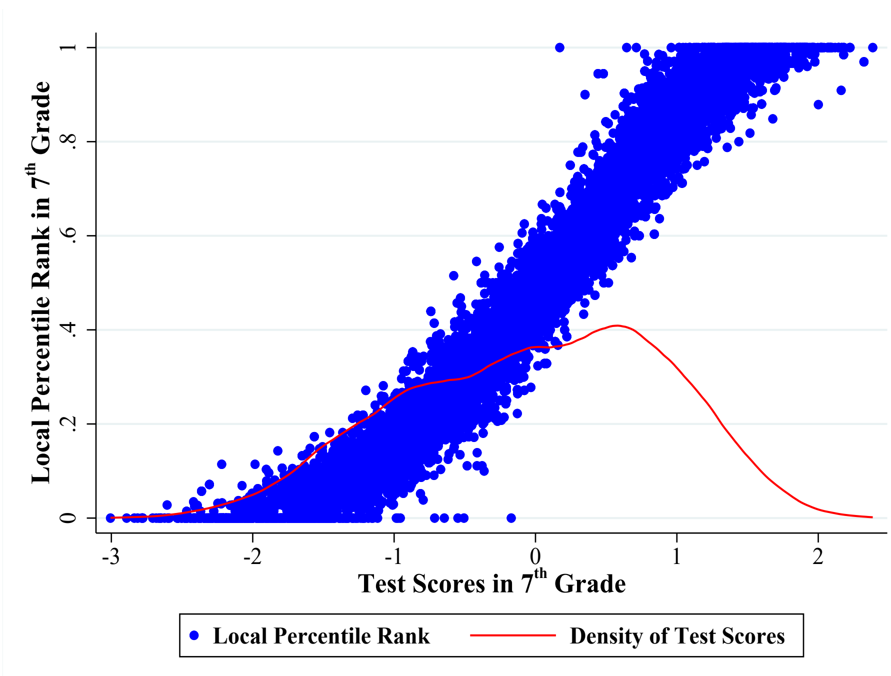


(b) Panel B: Test Score Distributions Across Math Classes

Notes: These figures show the relation between class rank and test score distribution. In Panel A, the classes have the same mean, but Class Y has a smaller variance. A student with score B above the mean will have a higher rank in Class Y than Class X while A student with score A below the mean will have a higher rank in Class X than Class Y. In Panel B, the graph presents data from three middle school math classes that have the same mean (-0.09) of the standardized test score. Given the different test score distributions of three classes, each student who has the same standardized score of 0.19 has different ranks of 0.56 (Class 1), 0.63 (Class 2), and 0.61 (Class 3).

My identifying variation includes variation in the composition of students' test scores across randomly assigned classrooms carried out in South Korean middle schools. Given the assignment rule, as described in subsection 2.2.1, one may be concerned that conditional on peer test score distributions, the variation in rank across classes is not enough for identification. Figure 2.2 shows the distributional variation in percentile ordinal ranks and 7^{th} grade test scores de-meaned by classroom–subject cells (classes) in my data, depicting the relationship between class rank and the (de-meaned) test score. As students move to the mean (zero on the X-axis) of the test score in their classes, their ranks range from 0.3 to 0.6 (on the Y-axis). Hence, this figure does not provide evidence of very local variation that students marginally change their class rankings for given test scores.

Figure 2.2: Variation in Rank and Test Scores Across Classroom–Subject Cells



Notes: This figure presents the joint distribution of class rank (on the Y-axis) and test scores (on the X-axis), and the density of test scores in 7^{th} grade. Test scores are the de-meaned test scores by 7^{th} grade classroom–subject cell.

2.3.3 Threats to Identification

To interpret the rank parameter (β_1) in Equation (2.2) as causal, an ideal setting is to randomly allocate students who have the same test score to classes that have the same class inputs but different test score distributions, which leads them to have different ranks. In order to approximate this in my empirical specification (Equation (2.2)), I make the following conditionally independence assumption based on the framework of the potential outcomes causal model (34):

$$y_{ijs}^{T}(r) Rank_{ijcs} | f(score^{7th}), \theta_{jc} \tag{2.4}$$

where $y_{ijs}^{T}(r)$ is the potential outcome for student (i) as a function of potential rank (r). This means that rank is not related to potential outcomes conditional on a function of the test score in 7^{th} grade and classroom by subject fixed effects.[9] However, this assumption could be violated if unobserved shocks that are not captured by $f(score^{7th})$ and θ_{jc} differently affect the potential outcomes. That is, the assumption (2.4) requires that a student's rank is as good as random after conditioning on $f(score^{7th})$ and θ_{jc}. To test for violations of assumption (2.4), I examine whether class rank in 7^{th} grade is correlated with student or teacher characteristics conditional on cubic in the 7^{th} grade test score and classroom by subject fixed effects. This test is analogous to a balancing test in the context of randomized control trial. The intuition is that if the observable students or teacher characteristics appear to be balanced conditional on the 7^{th} grade test score and classroom by subject fixed effects, then it might be reasonable that the unobservable student or teacher characteristics are as well. Table 2.2 shows that no estimates are statistically significant at the conventional level, suggesting that the effect of rank made by idiosyncratic variation in the test score distribution across classes that have the same inputs is not associated with student or subject teacher characteristics. Hence, I conclude that class rank is quasi-randomly assigned to students

[9]Alternatively, this assumption means that, for example, the mean potential outcome for $y_{ijs}^{T}(1)$ (or $y_{ijs}^{T}(0)$) is the same for both the top of class ($r = 1$) and the bottom of class ($r = 0$) conditional on a function of the test score in 7^{th} grade and classroom by subject fixed effects: $E[y_{ijs}^{T}(1)|r = 1, f(score^{7th}), \theta_{jc}]$=$E[y_{ijs}^{T}(1)|r = 0, f(score^{7th}), \theta_{jc}]$ or $E[y_{ijs}^{T}(0)|r = 1, f(score^{7th}), \theta_{jc}]$=$E[y_{ijs}^{T}(0)|r = 0, f(score^{7th}), \theta_{jc}]$. If these two conditions are satisfied, then the selection bias and the heterogeneous treatment effect bias would be eliminated. Hence, the average treatment effect (ATE) can be written as: ATE=$E[y_{ijs}^{T}(1)|r = 1, f(score^{7th}), \theta_{jc}]$ – $E[y_{ijs}^{T}(0)|r = 0, f(score^{7th}), \theta_{jc}]$.

conditional on the test score distribution and classroom by subject fixed effects.

Table 2.2: Effects on Student and Teacher Characteristics

	(1) Female Student	(2) Single Parent	(3) # of Siblings	(4) Parent BA+	(5) Free Lunch
7^{th} Grade Rank	-0.056	0.009	0.005	0.083	0.006
	(0.062)	(0.041)	(0.091)	(0.054)	(0.033)
Observations	13,544	13,544	13,239	12,479	13,541
R^2	0.234	0.052	0.034	0.183	0.095
	(6) Female Teacher	(7) Teacher Age	(8) Teacher College	(9) Teacher Master +	(10) Admin Teacher
7^{th} Grade Rank	-0.047	-0.101	0.019	-0.091	-0.039
	(0.048)	(0.069)	(0.062)	(0.065)	(0.047)
Observations	10,206	10,118	10,120	10,140	10,059
R^2	0.724	0.699	0.724	0.721	0.724

Notes: Each cell represents results from a separate regression including cubic in 7^{th} grade test score and classroom by subject fixed effects. In Columns (1)-(10), dependent variables are indicators for student gender, for a single-parent household, number of siblings, and indicators for having at least one parent with a BA degree or higher and for free lunch, for teacher gender, for teacher age over 40, for teacher's graduation from teachers college, for teacher having a Master's degree or higher, and for administrative teacher. Standard errors, shown in parentheses, are clustered at the classroom level.
*p<0.10, ** p<0.05, *** p<0.01.

Next, the rank parameter (β_1) in Equation (2.2) would be biased if students self–select into different schools conditional on their expected rank in their schools. For instance, strong students have a lower rank in their classes in a high-quality middle school where they expect to have a higher chance to go to a good high school, which results in a downward bias in the rank estimate. To confirm non-random sorting of students into classrooms, I test whether the average characteristics of classroom peers are associated with the student's own characteristics. I perform this by using the following equation:

$$x_{ics} = \alpha + \beta \frac{1}{n_c - 1} \sum_{i \neq j} (x_{jcs}) + \gamma \frac{1}{n_s - 1} \sum_{i \neq j} (x_{js}) + \lambda_s + \varepsilon_{ics} \tag{2.5}$$

where x_{ics} is student i's characteristic in randomly assigned classroom c within school s. n_c and n_s are the number of students in classroom c and school s, respectively. The term including x_{js} on the right-hand side represents the mean of the characteristic within the school, which is necessary to alleviate a negative mechanical bias that is inherent in a typical randomization test of peer assignment (35). School fixed effects (λ_s) account for random assignment within a school. ε_{ics} is the idiosyncratic error term. Table 2.3 indicates that a student's characteristics are not correlated with classroom peers' characteristics, suggesting random assignment of students into their peers' classroom.

Table 2.3: Randomization Tests for Peer Group Assignment

	(1) Female Student	(2) Single Parent	(3) # of Siblings	(4) Parents BA+	(5) Free Lunch
Avg. Classroom Peers	-0.029	-0.057	-0.125	0.001	-0.001
Characteristics	(0.164)	(0.068)	(0.109)	(0.018)	(0.036)
Observations	4.544	4.544	4,439	4,185	4,543
R^2	0.807	0.723	0.830	0.969	0.959

Notes: Each column represents results from a separate regression, including school fixed effects and the leave-out mean of school peers' characteristic. I also control for the school peers' leave-out mean (school-level mean for the relevant characteristics excluding the individual's own values) to overcome the negative mechanical bias of the randomization tests (35). In each column, dependent variables are indicators for student gender, for a single-parent household, number of siblings, and indicators for having at least one parent with a BA degree or higher and for free lunch. Standard errors, shown in parentheses, are clustered at the school level.
*p<0.10, ** p<0.05, *** p<0.01.

An additional concern about the identification of rank effects over time is that attrition from the sample is likely to bias the results. If attrition is systematically associated with class ranks in

7^{th} grade (i.e. if a low ranked student is more likely to exit the sample), then the estimated results would reflect the sample selection. Table 2.4 shows that there is no economically meaningful correlation between the average class ranks in 7^{th} grade and the likelihood of both attrition and cumulative attrition of the sample. Hence, the attrition problem of the dataset over time would not bias the results.

Table 2.4: Effects on Student Attrition

	(1) 8th Grade	(2) 9th Grade	(3) 10th Grade	(4) 11th Grade	(5) 12th Grade	(6) Post Secondary
A. Non-cumulative Attrition						
Avg. 7^{th} Grade Rank	0.085	0.061	-0.104	-0.098	-0.073	0.045
	(0.064)	(0.084)	(0.141)	(0.146)	(0.159)	(0.161)
Observations	4,471	4,471	4,471	4,471	4,471	4,471
R^2	0.045	0.050	0.137	0.125	0.126	0.095
B. Cumulative Attrition						
Avg. 7^{th} Grade Rank	0.085	0.043	-0.077	-0.089	-0.087	-0.132
	(0.064)	(0.085)	(0.141)	(0.145)	(0.154)	(0.159)
Observations	4,471	4,471	4,471	4,471	4,471	4,471
R^2	0.045	0.052	0.137	0.131	0.130	0.102

Notes: Each column represents results from a separate regression including cubic in 7^{th} grade test scores for mathematics, English, and Korean, and classroom fixed effects. In Panel A, the dependent variables are indicators for a student leaving the sample in each of the years (8th–12th grades and postsecondary period). In Panel B, the dependent variables are indicators for cumulative attrition as opposed to attrition in each year. The average 7^{th} grade rank is defined as the average rank across those three subjects, conditional on no missing values in rank. Standard errors, shown in parentheses, are clustered at the classroom level.
*p<0.10, ** p<0.05, *** p<0.01.

2.4 Results

2.4.1 Graphical Evidence on Rank Effects

I begin by showing a graphical relationship between class rank in 7^{th} grade and a student's outcomes. Specifically, I first regress these outcomes on a third-order function of students' test scores and classroom by subject fixed effects. I then regress ordinal rank in 7^{th} grade on a third-order function of students' test scores and classroom by subject fixed effects. Doing these tasks provides the residuals in a student's outcomes or class rank in 7^{th} grade. In an empirical matter, these residuals also capture identifying variation in the outcomes or class rankings.

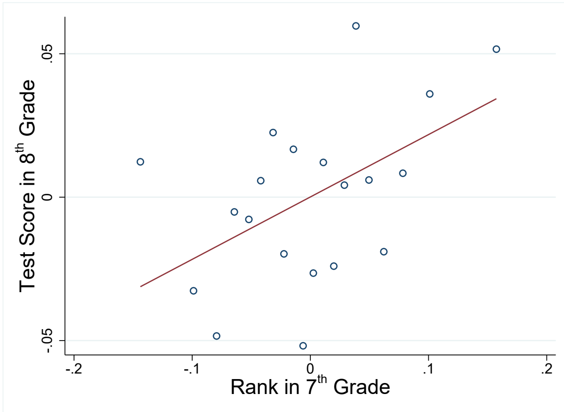
The visual relationships are provided in Figures 2.3–2.5. In each panel, the values on the X-axis represent the residualized rankings in 7^{th} grade. The values on the Y-axis represent the residualized outcomes. Each panel displays local averages of the test score in a given grade by the distribution of ordinal rank, as represented by the blue circles. I also fit a linear line to the underlying data, as represented by the red line.

Figure 2.3 shows the relationship between ordinal rank in 7^{th} grade and test scores from 8^{th} through 12^{th} grade. The upward-sloping linear lines suggest that ordinal rank in 7^{th} grade is positively associated with test scores in the future (Panels A, B, C, D, and E). This visual relationship indicates that a higher ordinal rank persistently correlates with higher test scores.

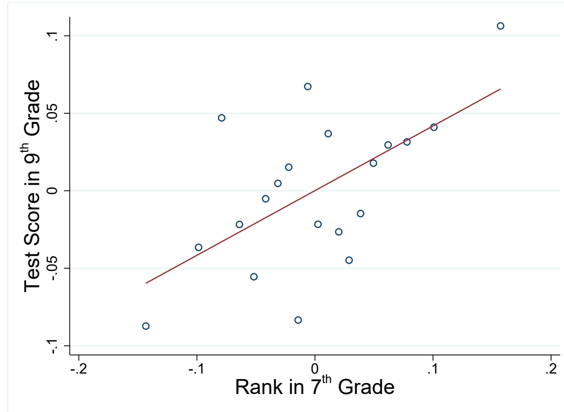
Figure 2.4 depicts the relationship between class rank in 7^{th} grade and STEM-related outcomes surveyed in high school (11^{th} grade). Looking at the linear lines suggests that ordinal rank is positively correlated with the likelihood of a math-science track choice, taking at least one advanced math course, or interest in majoring in a STEM field during postsecondary periods, whereas it is negatively associated with attending a STEM-focused high school.

Figure 2.5 displays the relationship between class rank in 7^{th} grade and postsecondary outcomes. Based on the direction of the linear lines, class rank is likely to change the outcomes, except for the likelihood of being employed at age 18 or 19 (Panel C), applying to a STEM major (Panel D), or being accepted conditional on applying to a STEM field (Panel E). However, limiting observations to students who applied to a university rather than a community college provides the positive relationships between ordinal rank and STEM-related outcomes (Panels G, H, and I)

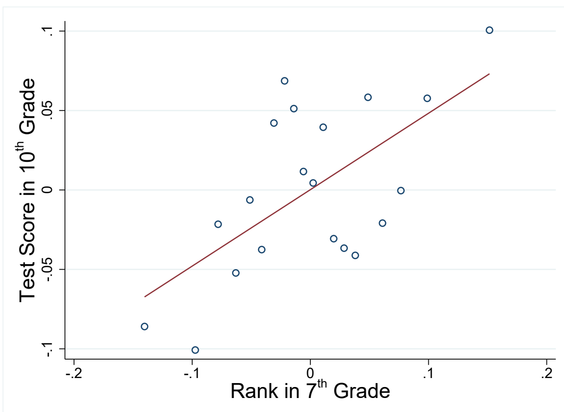
Figure 2.3: Relationship between Rank and Test Scores



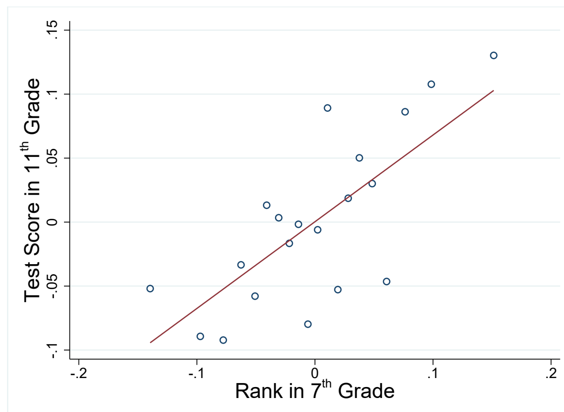
(a) Panel A: 8^{th} Grade



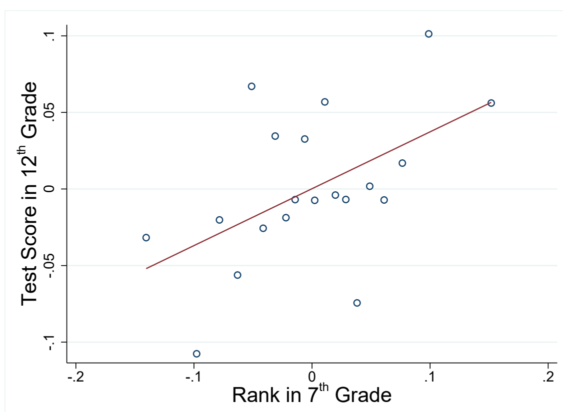
(b) Panel B: 9^{th} Grade



(c) Panel C: 10^{th} Grade



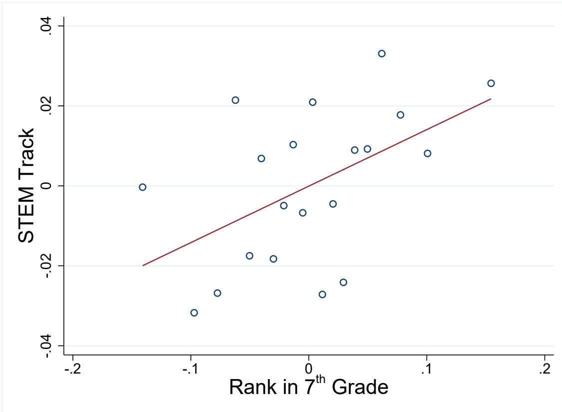
(d) Panel D: 11^{th} Grade



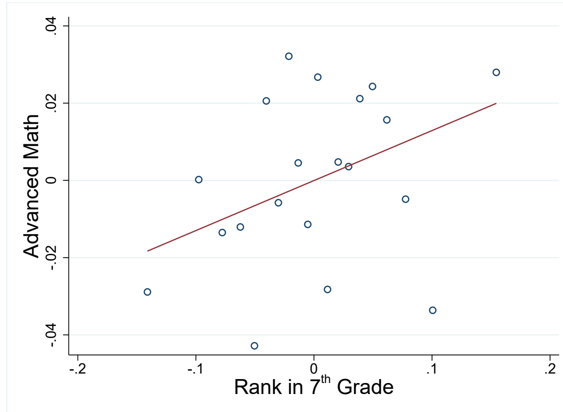
(e) Panel E: 12^{th} Grade

Notes: This figure presents the correlation between class rank in 7^{th} grade on the x-axis and test scores from 8^{th} grade through 12^{th} grade on the y-axis. Both variables are residualized with respect to a third-order polynomial of a student's 7^{th} grade test scores and classroom–subject fixed effects.

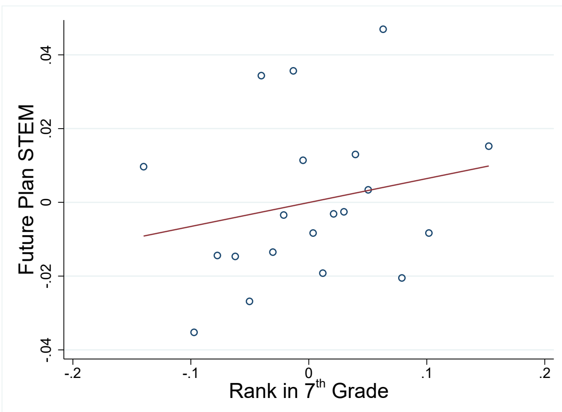
Figure 2.4: Relationship between Rank and STEM Outcomes in High School



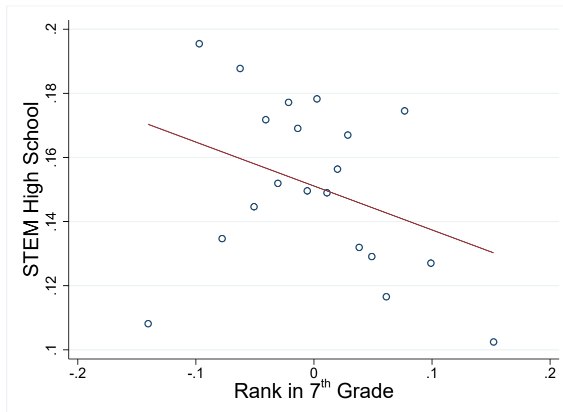
(a) Panel A: STEM Track



(b) Panel B: Advanced Math



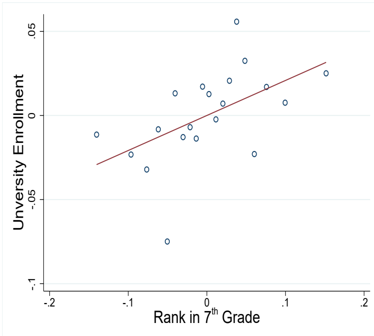
(c) Panel C: Future Plan STEM



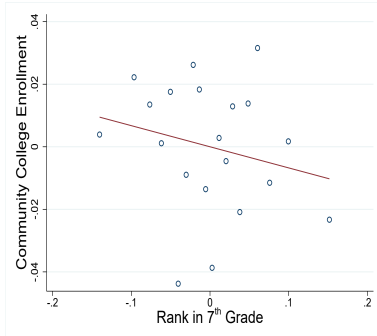
(d) Panel D: STEM High School

Notes: This figure presents the correlation between class rank in 7^{th} grade on the x-axis and STEM outcomes in high school on the y-axis. These STEM outcomes include indicators for (a) the choice of the math–science track in 11^{th} grade, (b) taking at least one advanced math course required for admission to a STEM postsecondary degree program, (c) the self-reported desire to seek a STEM postsecondary degree conditional on having decided a major, and (d) attending a STEM-oriented high school. Both variables are residualized with respect to a third-order polynomial of a student's 7^{th} grade test scores and classroom–subject fixed effects.

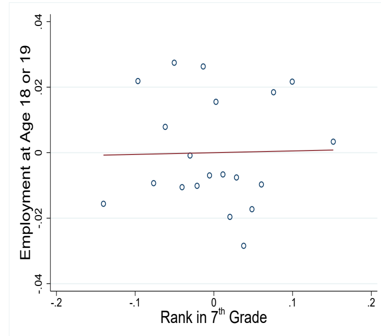
Figure 2.5: Relationship between Rank and Postsecondary Outcomes



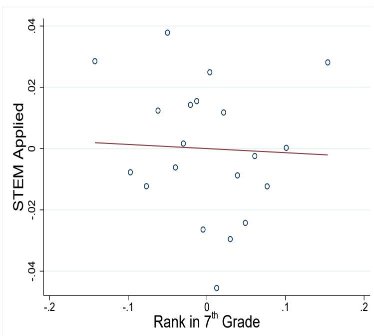
(a) Panel A: University Enrollment



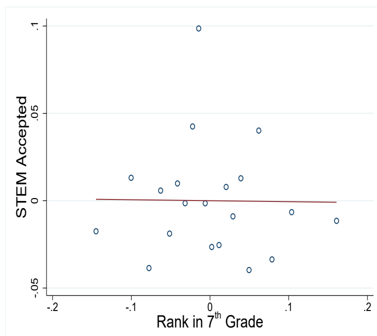
(b) Panel B: Community College Enrollment



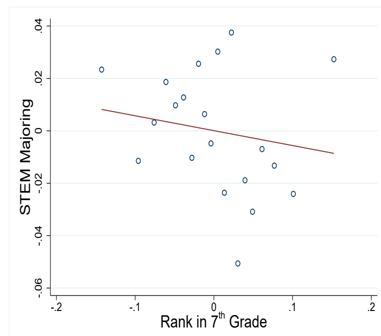
(c) Panel C: Employed at Age 18 or 19



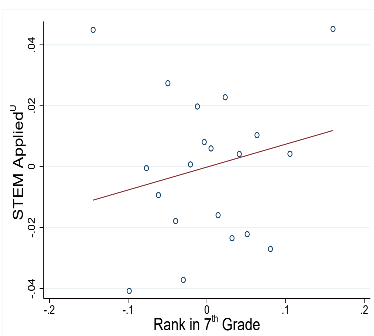
(d) Panel D: STEM Applied



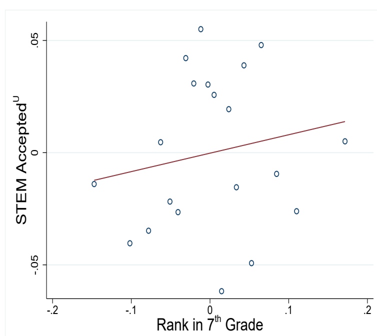
(e) Panel E: STEM Accepted



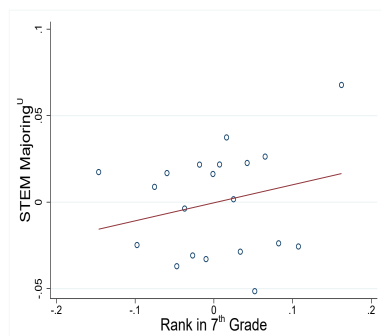
(f) Panel F: STEM Majoring



(g) Panel G: STEM Applied^U



(h) Panel H: STEM Accepted^U



(i) Panel I: STEM Majoring^U

Notes: This figure presents the correlation between class rank in 7^{th} grade on the x-axis and postsecondary outcomes on the y-axis. These postsecondary outcomes include the indicators for (a) attending a university, (b) attending a community college, (c) being employed at age 18 or 19, (d) applying for a STEM major, (e) being accepted conditional on applying for a STEM major, (f) majoring in a STEM field during a university or community college, (g) applying to a university for majoring in a STEM field, (h) being accepted conditional on applying to a university for majoring in a STEM field, and (i) majoring in a STEM field during a university. Both variables are residualized with respect to a third-order polynomial of a student's 7^{th} grade test scores and classroom–subject fixed effects.

Overall, class rank graphically leads to changes in the majority of a student's outcomes. I now turn to estimating Equations (2.2) and (2.3) to obtain rank parameters in the following sections.

2.4.2 Short-term Effects on Test Scores

The test score is the outcome of main interest. The effects of class rank in 7^{th} grade on test scores in 8^{th} grade are presented in Table 2.5. Column (1) shows that rank at the top of the class compared to the bottom, conditional on achievement and classroom by subject fixed effects, will improve the test score by 21.8% of a standard deviation. Controlling for student and teacher characteristics in Column (3) does not change the magnitudes of the point estimates by much. This suggests that conditional on the test score distribution and classroom by subject fixed effects, class rank effects are uncorrelated with observable student and teacher characteristics. Hence, identifying variation in class rank – which is obtained after controlling for a third-order function of students' test scores and classroom by subject fixed effects – seems to be quasi-random.

Table 2.5: Effects on a Student's Test Score in 8^{th} Grade

	(1)	(2)	(3)
7^{th} Grade Rank	0.218**	0.248**	0.227*
	(0.108)	(0.108)	(0.133)
Observations	12,913	11,852	8,577
R^2	0.528	0.531	0.532
Cubic in 7^{th} Grade Test Score	Y	Y	Y
Classroom by Subject Fixed Effects	Y	Y	Y
Student Controls	N	Y	Y
Teacher Controls	N	N	Y

Notes: Each column represents results from a separate regression. Student controls are indicators for student gender, for a single-parent household, number of siblings, and indicators for having at least one parent with a BA degree or higher and for free lunch. Teacher controls are indicators for teacher gender, for teacher age over 40, for teacher's graduation from teachers college, for teacher having a Master's degree or higher, and for administrative teacher. Standard errors, shown in parentheses, are clustered at the classroom level.
*p<0.10, ** p<0.05, *** p<0.01.

In this analysis, I include a cubic for the test scores in the specification (Equation (2.2)). However, if the (unobserved) misspecification error is correlated with class rank, then my estimate for the rank effects would be biased ((2); (26)). To check the robustness of my rank parameter (in Column (1) of Table 2.5), I estimate five specifications with increasingly higher-order polynomials for the test scores in 7^{th} grade. The estimates are shown in Appendix Table D1. Once a cubic function in the test score used in Equation (2.2) is included, utilizing additional polynomials does not make a significant difference to the estimated coefficients. A misspecified distribution function of the test score in 7^{th} grade would not threaten my estimate for the rank effect in Table 2.5.

Measurement error is another problem that would produce biased rank parameters. There are two types of errors in test scores.[10] The first one is a systematic error in test scores due to the situation that both teacher and peer effects affect the test scores but not the rank. I confirm that this kind of measurement error would not be problematic through the data-driven evidence in Appendix B.

The second one is a non-systematic error where a noisy measure of the test score would create false effects of the rank on test scores. For example, if students get a mistakenly high (or low) test score, they also have a falsely high (or low) rank. This situation would lead them to have lower (or higher) observed growth in test scores, so the estimated coefficient on rank effect would be downward biased. To assess the importance of this problem, I perform Monte Carlo experiments.[11] I find that allowing additional measurement error in test scores would non-linearly attenuate the estimated coefficient (in Column (1) of Table 2.5) toward zero. However, the magnitude of the downward bias does not seem to be large. I therefore conclude that the main estimate is not substantially affected by such error.

2.4.3 Longer-term Effects on Test Scores

The longer-term effects of class rank in 7^{th} grade on test scores, which last for five years, are presented in Table 2.6. Panel A shows the results by estimating Equation (2.4). Panel B presents

[10] Existing literature ((24); (2), (26)) has addressed the importance of the impact of measurement error in terms of the identification of the rank effects. Here, I refer to the classification of (2) to define the measurement error problem.

[11] Details for the Monte Carlo exercises are provided in Appendix A.

the results by estimating the specification that includes the student and subject teacher controls in Equation (2.3). Because my identifying assumption has proven to be true in subsection 2.3.3, I mostly focus on the results of Panel A, which relies on the most parsimonious specification. The estimates in Panel B complement these results by showing the robustness of the estimated rank parameters.

Table 2.6: Effects on Test Scores Over Time

	(1) Test Score 8th Grade	(2) Test Score 9th Grade	(3) Test Score 10th Grade	(4) Test Score 11th Grade	(5) Test Score 12th Grade
A. Controls for Classroom by Subject FEs					
7^{th} Grade Rank	0.218**	0.419***	0.488***	0.688***	0.378**
	(0.108)	(0.141)	(0.164)	(0.168)	(0.156)
Observations	12,913	12,312	10,247	10,085	9,733
R^2	0.528	0.261	0.195	0.247	0.137
P-Value	-	0.00	0.00	0.00	0.01
B. Controls for Classroom by Subject FEs, Stu Ctrls & Tch Ctrls					
7^{th} Grade Rank	0.227*	0.469***	0.451**	0.528**	0.336*
	(0.133)	(0.165)	(0.220)	(0.233)	(0.172)
Observations	8,577	8,174	6,751	6,633	6,393
R^2	0.532	0.306	0.219	0.285	0.161
P-Value	-	0.00	0.00	0.00	0.03

Notes: Each cell represents results from a separate regression including cubic in 7^{th} grade test score. Dependent variables in Columns (1)–(5) are standardized test scores within a subject and a year, from 8^{th} through 12^{th} grade, respectively. The P-value represents the significance of the difference between the point estimate in 8^{th} grade and that in each grade. Student controls are indicators for student gender, for a single-parent household, number of siblings, and indicators for having at least one parent with a BA degree or higher and for free lunch. Teacher controls are indicators for teacher gender, for teacher age over 40, for teacher's graduation from teachers college, for teacher having a Master's degree or higher, and for administrative teacher. Standard errors, shown in parentheses, are clustered at the classroom level.
*p<0.10, ** p<0.05, *** p<0.01.

I find that a positive effect of class rank in 7^{th} grade on test scores continuously exists for five years. The point estimates in Panel A show that moving from the bottom of the class to the top, conditional on achievement and classroom by subject fixed effects, improves test scores by

21.8% (8^{th} grade), 41.9% (9^{th} grade), 48.8% (10^{th} grade), 68.8% (11^{th} grade), and 37.8% (12^{th} grade), respectively. The first four estimates among these estimates suggest that the impact of class rank in 7^{th} grade will be larger over time. In addition, there are statistically significant differences between the rank effect on the test score in 8^{th} grade and the remainders in the following years. This is supported by the p-values for differences between each coefficient in Columns (2)–(5) and that in Column (1). Column (5) shows that the rank effect on the test score largely fades out after 11^{th} grade, but is still statistically significantly greater than the effect in 8^{th} grade (p-value=0.01). In addition, I limit myself to the balance sample and present the result in Panel B of Appendix Table F5. Doing so does not change the main estimates (Panel A of Table 2.6) significantly.

My findings above reinforce the existing literature ((21); (2); (26)) that documents a significant and positive effect of students' rank on later test scores (Appendix Table F1). In particular, the results obtained by (2) suggest that a one standard deviation increase in rank at the end of primary school improves test scores three years later by around 0.08 standard deviations or five years later by around 0.07 standard deviations. These effects are smaller in the magnitude than my effects that a one standard deviation (0.30) increase in rank in middle school (7^{th} grade) raises the 10^{th} grade test score by around 0.15 ($0.30 \times 0.49 \doteqdot 0.15$) standard deviations or the 12^{th} grade test score by around 0.11 ($0.30 \times 0.38 \doteqdot 0.11$) standard deviations.

One may be concerned that these results could be driven by teacher or peer effects that generate false rank effects. If these effects have a temporary impact on the test score in 7^{th} grade but do not change the class rank, then the effects of class rank that might pick up information related to true ability rank on the test score would be spurious.[12] Appendix Section B addresses this issue and presents a set of tests that yields evidence against such effects. The results of those tests show that the false rank effects by teachers or peers might not be significantly driving the main results in

[12] As one of the probable teacher effects, I imagine that an English teacher improves all students' test scores by 5 points on average in English class. As a result, students A and B obtain scores 90 and 65 in that grade, respectively. Based on these scores, those students have the corresponding ordinal rankings as 1 and 20 in that class. However, these rankings can mostly reflect the true ability rankings. I now assume that the teacher will no longer teach them in later years; hence, students' true ability will significantly affect their test score growth. Perhaps student A will continuously get a high English score, whereas student B will experience a deep fall in that score. This might lead me to observe a falsely positive impact of class rank on future test scores during the analysis.

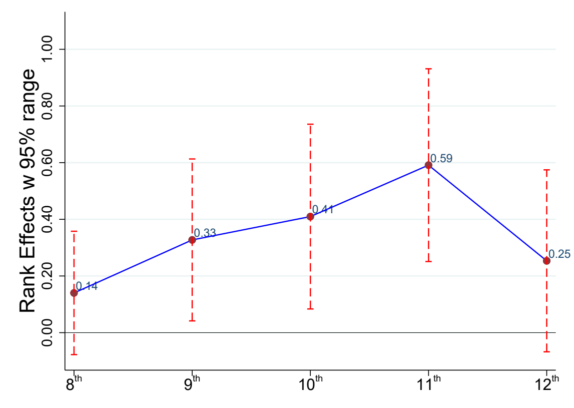
Panel A of Table 2.6.

Next, I investigate the extent to which the effects of class rank in 7^{th} grade on students' test scores might vary by student gender, class size, students' relative rank position, and household income. To do so, I allow the rank effects to be heterogeneous by student gender, class size, students' relative rank position, and household income in Equation (2.3).[13] The results are shown in Figures 2.6–2.9. Each figure displays the point estimates and 95%–confidence intervals

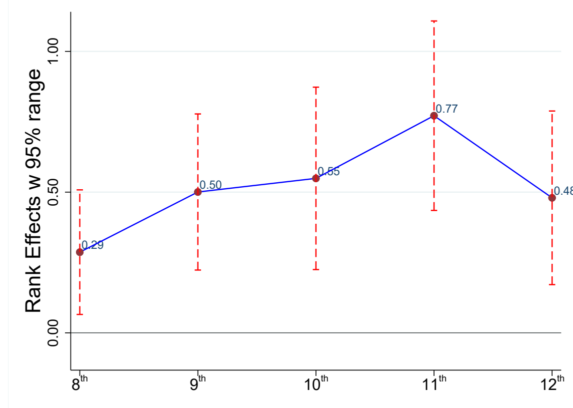
I begin with the heterogeneous effects of class rank by student gender. In Figure 2.6, the point estimates show that female students' test scores are more affected by their class rank than are male students from 8^{th} to 12^{th} grade, though the effect of class rank is not statistically significant for male students in 8^{th} and 12^{th} grade. In terms of statistical significance, these coefficients are different from each other. This finding might be different from some of the previous studies ((2); (26)) documenting that achievement or ability rank effects on test scores are more noticeable for male students than female students.

[13]Specifically, I perform this analysis by interacting the rank variable with dummies for these characteristics.

Figure 2.6: Rank Effects by Student Gender



(a) Panel A: Male Student

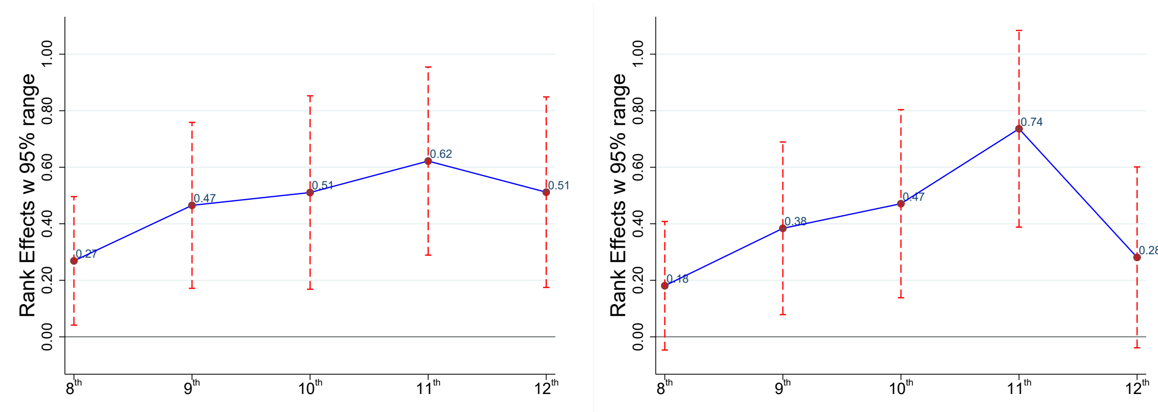


(b) Panel B: Female Student

Notes: These graphs depict estimated effects of class rank in 7^{th} grade on test scores by student gender. The numbers labelled above the circles are the coefficients of class rank. 95% confidence intervals are shown with dashed lines. The p-values of the F-test for the difference between the coefficients in each grade are 0.001 (8^{th} grade), 0.000 (9^{th} grade), 0.000 (10^{th} grade), 0.000 (11^{th} grade), and 0.000 (12^{th} grade), respectively.

I present the results of the heterogeneous effects by class size in Figure 2.7. If students in a relatively larger class are less aware of their achievement rank than those in a relatively smaller class, then the rank effects would be more pronounced for students in a relatively smaller class. For doing this, I first differentiate classes based on their median class size in the sample. The point estimates indicate that the students in the class below the median size are more likely to be affected by rank than those in the class above the median size from 9^{th} to 10^{th} grade. The effect in the class above the median size is temporarily larger than that in the class below the median size in 11^{th} grade, but the coefficients are not statistically significantly different from each other.

Figure 2.7: Rank Effects by Class Size



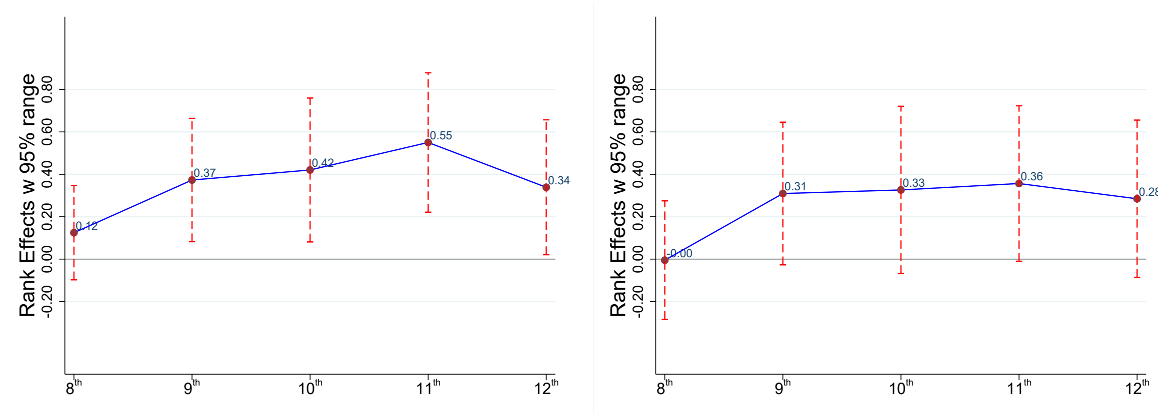
(a) Panel A: Below Median Class Size

(b) Panel B: Above Median Class Size

Notes: These graphs depict estimated effects of class rank in 7^{th} grade on test scores by class size. The numbers labelled above the circles are the coefficients of class rank. 95% confidence intervals are shown with dashed lines. The p-values of the F-test for the difference between the coefficients in each grade are 0.059 (8^{th} grade), 0.008 (9^{th} grade), 0.013 (10^{th} grade), 0.221 (11^{th} grade), and 0.006 (12^{th} grade), respectively.

In addition, I estimate the effects for students who fall into the high and low positions of class rankings distribution in 7^{th} grade separately. I define these rank positions as the rank within the top 50^{th} percentile and the bottom 50^{th} percentile. Figure 2.8 shows that a positive impact of class rank in 7^{th} grade on test scores in 8^{th} grade is more pronounced for students who are located in the high positions of the rank distribution than those who fall into low positions. This pattern seems to be persistent until 11^{th} grade. However, all coefficients are statistically insignificant or weak for students in the low positions.

Figure 2.8: Rank Effects by Relative Rank Position



(a) Panel A: High Rank Position

(b) Panel B: Low Rank Position

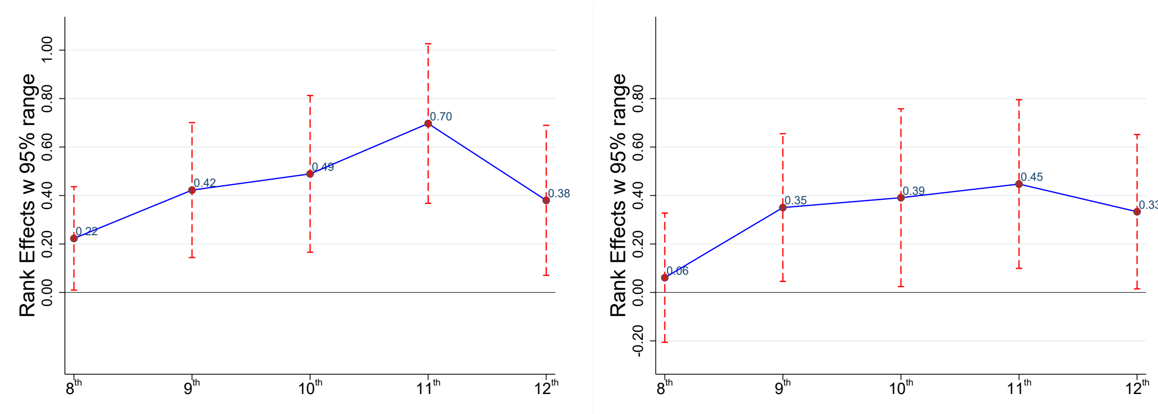
Notes: These graphs depict estimated effects of class rank in 7^{th} grade on test scores by a student's relative rank position. The high and low ranks positions represent the class rank within the top 50^{th} percentile and the bottom 50^{th} percentile. The numbers labelled above the circles are the coefficients of class rank. 95% confidence intervals are shown with dashed lines. The p-values of the F-test for the difference between the coefficients in each grade are 0.029 (8^{th} grade), 0.212 (9^{th} grade), 0.133 (10^{th} grade), 0.002 (11^{th} grade), and 0.400 (12^{th} grade), respectively.

I show the differential effects of class rank by household income in Figure 2.9. A student's household income is proxied by whether he or she receives free school meals.[14] That is, I consider students who get free school meals as those in the low-income household. Panel A shows rank parameters for students who receive free school meals while Panel B shows rank parameters for students who do not. For the test score in 8^{th} grade, the impact is larger in students in the low-income household; however, the effect in the high-income household is not statistically significant. The impact of class rank on test scores is stronger for students in the low-income household from 8^{th} to 11^{th} grade. In addition, the differences between rank effects in these grades are statistically significant at the 1% level. This finding might be similar to (2) and (21) who find that a free school lunch student gains more from being highly ranked in the UK and US.[15]

[14]In 2010, South Korean middle school students received free school meals if their household income was low or below the poverty line.

[15](21) document that this result might be because students in the more disadvantaged group have low self-

Figure 2.9: Rank Effects by Household Income



(a) Panel A: Free School Meal

(b) Panel B: No Free School Meal

Notes: These graphs depict estimated effects of class rank in 7^{th} grade on test scores by whether a student receive free school meals. The numbers labelled above the circles are the coefficients of class rank. 95% confidence intervals are shown with dashed lines. The p-values of the F-test for the difference between the coefficients in each grade are 0.005 (8^{th} grade), 0.005 (9^{th} grade), 0.005 (10^{th} grade), 0.000 (11^{th} grade), and 0.054 (12^{th} grade), respectively.

2.4.4 Effects on STEM Outcomes in High School and Postsecondary Outcomes

STEM (Science, Technology, Engineering, and Math) education is an important contributor to innovation and productivity growth (36). Therefore, studying on how a student's class rank for math in 7^{th} grade will affect STEM-related paths and decisions later is particularly intriguing.

I examine the effects of students' class rankings in 7^{th} grade on STEM outcomes that are self-reported by them in 11^{th} grade and present the results in Table 2.7. In Panel A, the specification, like Equation (2.3), is used for the estimation. Column (1) of Panel A shows that class rankings in 7^{th} grade will positively affect a student's math–science track choice in 11^{th} grade. This effect is substantial but just statistically significant at the 10% level. These rankings will also substantially reduce the likelihood of students' STEM–focused high school attendance (Column (4)) and this effect is statistically significant at the 5% level. This result might reflect that a student with a higher

confidence or a different information set about academic achievements, and weight their school experience more heavily than students in the less disadvantaged group.

rank would be less likely to attend a STEM–focused high school because a STEM–focused high school is usually worse than a general academic high school in terms of school quality in South Korea. These rankings will not have statistically significant effects on taking at least one advanced math course (Column (2)) or being interested in majoring in a STEM field during postsecondary periods (Column (3)).

Table 2.7: Effects on STEM Outcomes

	(1) STEM Track	(2) Advanced Math	(3) Future Plan STEM	(4) STEM High School
A. Main Effects				
7^{th} Grade Rank	0.145*	0.133	0.069	-0.122**
	(0.082)	(0.084)	(0.104)	(0.057)
Observations	8,562	8,615	6,213	9,760
R^2	0.096	0.088	0.106	0.129
Dep.Var.Mean	0.313	0.252	0.332	0.151
B. Heterogeneous Effects by Subject				
7^{th} Grade Math Rank	0.406***	0.358**	0.371**	-0.096
	(0.155)	(0.142)	(0.182)	(0.097)
7^{th} Grade Eng Rank	-0.029	0.151	-0.178	-0.074
	(0.112)	(0.119)	(0.143)	(0.094)
7^{th} Grade Kor Rank	-0.045	-0.164	-0.042	-0.098
	(0.103)	(0.125)	(0.143)	(0.087)
Observations	2,831	2,848	2,055	3,229
R^2	0.151	0.107	0.161	0.150
Dep.Var.Mean	0.313	0.251	0.332	0.151
Original P-Value for Math Rank	0.004	0.007	0.027	0.331
Adj. P-Value (*Wyoung*) for Math Rank	0.014	0.019	0.049	0.297
Odds Ratio for Math Rank	5.436*	7.589***	5.273*	0.385
	(4.749)	(5.961)	(5.071)	(0.392)

Notes: Each cell represents results from a separate regression including cubic in 7^{th} grade test score and classroom by subject fixed effects (Panel A), or classroom fixed effects (Panel B). In each Column, the dependent variables are indicators for (1) the choice of the math–science track in 11^{th} grade, (2) taking at least one advanced math course required for admission to a STEM postsecondary degree program, (3) the self-reported desire to seek a STEM postsecondary degree conditional on having decided a major, and (4) attending a STEM-oriented high school. In Panel B, I present the results for testing multiple hypotheses on the effects of ordinal rank in mathematics. For doing this task, I calculate family-wise adjusted p-values based on 1,000 bootstraps clustering standard errors at the classroom level for Westfall and Young correction. Odds ratios for math rank are earned by logit specifications. Standard errors, shown in parentheses, are clustered at the classroom level.
*p<0.10, ** p<0.05, *** p<0.01.

I turn to the results in Panel B of Table 2.7. Here, I break down the rank effects by subject. While decreased sample sizes lead to the loss of statistical power, these results show that a student with a higher math rank in 7^{th} grade will be more likely to choose the math–science track (Column (1)) and at least one advanced math course (Column (2)). It also shows the interest in majoring in a STEM field in college (Column (3)). These effects are substantial and statistically significant at the conventional levels. A higher math rank in 7^{th} grade will reduce the likelihood of a STEM-focused highs school attendance, though the corresponding coefficient is imprecisely estimated. I find no evidence of statistically meaningful patterns in other subjects. Because the four outcomes in Columns (1)–(4) are closely related, I additionally perform the task to verify if the estimated coefficients on math rank are still statistically significant with multiple hypothesis testing correction. Given the p-values, listed in the bottom of line of the table, I find that those coefficients in Columns (1)–(3) are still statistically significant at conventional levels when the correction is applied.[16] In the previous section, I found no statistically significant effect of math rank in 7^{th} grade on math scores in later years. However, this rank will substantially affect the changes in STEM-related outcomes that are important to determine a career path. Interestingly, these findings suggest that a higher math rank plays a significant role in preparing for postsecondary STEM study, though I cannot strongly conclude its effect on the math score growth. I also report odds ratios for math rank from logit estimation at the bottom of the table. In each logit specification, I include ranks in each subject and classroom fixed effects. Odds ratios imply that moving from the bottom to the top in the distribution of math rank increases the probabilities of choosing the math–science track, selecting at least one advanced math course, or showing interest in majoring in a STEM field in college by 443% (Column (1)), 658% (Column (2)), or 527% (Column (3)). These odd ratios are statistically significant at conventional levels. However, doing so decreases the likelihood of a STEM-focused highs school attendance by 61% (Column (4)). The corresponding odds ratio is not statistically significant. Overall, I conclude that my OLS results for math rank effects are robust in

[16] I utilize the method of (37). This methodology adjusts p-values for multiple hypotheses using randomization or bootstrap inference to calculate the joint distribution of p-values inference and allows me to check the robustness of the results to randomization inference.

the direction of the effects to those from logit estimation.[17]

Next, I explore the rank effects on postsecondary outcomes. Regarding these outcomes, a follow-up survey was conducted two months after the students left high school. The results are presented in Table 2.8. Columns (1)–(3) show the results of the effects of class rank in 7^{th} grade on the probability of attending a university, community college, or being employed at age 18 or 19. Columns (4)–(6) show the rank effects on the likelihood that students applied to a STEM major, the likelihood of being accepted conditional on applying to a STEM field, and the likelihood that students actually major in a STEM field. Columns (7)–(9) show the rank effects on the likelihood that students applied to a university in a STEM field, the likelihood of being accepted conditional on applying to a university for majoring in a STEM field, and the likelihood that students actually major in a STEM field during a university. Panel A of the table shows the positive effect of class rank in 7^{th} grade on the probability for attending a university.[18] Column (2) reports the negative effect of class rank in 7^{th} grade on the probability for attending a community college. This effect is not statistically distinguishable from zero. In Column (3), I find no meaningful effects of ordinal rank in 7^{th} grade on being employed at age 18 or 19: all estimated coefficients are statistically insignificant. Moreover, the estimated coefficient is very small. Columns (4)–(6) show that a higher class rank in 7^{th} grade will have negative effects on STEM-related outcomes, but these effects are small and statistically insignificant. In Columns (7)–(9), the corresponding estimates change to be positive and fairly sizable. However, these estimated coefficients are not statistically significant.

[17] My logit estimates are much greater than OLS estimates in terms of the magnitude. For instance, the logit estimate for the probability of choosing the math–science track is 1.693 (s.e.=0.874), which is forth times more effect size than the OLS estimate (0.406). Other three logit estimates show the similar pattern. Because my logit specification includes the classroom fixed effects, the results from logit estimation suffer from the incidental parameter problem in the process of the (maximum likelihood) estimation ((38); (39)). Hence, I still prefer OLS estimates over logit estimates in terms of the results.

[18] I am agnostic on the exact reason why class rank in middle school has a positive effect on the likelihood of attending a university later. However, this result might be associated with the persistence of its impact on test scores, self-confidence, and study effort. Specifically, by estimating Equation (2.3), I find that the estimates for rank effects on test scores, self-confidence, and study effort in 12^{th} grade are 0.038 (s.e.=0.008), 0.040 (s.e.=0.011), and 0.076 (s.e.=0.010), respectively.

Table 2.8: Effects on Postsecondary Outcomes

	(1) University	(2) Community College	(3) Employed	(4) STEM Applied	(5) STEM Accept	(6) STEM Majoring	(7) STEM Applied^U	(8) STEM Accept^U	(9) STEM Majoring^U
A. Main Effects									
7^{th} Grade Rank	0.219**	-0.071	0.006	-0.014	-0.007	-0.063	0.085	0.111	0.125
	(0.107)	(0.085)	(0.059)	(0.110)	(0.179)	(0.134)	(0.147)	(0.230)	(0.199)
Observations	6,443	6,443	6,443	5,555	2,508	4,233	3,827	1,848	2,627
R^2	0.149	0.120	0.116	0.106	0.183	0.140	0.136	0.235	0.210
Dep.Var.Mean	0.407	0.247	0.080	0.392	0.796	0.385	0.409	0.756	0.393
B. Heterogeneous Effects by Subject									
7^{th} Grade Math Rank	0.126	0.122	-0.062	0.234	0.268	0.260	0.489**	0.464	0.555**
	(0.180)	(0.145)	(0.119)	(0.198)	(0.314)	(0.211)	(0.229)	(0.434)	(0.270)
7^{th} Grade Eng Rank	0.011	0.180	0.020	-0.053	-0.149	-0.216	-0.003	-0.214	-0.085
	(0.148)	(0.143)	(0.095)	(0.164)	(0.246)	(0.208)	(0.189)	(0.294)	(0.266)
7^{th} Grade Kor Rank	0.252*	-0.352***	0.081	-0.167	-0.086	-0.111	-0.180	0.046	-0.018
	(0.146)	(0.114)	(0.076)	(0.179)	(0.253)	(0.206)	(0.216)	(0.289)	(0.266)
Observations	2,130	2,130	2,130	1,836	828	1,398	1,267	611	869
R^2	0.183	0.146	0.131	0.133	0.201	0.167	0.191	0.265	0.267
Dep.Var.Mean	0.407	0.247	0.080	0.392	0.796	0.385	0.409	0.756	0.393

Notes: Each cell represents results from a separate regression including cubic in 7^{th} grade test score and classroom by subject fixed effects (Panel A) or classroom fixed effects (Panel B). In each column, dependent variables are the indicators for (1) attending a university, (2) attending a community college, (3) being employed at age 18 or 19, (4) applying for a STEM major, (5) being accepted conditional on applying for a STEM major, (6) majoring in a STEM field during a university or community college, (7) applying to a university for majoring in a STEM field, (8) being accepted conditional on applying to a university for majoring in a STEM field, and (9) majoring in a STEM field during a university. Standard errors, shown in parentheses, are clustered at the classroom level.
*p<0.10, ** p<0.05, *** p<0.01.

I further investigate the rank effects on the same outcomes by breaking them down by subject again. The results are shown in Panel B of Table 2.8. Large decreases in the sample size lead to reduced statistical power. In Columns (1) and (2), Korean rank in 7^{th} grade will increase the likelihood of attending a university, whereas it will decrease the likelihood of attending a community college. Math rank in 7^{th} grade will increase the probability of attending both a university and a community college, but these effects are statistically insignificant. In Column (3), English and Korean ranks in 7^{th} grade have a positive effect on being employed at age 18 or 19. The corresponding coefficients are imprecisely estimated. In Columns (4)–(6), the positive effects of math rank in 7^{th} grade on STEM-related outcomes are substantial. However, these effects are not statistically significant. Columns (7) and (8) show that math rank in 7^{th} grade will substantially increase the likelihood of applying a STEM field in a university and being subsequently accepted. However, the math rank effect on being accepted is not statistically significant (Column (8)). Finally, Column (9) shows math rank in 7^{th} grade has a positive and significant impact on majoring in a STEM field while attending a university.

2.5 Potential Mechanisms

In combination, my results show a persistent impact of class rank in 7^{th} grade on a student's later outcomes during secondary school, along with a substantially positive effect of math rank in 7^{th} grade on preparation for postsecondary STEM study. A higher class rank in 7^{th} grade will increase the likelihood that students would enroll in a university. There are significant effects of math rank in 7^{th} grade on applying to a STEM major in a university and actually majoring in a STEM field. This implies that improved STEM outcomes during high school due to a higher math rank may generate meaningful changes in actual outcomes in higher education. However, the underlying mechanisms still remain opaque. I examine a number of the potential mechanisms that might explain why these results emerge over time, including student-, teacher-, and parent-oriented driving forces.

2.5.1 Self-Confidence, Self-Expectation, Class Engagement, and Effort

A higher rank may raise a student's self-confidence in their studies or self-expectation about their future educational attainment. I first test this possibility by examining whether class rank in 7^{th} grade has a positive lasting effect on a student's self-confidence. In the SELS2010 data, students are asked to report their level of agreement to the following statement: "I am able to understand difficult things in my studies." They can select a response from the following Likert-type scale: 1."Completely Disagree," 2. "Disagree," 3. "Normal," 4."Agree," 5."Completely Agree." Panel A of Table 2.9 shows that the effect on students' self-confidence is positive and statistically significant in 9^{th} grade. This positive effect exists until 12^{th} grade.

I next examine whether rankings in 7^{th} grade have similar effect on a student's self-expectation about attending college. Panel B of Table 2.9 shows positive but relatively small effects on this outcome. The estimated coefficients are mostly statistically insignificant; thus, I find no evidence that this is driving my results in a meaningful way. In addition, I explore how the rank effects will change students' class engagement over time. I construct a series of class engagement indices for each year by using principal component analysis.[19] Panel C of Table 2.9 shows that the majority of the point estimates are positive and substantial, but statistically indistinguishable from zero.

Finally, a student's higher rank may subsequently trigger more efforts in their studies. I make a series of student effort indices by using principal component analysis.[20] Panel D of Table 2.9 shows the effects of class rank in 7^{th} grade on the study effort index. The effect is positive and substantial in 8^{th} grade. This effect persists until 12^{th} grade, though the point estimate temporarily loses statistical significance in 11^{th} grade.

[19] To construct this index, I use the student's responses to four questions: 1) whether he or she is focused in class, 2) whether they participate in class, 3) whether they skip homework assignments or not, and 4) whether they review what they learned in class. Students can select a response from the following: 1."Completely Disagree," 2. "Disagree," 3. "Normal," 4."Agree," 5."Completely Agree."

[20] This index is generated, based on a student's responses to three questions: 1) whether he or she does their best to completely understand the class materials, 2) whether they do their best to study at maximum, and 3) whether they endeavor to keep their study schedule. Students can select a response from the following: 1."Completely Disagree ," 2. "Disagree," 3. "Normal," 4."Agree," 5."Completely Agree."

Table 2.9: Effects on Confidence, Expectation, Engagement, and Effort

	(1) 8th Grade	(2) 9th Grade	(3) 10th Grade	(4) 11th Grade	(5) 12th Grade
A. Self-Confidence					
7^{th} Grade Rank	0.078	0.453***	0.298*	0.322*	0.397**
	(0.159)	(0.162)	(0.171)	(0.167)	(0.168)
Observations	12,841	12,368	10,331	10,082	9,817
R^2	0.058	0.136	0.092	0.111	0.108
Dep.Var.Mean	3.748	3.323	3.334	3.332	3.385
B. Self-Expectation					
7^{th} Grade Rank	0.016	-	0.051	0.135*	0.084
	(0.052)	-	(0.069)	(0.074)	(0.076)
Observations	12,778	-	10,048	10,057	9,803
R^2	0.112	-	0.157	0.146	0.151
Dep.Var.Mean	0.849	-	0.757	0.764	0.804
C. Class Engagement					
7^{th} Grade Rank	0.224	0.428*	0.362	0.254	0.360
	(0.204)	(0.231)	(0.255)	(0.272)	(0.323)
Observations	12,765	12,375	10,278	10,072	9,823
R^2	0.166	0.170	0.138	0.132	0.092
Dep.Var.Mean	0.001	0.001	0.002	0.001	0.006
D. Study Effort					
7^{th} Grade Rank	0.461**	0.615***	0.420**	0.361	0.716***
	(0.200)	(0.203)	(0.207)	(0.249)	(0.264)
Observations	12,847	12,347	10,319	10,093	9,808
R^2	0.116	0.130	0.123	0.132	0.127
Dep.Var.Mean	0.001	0.001	0.002	0.000	0.002

Notes: Each cell represents results from a separate regression, including cubic in 7^{th} grade test score and classroom by subject fixed effects. Dependent variables in Columns (1) through (5) are students' reported self-confidence (Panel A), self-expectation (Panel B), class engagement index (Panel C), and student effort (Panel D) for 8^{th} through 12^{th} grades. A student's response to his or her self-confidence is consisted of the five levels (1. completely disagree; 2. disagree; 3. normal; 4. agree; 5. completely agree) of agreement to the statement "I am able to understand difficult things in my studies." A student's self-expectation is the indicator for being interested in attending college. The class engagement index is constructed by principal component analysis that extracts common variation for each year based on the student's response to four questions: 1) whether he or she is focused in class, 2) whether they participate in class, 3) whether they do not skip homework assignments, and 4) whether they review what they learned in class. The student effort index is also constructed by principal component analysis that extracts common variation for each year based on the students' responses to three questions: 1) whether they do the best to completely understand the class materials, 2) whether they do the best to study at maximum, 3) whether they endeavor to keep their study schedule. *p<0.10, ** p<0.05, *** p<0.01.

2.5.2 Teacher Behavior

Another possible mechanism for the persistent effect of rank is that the student's rank may change teachers' behaviors. The SELS2010 data asked students to report if their subject teachers (for mathematics/English/ Korean) teach them well, present information in a way that is easy to understand, check the homework thoroughly, and ask them if they understand the lessons thoroughly. Using these questions, I construct an index of teacher behavior through principal component analysis.

Table 2.10 shows that most of the rank effects on teacher behavior indices are substantial. However, all coefficients from 8^{th} to 12^{th} grade are imprecisely estimated; thus, I cannot strongly conclude that the persistent effects on the teacher behavior index drive the main results.

Table 2.10: Effects on Teacher Behavior

	(1) 8th Grade	(2) 9th Grade	(3) 10th Grade	(4) 11th Grade	(5) 12th Grade
7^{th} Grade Rank	0.108	0.381	0.312	0.361	0.316
	(0.230)	(0.287)	(0.255)	(0.265)	(0.299)
Observations	12,809	12,368	10,199	10,030	9,823
R^2	0.116	0.109	0.076	0.080	0.057
Dep.Var.Mean	0.002	0.001	0.003	0.002	0.006

Notes: Each column represents results from a separate regression, including cubic in 7^{th} grade test score and classroom by subject fixed effects. Dependent variables in Columns (1) through (5) are standardized teacher behavior indexes for 8^{th} through 12^{th} grades. The teacher behavior index is constructed by principal component analysis that extract common variation for each subject and each year based on the students response to seven questions: 1) whether a teacher has enthusiasm towards class, 2) whether a teacher has enough knowledge about the class, 3) whether a teacher provides an adequate-level lecture, 4) whether a teacher expects students to study hard, 5) whether a teacher expects students to have a high academic performance, 6) whether a teacher checks out students' homework, and 7) whether a teacher checks students' level of understanding the lecture.
*p<0.10, ** p<0.05, *** p<0.01.

2.5.3 Parental Investment and School Choice

Parental response to a child's education may vary with their academic rank in school. I use information on parental investment behavior from the SELS2010 data. Specifically, such information includes the levels of parents' fostering a learning environment at home, checking homework, directly assisting, advising study methods, utilizing private tutoring, spending money on studies, managing the schedule, and encouraging studies. Based on these behaviors, I create an index of parental investment by using the first components from the principal component analysis. Table 2.11 shows that class rankings in 7^{th} grade will substantially increase parental investment even until high school (11^{th} grade). Although the point estimate in the short-run (8^{th} grade) is imprecisely estimated, the estimated size of this coefficient is large. However, the negative coefficient for the rank effect in (12^{th} grade) indicates that parents reduce investments if their child has a higher rank among classmates. This coefficient is statistically insignificant but fairly sizable. One possible explanation for this effect is that parents who prepare for their children's college admissions in 12^{th} grade strategically invest the weaker subject more; thus, the rank effect on parental investment would be reversed between 11^{th} and 12^{th} grade.

Table 2.11: Effects on Parental Investment

	(1) 8th Grade	(2) 9th Grade	(3) 10th Grade	(4) 11th Grade	(5) 12th Grade
7^{th} Grade Rank	0.503	0.837***	0.834**	0.697**	-0.131
	(0.320)	(0.274)	(0.327)	(0.336)	(0.416)
Observations	11,764	11,177	9,323	9,204	8,929
R^2	0.121	0.114	0.095	0.094	0.089
Dep.Var.Mean	-0.001	0.002	-0.001	0.002	0.002

Notes: Each column represents results from a separate regression, including cubic in 7^{th} grade test score and classroom by subject fixed effects. Dependent variables in Columns (1) through (5) are standardized parental investment indexes for 8^{th} through 12^{th} grades. The parental investment index is constructed by principal component analysis that extracts common variation for each year based on the parents response to eight questions: 1) whether parents make the study atmosphere, 2) whether parents check out their homework, 3) whether parents directly teach them, 4) whether parents advise them about the study method, 5) whether parents collect information on private tutoring, 6) whether parents pay for their study, 7) whether parents manage their study schedule, and 8) whether parents encourage them to study hard.
*p<0.10, ** p<0.05, *** p<0.01.

Next, students can apply to their preferred high school in Seoul. Indeed, parents also intervene in their children's application process. A parent may therefore move his or her children with a higher rank to a more preferred or higher-quality school for better academic success. Also, the previous studies addressed the effectiveness of single-sex schools to get better academic performance for students in South Korea ((40); (41); (42)); thus, estimating the relationship between class rank and single-sex high school choices might be the task to catch the additional driver. Accordingly, in Table 2.12, I examine whether students attend in a more-preferred high school, whether they attended a single-sex high school, and whether they attended a higher-quality high school. Column (1) shows the positive effect of class rank on the likelihood that the student attends a first-round academic high school that is known to be prestigious is substantial and statistically significant. In Column (2), the positive impact on single-sex high school choices is not precisely estimated but fairly sizable. Column (3) shows the effect on the likelihood that a student attended high school, displaying above the median of the percent of 11^{th} grade students who obtained "Above Basic" grade in a given subject on the National Assessment of Educational Achievement (NAEA) test.

To avoid the influence of the sample students themselves on these grades, I use statistics for students from one cohort above those students. The higher class rank increases the attendance in a higher-quality school, but the impact seems to be small.

Table 2.12: Effects on School Choice

	(1) 1st Round HS	(2) Single-Sex HS	(3) Above Basic
7^{th} Grade Rank	0.186**	0.080	0.024***
	(0.081)	(0.079)	(0.007)
Observations	8,285	10,550	7,992
R^2	0.100	0.207	0.282
Dep.Var.Mean	0.215	0.544	0.934

Notes: Each column represents results from a separate regression, including cubic in 7^{th} grade test score and classroom by subject fixed effects. Dependent variables in Columns (1)–(3) are indicators for whether a student's high school is the high school in the first-round application, whether a student's high school is the single-sex school, and the percent of 11^{th} grade students with Above Basic Performance in National Assessment of Educational Achievement (NAEA) exam.
*p<0.10, ** p<0.05, *** p<0.01.

2.5.4 Discussion on Mechanisms

2.5.4.1 Probable Determinants of Class Rank

Regarding class (local) rank, a fundamental question asks, "What kinds of factors determine a higher rank in class?". While the definition of class rank directly or indirectly speaks to its correlations with students' cognitive ability or peer quality, I further examine household income and teacher gender effects.

First, prior work has documented the positive effects of household income on children's educational outcomes ((43); (44); (45)). Because class rank is closely associated with a student's academic performance, I investigate the possibility of a positive relationship between household income and class rank.[21] To do so, I use a rank-rank specification similar to (29):

$$R_{ic} = \alpha + \beta_c H_{ic} + \varepsilon_{ic} \tag{2.6}$$

where H_{ic} denotes a student (i)'s (monthly) percentile rank in the household income distribution in class (c). R_{ic} is a student's percentile rank as I defined in Subsection 2.2.2. β_c shows the estimated slope of the rank–rank relationship. Panel A of Appendix Figure E2 plots a student's class rank by household income rank bins. In this graph, the OLS estimate for the slope of the red line (the rank–rank relationship) is 0.230, which suggests that a 10 percentile point increase in household income rank is associated with a 2.3 percentile point increase in a student's class rank. This finding suggests that a student's class rank is positively correlated with parental income rank in class.[22] However, note that this result is descriptive, not causal.

Second, I focus on teacher gender effects according to student gender. Because the South Korean educational system provides the random assignment of subject teachers to students within classes, using this randomization allows me to propose the causal effect of teacher–student gender matches on class rank. To estimate contemporaneous effects of teacher–student gender matches on class (ordinal) rank in 7^{th} grade, I use a difference-in-difference style approach with the following specification:

$$R_{ijbs} = \alpha + \beta_1 FemStu_i + \beta_2 FemTea_j + \beta_3 FemStu_i \times FemTea_j + X_i^{'}\delta_1 + T_j^{'}\delta_2 + \gamma_{bs} + \varepsilon_{ijbs} \tag{2.7}$$

where y_{ijbgs} is a 7^{th} grade test score of student i taught by teacher j for subject b at school s. $FemStu_i$ and $FemTea_j$ are indicator variables for whether student i and subject teacher j is female, respectively. $X_i^{'}$ is a vector of student characteristics, including indicators for a single-parent household, the number of siblings, and indicators for having at least one parent with a BA degree

[21] In SELS2010, a parent's questionnaire includes monthly household income. Unfortunately, due to limited information, I cannot know whether this is the pre- or post-tax income. Moreover, because this income was self-reported by parents, my OLS estimates for the rank–rank correlation could be attenuated by the measurement error problem.

[22] Panels B and C of Appendix Figure E2 suggest that this correlation might be driven by the positive correlation between household income and parental investment ranks or household income and private tutoring ranks in class.

or higher and for having a free lunch. $T_j^{'}$ is a vector of subject teacher characteristics, including indicators for teachers' having less than five years of teaching experience, for teachers' graduation from teachers college, for teachers having a Master's degree or higher, and for administrative teachers. I add school by subject fixed effects (γ_{bs}) to compare students in a subject within a school.[23] The robust standard errors are clustered at the school level to account for arbitrary correlations among students within the same schools.

The results are reported in Appendix Table F6. In Column (1), the point estimate for female student indicates that female students are more likely to have a higher class rank than male students.[24] The point estimate for a teacher–student gender match (female student $\times$ female teacher) suggests that a female subject teacher increases female students' class rank by 0.046. This estimate is statistically significant at the 1% level and robust to the inclusion of student and teacher controls (Columns (2) and (3)). However, one may be concerned that this finding could be driven by other teacher characteristics correlated with teacher gender (46). To address this concern, I include the interactions between student or teacher gender indicators with teacher controls ($T_j^{'}$) in Equation (2.7). Intuitively, this task helps identify whether certain teacher characteristics have a different impact on students across gender. Columns (4) and (5) show that my baseline estimates are robust to the inclusion of these interactions; thus, observed teacher–student gender matches effects on class rank are driven primarily by a teacher's gender rather than by other teacher characteristics. This finding also complements to the existing evidence for a positive effect of having a female teacher on a female student's academic achievement ((47); (32)), noncognitive skills (46), and self-perceived rank (26).

2.5.4.2 The Mechanism of Increased Confidence Effects on Test Scores

As noted earlier, class rank has a positive and persistent impact on self-confidence. Increased confidence may improve students' test scores as they learn about their subject-specific abilities

[23]Including school by subject fixed effects helps to exploit within-cluster (school–subject cell) random variation in teacher–student gender interactions.

[24]Despite this finding, I note that this effect might be driven by a handful of female teachers, rather than a systematic impact of them. The detail for this result is presented in Appendix Section B.1.1.

through local (class) rank in a subject in addition to their absolute achievement (the learning–about–ability effect).[25] For example, students who have a high rank in math class in middle school can believe that they have a high ability in math; thus, they would devote more effort to math in high school without considering the opportunity cost for that task and would then be more likely to misallocate study effort across subjects. As a result, they obtain lower average test scores compared to those who are less likely to do so. To indirectly test for this possibility, I perform the empirical analysis by including the degree of misinformation that is defined as the absolute difference between local rank and global (in-sample) rank in the specification.[26] In the specification, I use average test scores and average local rankings across three subjects (math, English, and Korean) as the dependent variable and the independent variable of interest, and include the test score distribution of each subject and the classroom fixed effects.

In Appendix Table E8, the results show that a higher difference (the degree of misinformation) between local rank and global rank in 7^{th} grade reduces average test scores in all later years along with a statistically significant effect in 11^{th} grade. Hence, the learning–about–ability effect might exist in 11^{th} grade. This evidence may be different from (2)'s finding that no effect of misinformation about rank would statistically significantly impact the change in the average test score in the UK.[27] In particular, the learning–about–ability effect appears to exist for male students in 8^{th} grade and persists until 12^{th} grade (Panel C of Appendix Table F8).

[25](2) propose two distinct mechanisms behind increased confidence effects due to a higher rank on test scores: the learning–about–ability effect and the non-cognitive skill effect. Regarding these mechanisms, I describe the details in Appendix Section B.

[26]To obtain the coefficient for the effect of misinformation, I replicate the analysis of (2). Using the learning hypothesis, they document that students with large differences between local and global (national) ranks would have more distorted information about their true abilities, assuming national test scores are a good measure of ability. These students would be more likely to misallocate study effort across subjects. Hence, these students would achieve lower average grades in all subjects compared to students whose local ranks that nearly equal to global ranks. They also address that the estimate on the degree of misinformation should be negative if increased confidence could improve student test scores via the learning-about-ability effect.

[27]Table 8 of their paper shows that Age-11 misinformation increases Age-14 test scores without controlling for student characteristics, but decreases these scores with controlling for student characteristics. However, corresponding estimates for the effect of misinformation are statistically insignificant in this table.

2.5.4.3 Parental Investment and Local Distortions

My analysis of the parent–side mechanism shows that a higher class rank of children can increase parental investment in the future. This result suggests that if parents know that their child is highly ranked in class, then they may tend to encourage the child to study more or help the child to do well. This evidence is rare in the literature: most prior studies find no significant or negative effects of rank in school on parental investment ((23); (2); (26)).

To better understand the effect of class rank in 7^{th} grade on parental investment, I focus on local distortions in parental beliefs about child skill: parents of children attending schools with low (high) average will be inclined to believe that their child is higher (lower) in the overall skill distribution than they actually are (48). In addition, parental beliefs about child skill are associated with parental investment; in particular, the parental belief uncertainty (e.g. belief inaccuracies or parental overconfidence) could lead to parental investment deviated from its optimal levels (49). My hypothesis is that if parents learn child skill through class rank that is negatively associated with school average skill, then information on class rank may contribute to generate local distortions in parental beliefs about child skill.[28]

Because the data I use do not include direct information on parental beliefs about child skill, I cannot directly test whether local distortions in parental beliefs explain the effects of class rank in 7^{th} grade on parental investment.[29] Instead, I reinterpret rank effects on parental investment in terms of local distortions in parental beliefs. According to a simple framework for rank effects on parental investment, these effects can be discomposed into three parts: the relationship between parental investment and parental beliefs about child skill, the relationship between school quality and parental beliefs about child skill, and the reciprocal of the relationship between school quality and class rank.[30] Note that my identification of rank effects compares students in different classes

[28] For instance, if students have high (low) test scores but low (high) class rank, then their parents could think that their school peers have high (low) average test scores. Using this intuition, I provide a simple model of parental learning in Appendix C.2.

[29] To provide evidence of local distortions in parental beliefs about child skill, (48) use the parent's response on the question "Does your child learn, think, and solve problems better, as well, slightly less well, or much less well than other children his/her age?" Unfortunately, my data do not contain this kind of response.

[30] I provide the detail of how to develop this framework in Appendix Subsection C.2.1.

who have the same test score. These students' class ranks are generally negatively correlated with school quality; thus, the sign of the reciprocal of the correlation between school quality and class rank is negative. In addition, the effects of class rank on parental investment are positive until 11^{th} grade in Table 10. For simplicity, I assume that parents who believe that their child has a higher skill choose a more increased investment. To obtain the positive effect of class rank on parental investment, there should be local distortions in parental beliefs about child skill.[31] I am still agnostic on the role of class rank in driving these results. However, the framework suggests that if information on class rank contributes to generate local distortion in parental beliefs about child skill, then my rank effects on parental investment could be driven by local distortions of parents whose investment is positively associated with their beliefs about child skill.

2.5.4.4 Higher–preferred Schools and Future Academic Gains

While a higher class rank increases the probability of attending higher–preferred high schools, it is still ambiguous if doing so will actually lead to students' academic success. If these schools are "better", then they will have better peers in the classroom, which creates a negative effect on academic performance by lowering a student's rank ((50); (51); (21)). However, few papers document that a student can benefit from high-quality peers in the classroom ((52); (53)). Despite evidence of a positive rank effect on higher-preferred school choices, I still remain agnostic on the relationship between school choices associated with rank and future academic gains or educational attainments.

To indirectly estimate the net gains of attending a school with better performance, I perform the thought experiment by grouping schools into "high-quality", "middle-quality", and "low-quality" schools in terms of mean test scores in 7^{th} grade.[32] Here, I assume that a higher-preferred school has better peers in class. Appendix Table F9 shows a negative correlation between school quality and class rank of students: the values in columns of the table are the mean (percentile) rank in each test score decile in each school type. Specifically, for each decile, the mean rank of students

[31] This means that $\frac{\partial P^B(c_{ij},s_j)}{\partial s_j}$ should be negative in Appendix Equation C.10.

[32] My experiment is inspired by (21); however, I consider only average test scores of each school as school quality.

decreases when moving from a "low-quality" school to a "high-quality" school. Students in the higher distribution (decile) of test scores have higher ranks in their classes. The mean value-added scores are higher when moving from a "low-quality" school to a "high-quality" school.[33] Appendix Table F10 presents the net gains of moving from a "low-quality" school to a "high-quality" school net of rank effects by conducting simulations. I classify a student in decile 1, decile 5, and decile 10 into a low-achieving student, a middle-achieving student, and a high-achieving student, respectively. Panel D of this table indicates that a middle-achieving student obtains net losses or small net gains by attending a "high-quality" school over time, while a low- or high-achieving student has net gains.[34] From these simulation results, I cannot rule out the possibility that some students who attend the better school by school choices would have the reduced rank and, hence, obtain academic net losses in the future. Therefore, I do not overstate a positive impact on the probability of attending the high school in the first–round application (Column (1) of Table 2.12) when interpreting it as the potential driver of a positive rank effect on future academic success.

2.6 Conclusion

My study complements the growing number of studies on the existence of rank effects by constructing classroom–subject cells and then computing local rank within such cells for the main analysis. I show that class rank is quasi-randomly assigned to students conditional on cubic in test score and classroom by subject fixed effects. The random assignment of students to classrooms allows me to eliminate concerns about the self-selection problem that confounds "true rank" effects.

In this paper, I find that a student's class rank in 7^{th} grade continuously positively impacts

[33] For each grade, I regress test scores from 8^{th} to 12^{th} grade on a cubic of test scores in 7^{th} grade and middle school fixed effects. A school value-added score in each grade is recovered from a school fixed effect. Then, I compute the mean school value-added score by averaging school value-added scores for those grades.

[34] To calculate each value in Panel D, I first compute value-added gains (0.254=0.145-(-0.109) from the bottom of Appendix Table F9) from a "low-quality" school to a "high-quality" school (Panel A). Then, I obtain correlations between class rank and test scores by regressing test scores from 8^{th} to 12^{th} grade on class rank in 7^{th} grade (Panel B). Attending the better school would reduce students' expected class rank by 0.049 (low-achieving student), 0.275 (middle-achieving student), and 0.087 (high-achieving student) (Panel C). Lastly, I obtain net gains (losses) by adding value-added gains to the rank effect (correlations between class rank and test score (Panel B) $\times$ expected class rank changes) over time: specifically, 0.159=0.254+((1.943)$\times$(-0.049)) for a low-achieving student.

academic performance from 8^{th} to 12^{th} grade. A higher math rank in 7^{th} grade increases the 11^{th} grade reported likelihood that students choose the math–science track, take higher-level math courses, and show interest in obtaining a STEM degree in college. I also find that a higher class rank in 7^{th} grade raises the likelihood of attending a university. In addition, a higher math rank in 7^{th} grade increases the likelihood of applying and actually majoring in a STEM field at a university. The changes in students' self-confidence and effort provisions, parental investment, and students' attendance in a preferred high school might be the potential drivers to explain my results.

My findings provide strong evidence that a higher class rank plays a significant role in obtaining future academic gain or choosing the later career path of students. My evidence also implies that students with a high achievement but a low rank in class may obtain a serially persistent loss in test scores compared to students with the same achievement but a high rank in class. Hence, parents and children need to consider the potential trade–off between attending more selective schools and a lower rank. In addition, using the salience of rank, policy makers could develop an effective way to give more support to lower–ranked students in class in order to offset the persistent and negative effects of their rank.[35] Doing so may help close an observed gap in educational outcomes over the education cycle.

[35] Similar to the argument of (23), providing low-ranked students in high-performing schools with information on their absolute position in the national achievement distribution could be an effective way to improve their academic performance. Given my finding in Subsection 2.5.4.2, it is expected that the persistent effect of this policy using the salience of global rank would exist for male students who has a much lower class rank than global rank.

3. WATCH OUT FOR MISBEHAVING BOYS! CONTAGION EFFECTS IN RANDOMLY ASSIGNED CLASSROOMS

3.1 Introduction

In education, there is a well-documented gender gap in noncognitive skills, especially with boys performing worse than girls on behavioral factors (3). Behavioral difficulties among boys in school, including the inability to pay attention in class, to work with other pupils, to follow homework or class materials and to seek help from other classmates can help explain boys' disadvantage for academic and labor market success (54). Boys are more likely to engage in aggressive behavior than girls in school (55). This gender disparity is revealed in school suspension statistics: during 2013$\sim$2014, more than twice as many male students (7.3%) than female students (3.2%) received one or more school suspensions (U.S. National Center for Education Statistics, 2019).

Nonetheless, there exists little causal evidence of the impacts of misbehaving peers on misbehaviors of others. A small body of past works document that aforementioned traits of boys have spillover effects on other students' behavioral outcomes more likely than those of girls. Boys with female sounding names are more likely to engage in misbehaviors, thereby increasing classmates' disciplinary problems (11). Male students who are exposed to domestic violence also lead classmates to violate school rules (12). These studies suggest that those students' behavioral problems eventually generate negative externalities in the classrooms, thereby reducing other classmates' academic achievements.

Disentangling "true" peer effects from other confounding factors can be a significant challenge in peer effects studies. Addressing non-random sorting (or the "selection problem") is important for identification. If individuals with similar personal and family backgrounds self-select into a peer group, then students' outcomes across classrooms or schools might not be truly attributable to the effect of their peers. Two major strategies have been put forth to tackle the problem of self-selection. The first one exploits the random assignment of individuals into peer group ((56);

(57); (58); (59)). This strategy is mostly observed in the literature on US higher education because the truly random assignment of children to classrooms is rare in US primary and secondary school setting with exception of Project STAR. The second one, introduced by (8), creates a novel and compelling method to estimate the causal effects of school peers that relies on the comparison of cohort-to-cohort variation in peer measure within school by grade and by year. This strategy has since been used in several peer effect papers ((12); (1)), though it usually requires large administrative panel data sets and controlling for a series of fixed effects.

The purpose of this paper is to contribute to the literature by presenting new evidence that misbehaving boys have spillover effects on other students' behavioral outcomes by adopting the first strategy: students in middle schools are randomly assigned to a physical homeroom classroom in which they interact with the same classmates throughout the entire school day. This setting offers ideal circumstances for avoiding concerns of endogenous sorting and differences in factors that might be correlated with peer composition. Moreover, I use a dataset that includes a record of the students' misbehaviors in the class. These misbehaviors are reported by the students and represent how frequently they misbehave in the class. The method of (4) allows computation of a single misbehavior composite index (MCI) that captures the common variation in those misbehaviors and to estimate the contagion effects in a straightforward way.

In addition, I overcome the reflection problem – in which the direction of peer effects is opaque (60) – by linking the reasons for classroom peers' misbehavior to family trouble like a single-parent household. This is supported by the literature ((61); (62); (63)) showing that children raised in a single-parent household are more likely to experience a number of emotional and behavioral problems such as depression, violence, and externalizing behavior. Because single-parent status of randomly assigned peers cannot directly affect other children's misbehavior, I instrument peers' average MCI with peers' having single-parent households to obtain exogenous variation in the peer measure.

I find that misbehaving male students statistically significantly increase the intensity of others' misbehaviors in the classroom, whereas misbehaving female students do not. Specifically, the increase of one single misbehaving male student among 30 students in the classroom leads to an increase of other students' MCI by 3.2% of a standard deviation. These results are robust to the inclusion of student, peer, and teacher characteristics, providing plausible evidence on random assignment. Moreover, adding one single misbehaving male student among 30 students into the classroom increases other male students' MCI by 5.3% of a standard deviation. This result is more pronounced for male pupils who live in low-income households.

By showing the non-linear effects of misbehaving peers by student gender in randomly assigned classrooms, the results reinforce the existing finding that boys are the main drivers for class disruption ((8); (9)). As a policy implication, these results provide educators with compelling reasons to address student misbehavior in the classroom; at least in this context, doing so might be effective if it mainly targets misbehaving boys.

3.2 Background and Data

3.2.1 Institutional Setting

Since 1969, when the "leveling policy" was first introduced in South Korea, student allocation to middle schools has been fairly simple and straightforward (31). Elementary school graduates move to local middle schools within the residence-based school attendance zones assigned by a lottery system.[36] Then, they are assigned to a physical homeroom classroom at the beginning of each academic year (generally in the first week of March). The homeroom is the physical place where children mostly remain throughout a school day, and where subject teachers visit to teach them. Regardless of whether a school is public or private, almost all middle schools implement a random assignment of homerooms.

The most common method of assignment within the school is to order students by previous academic performance and assign them one by one across homerooms. For instance, the top-ranked student is assigned to Classroom 1, the second-ranked student to Classroom 2, the third-

[36]The formulas used for school assignment are not known to the public; the assignment process involves random draws performed using software (64). Students are not permitted to change middle schools within the same zone, except in the case of residential relocations. Although students do transfer across different school attendance zones, they randomly assigned to a new school in the new zone

ranked student to Classroom 3, and so on.[37] Reporting the distribution of the previous academic performance to both students and their parents is strictly banned before classroom assignment is complete. This system prevents children from being compulsorily sorted into a homeroom peer group that is extremely skewed to high (or low) academic ability, naturally guaranteeing ability mixing within classrooms. This method yields as-good-as random classroom assignment of children to their own peer group with respect to at least students' non-academic characteristics. In addition to classmate assignment, homeroom teachers are also assigned, either by lottery or a committee, to specific homeroom classrooms (59). Homeroom teachers generally teach a specific subject themselves and are responsible for school discipline.

3.2.2 Sample Construction and Statistics

I use the 2012 Gyeonggi Education Panel Study (GEPS) data from the Gyeonggi Institute of Education, a research institute responsible for advising on educational policy in the Gyeunggi province (surrounding Seoul, South Korea).[38] GEPS surveys 7^{th} grade students in two classrooms randomly chosen within each school, and also collects information on these students' parents and homeroom teachers.

The initial dataset includes 4,051 7^{th} grade students. Dropping observations without information about parental marriage status (including legal divorce, separation, and death) or classroom information leaves 3,984 observations. Each student was asked to self-report their misbehavior related to classroom disruption or risky behavior.[39] Table 1 provides summary statistics for the main

[37](47) performed a survey to examine whether this rule is really applied in 197 South Korean middle schools by coordinating with local offices of education and found that almost all schools were applying this method of classroom assignment.

[38]Sixty-three schools were randomly chosen from the population of 624 middle schools in the GEPS dataset using a two-stage cluster sampling design.

[39]There could be concerns that students might under-report their bad behaviors compared with their teachers' reports. For example, (65) use multiple teachers' reports of children's misbehavior for their research. Unfortunately, GEPS does not include information on pupils' misbehaviors as reported by teachers, so I cannot thoroughly cross-check whether and to what extent students' self-reporting misbehaviors are consistent with teachers' reports of bad behaviors. However, I find that 81.3% of 7^{th} grade students who talk with others in class were not reported by homeroom teachers in a different dataset gathered by the Korea Education Longitudinal Study (KELS 2005). Hence, it appears that relying on self-reporting measures for misbehavior is useful for my research because I can obtain information on children's misbehavior that are not observed by teachers. Moreover, I mostly use this measure as the outcome variable in this paper. Thus, I believe that the measurement errors regarding misbehavior do not lead to inconsistency of the estimator, though these errors may inflate the standard errors of estimated coefficients of interest

sample of students used in the analysis. Male students constitute 51% of individuals in my sample. Table 3.1 also indicates that 5% of individuals are male students in a single-parent household, which is nearly the same as the percent of male peers in a single-parent household.

Table 3.1: Summary Statistics

	Mean	Std. Dev	Min	Max	N
A. Student Outcomes					
(Raw) Misbehavior Composite Index (MCI)	0.359	0.259	0.00	1.00	3,539
B. Student Characteristics					
Male	0.51	0.50	0.00	1.00	3,593
Dad w BA or Higher Degree	0.60	0.49	0.00	1.00	3,575
Mom w BA or Higher Degree	0.49	0.50	0.00	1.00	3,576
Household Income (million KRW)	4.92	4.30	0.00	99.99	3,537
C. Peer Characteristics					
Male Peers w Single Parent	0.05	0.05	0.00	0.32	3,593
Female Peers w Single Parent	0.05	0.05	0.00	0.32	3,593
Male Peers w Single Parent (Net of School Fixed Effects)	0.00	0.03	-0.12	0.14	3,593
Female Peers w Single Parent (Net of School Fixed Effects)	0.00	0.03	-0.09	0.18	3,593
D. Classroom Characteristics					
Female Homeroom Teacher	0.81	0.39	0.00	1.00	3,459
Homeroom Teacher over 40	0.29	0.45	0.00	1.00	3,459
Homeroom Teacher Experience below 5 Years	0.21	0.40	0.00	1.00	3,459
Homeroom Teacher's College	0.71	0.45	0.00	1.00	3,422
Homeroom Teacher's Post Graduate	0.38	0.49	0.00	1.00	3,459
Class Size	33.07	5.20	12	40	3,593

Notes: I restrict the sample to students who do not have experienced parental divorce, separation or death.

3.2.3 Misbehaving Children and Composite Index

The GEPS data set includes questions that ask children how frequently they misbehave in the class. Students were asked to answer the following 4 questions:

(66).

How frequently did you engage in the following misbehaviors during this year? Choose from "Never", "Once or Twice per year", "Once or Twice per semester", "Once or Twice per month", "Once or Twice per week", or "Nearly Every Day".

1. Chat with others in class 2. Defy teachers in class 3. Sleep in class 4. Goof off in class

Each component of these misbehaviors is measured with scale scores ranging from one ("Never") to six ("Nearly Every Day"), revealing the intensity of children's misbehavior. The overall distribution of children's propensity to misbehave across each component is summarized in Table 3.2. However, the distribution of the intensity of children's misbehavior is not balanced across each bad behavior. For example, the fraction of students who never chat with others in class is 6.08%, while 61.93% of students report they never defy teachers in class. I simplify how intensity is measured, creating a binary variable by collapsing the six answers in Table 3.2 into two categories: low and high tendency. Low tendency comprises "Never", "Once or Twice per year", and "Once or Twice per semester"; high tendency comprises "Once or Twice per month", "Once or Twice per week" and "Nearly Every Day". Note that this is not intended to proxy for a student's actual misbehaviors, but rather, as a (noisy) measure of how they perceive themselves to misbehave in the class. Table 3.3 provides the resulting narrower distribution of misbehavior tendency.

To obtain a more straightforward misbehavior measure, I construct a single composite index of these misbehaviors by using the aggregation method of (4). I name this the Misbehavior Composite Index (MCI). This MCI is then normalized to z-scores with means of zero and standard deviations of one.

Table 3.2: Distribution Over Children's (Raw) Propensity to Misbehave in a Class

	Never	Once or Twice per year	Once or Twice per semester	Once or Twice per month	Once or Twice per week	Nearly Every Day
Chat with others in class	6.12	7.30	6.96	14.48	36.57	28.57
Defy teachers in class	62.52	14.92	7.18	7.60	6.43	1.34
Sleep in class	32.87	13.78	9.46	15.09	21.26	7.53
Goof off in class	56.33	13.95	7.59	10.59	9.08	2.46

Notes: I restrict the sample to students who do not have experienced parental divorce, separation or death. Each value represents a percentage (%) of self-reported misbehavior frequency. In each row, the tendency of misbehavior increases from left to right.

Table 3.3: Distribution Over Children's (Transformed) Propensity to Misbehave in a Class

	Low Tendency	High Tendency
A. All Students		
Chat with others in class	20.38	79.62
Defy teachers in class	84.63	15.37
Sleep in class	56.11	43.89
Goof off in class	77.87	22.13
B. Male Students		
Chat with others in class	19.09	80.91
Defy teachers in class	79.37	20.63
Sleep in class	57.51	42.49
Goof off in class	79.25	20.75
C. Female Students		
Chat with others in class	21.73	78.27
Defy teachers in class	90.12	9.88
Sleep in class	54.65	45.35
Goof off in class	76.43	23.57

Notes: I restrict the sample to students who do not have experienced parental divorce, separation or death. This table is created by collapsing six answers of Table 2 into two categories: low tendency includes "Never", "Once or Twice per year" and "Once or Twice per semester"; high tendency contains "Once or Twice per month", "Once or Twice per week" and "Nearly Every Day". Each value represents a percentage (%) of self-reported misbehaving frequency in terms of those two categories. In each row, the tendency of misbehavior increases from left to right.

3.3 Identification Strategy

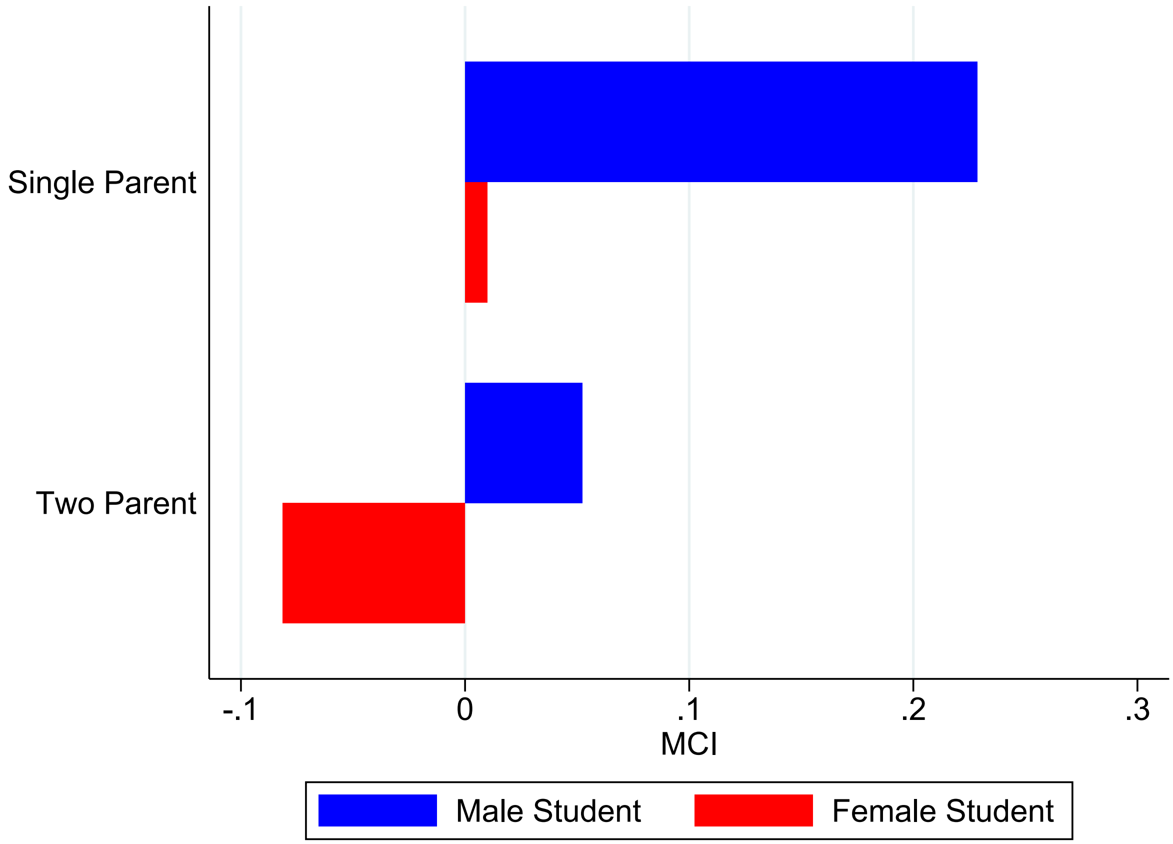
3.3.1 Empirical Approach

The simple way to investigate misbehaving peer effects is to estimate the effects of peers' (average) MCI on a student's own MCI. However, this would be biased by the reflection problem (60). The student and his or her peers affect each other's misbehavior in the classroom. To overcome the reflection problem, I develop an instrument for measuring misbehaving peers. Although employing lagged characteristics of my sample students (7^{th} grade children) is a plausible way to avoid the reflection problem, the dataset does not allow me to do so because it does not contain information on these characteristics prior to the sample year.[40]

Instead, I exploit a family attribute as the exogenous source of variation in peer quality in a manner similar to other papers ((69); (12); (1)). In particular, the dataset contains information about whether children have experienced parental divorce, separation or death. Some studies support that the effects of single parenting could lead children to suffer from behavioral, mental, or emotional problems. For instance, (71) indicate that children from single-parent families are more likely to engage in delinquent behavior and juvenile crime, and to exhibit depression and other disorders. (72) suggest that adolescents' self-esteem in single-parent families tends to be lower than that in the two-parent families. Panel C of Table 3.1 shows that the residual variation is about 0.03 for both male and female peers with a single parent after removing school fixed effects, accounting for about three-fifths (0.03/0.05) of the overall variation (0.05). This demonstrates sufficient residual variation in these two peer measures. Figure 3.1 displays conditional means for students with single parents and both parents graphically. Specifically, the blue and red bars represent mean residuals for male and female students after regressing the MCI on school fixed effects. This figure shows that students with single parents have a higher MCI than those with both parents. In addition, male students with single parents have a much higher MCI than female students with single parents.

[40]For example, (67), (68), (69), and (70) have used lagged outcomes or characteristics of peers as the variables of interest to resolve the reflection problem in the peer effects literature.

Figure 3.1: Single Parenting and MCI



Notes: This figure depicts the relationship between single parenting in the household and a student's misbehaving composite index (MCI). The values of MCI on x-axis are the conditional means of residuals from a regression that includes school-fixed effects.

I consider single parenting as a relevant proxy for misbehavior. Importantly, I use two instrumental variables, such as the proportion of male peers with a single parent and the proportion of female peers with a single parent to capture the effects on outcomes by peer gender. I thus estimate a reduced-form (RF) specification in which peer quality is directly substituted for such instrumental variables:

$$MCI_{ics} = \alpha + \frac{1}{n_c - 1}\sum_{i\neq j}(\beta_0 MaleSingleParent_{jcs} + \beta_1 FemaleSingleParent_{jcs} + \gamma X_{jcs}) \\ + \delta X_{ics} + \eta C_{cs} + \theta_s + \varepsilon_{ics} \quad (3.1)$$

where MCI_{ics} is the MCI for 7^{th} grade student i, randomly assigned to classroom c in school s. $MaleSingleParent_{jcs}$ is the indicator for 7^{th} grade male student j with a single parent, randomly assigned to classroom c in school s. $FemaleSingleParent_{jcs}$ is the indicator for 7^{th} grade female student j with a single parent, randomly assigned to classroom c in school s. X_{ics} is a vector of student observed characteristics. Student characteristics include student gender, parents' education, and household income. C_{cs} is a vector of classroom observed characteristics.[41] Classroom characteristics include homeroom teacher gender, whether the homeroom teacher is older than 40 years, whether the homeroom teacher has less than 5 years of teaching experience, whether the homeroom teacher graduated from a teachers college, whether the homeroom teacher has a post-graduate degree, and class size. S_s is a set of school fixed effects.[42] ε_{ics} is the idiosyncratic error term. I cluster standard errors at the classroom level to account for any arbitrary correlations across students within the same classroom.

As (73) noted, there is a negative mechanical correlation between own and peer characteristics when adding peer averages to the right hand side peer variable in the specification. To address this problem, I follow their solution; thus, I exclude students with single parents from the data.

3.3.2 Tests for Randomization

Both institutional setting and evidence in the literature ((31), (59)) support that the random assignment of peers takes place in South Korea middle schools. I provide additional evidence here.

I test whether the average characteristics of classroom peers are associated with the student's

[41] As (31) noted, omitting homeroom teacher characteristics in the specification for estimating the effects of classroom peers can cause the limitation to have robust estimates. However, my data contains information about homeroom teachers; therefore, I am able to overcome this limitation by controlling for classroom-level characteristics including homeroom teacher' background.

[42] Controlling for school fixed effects accounts for the random assignment of students and other resources to classrooms within schools, as well as any broader unobserved school factors.

own characteristics. Following (35), I also address the potential for a negative bias problem when regressing an individual's own characteristics on peers' characteristics. As (35) refer to (56)'s critique of typical randomization tests, sampling without replacement causes each individual to be removed from the "urn" from which the peer group is chosen. Peers for high-ability individuals are therefore chosen from a group with a slightly lower mean ability than the peers for low-ability individuals. This creates a mechanical negative correlation between one's own and one's peers' characteristics, and controlling for the leave-out mean for the population from which peers are drawn is a good way to rule out this problem. Hence, I examine the independence of characteristics:

$$x_{ics} = \alpha + \beta \frac{1}{n_c - 1} \sum_{i \neq j} (x_{jcs}) + \gamma \frac{1}{n_s - 1} \sum_{i \neq j} (x_{js}) + \lambda_s + \varepsilon_{ics} \tag{3.2}$$

where x_{ics} is student i's characteristic in randomly assigned classroom c within school s, and n_c and n_s are the number of students in classroom c and school s, respectively. The term on the right-hand-side that includes x_{js} represents the mean of the characteristic within the school, which is necessary to alleviate a negative mechanical bias. School fixed effects (λ_s) account for random assignment within a school, and ε_{ics} is the idiosyncratic error term. Table 3.4 indicates that only one of four student characteristics is statistically significantly correlated with the corresponding peer's characteristic, yielding evidence of as-good-as random assignment to the peer group.[43]

[43]By using the data including students with single parents, I additionally test for whether own single parenting is affected by peer single parenting. The point estimate is -0.047, but is not statistically significant at conventional levels. This provides evidence that within-school variation in the proportion of peers from a single-parent household is not systematically associated with one's own likelihood of having single parents, when I control for the school-level leave-out mean to correct for mechanical bias. This also indicates that single parenting in the household is an exogenous measure due to a non-existent feedback loop.

Table 3.4: Regression of Own Characteristics on Peer Characteristics

	(1) Male Student	(2) Dad BA+	(3) Mom BA+	(4) Household Income
Avg. Classroom Peers Characteristics	-0.142	0.065**	0.040	-0.032
	(0.123)	(0.029)	(0.037)	(0.024)
Observations	3,593	3,575	3,576	3,537
R^2	0.952	0.959	0.959	0.978

Notes: I restrict the sample to students who do not have experienced parental divorce, separation or death. Each column represents results from a separate regression, including school fixed effects and the leave-out mean of school peers' characteristic. I also control for the school peers' leave-out mean (school-level mean for the relevant characteristics excluding the individual's own values) to overcome the negative mechanical bias of the randomization tests. In each column, dependent variables are indicators for male students, for having a dad with a BA degree or higher, for having a mom with a BA or higher degree, and household income (in millions of KRW). Standard errors, shown in parentheses, are clustered at the school level.
*p<0.10, ** p<0.05, *** p<0.01.

In addition, I examine whether those instrumental variables are correlated with some of covariates. In Table 5, the result indicates that there is very little relationship between peers with a single parent and observed covariates. Of the eighteen estimates shown in Columns (1)–(9), four are significant at the 5% or 10% level, roughly consistent with random chance.

Table 3.5: Effects on Student and Classroom Characteristics

	(1) Dad BA+	(2) Mom BA+	(3) Household Income	(4) Female Teacher	(5) Teacher over 40	(6) Low Experience	(7) Teacher's College	(8) Post Gradate	(9) Class Size
Male Peers	0.206	0.277*	-0.556	1.670*	0.851	-0.847	0.885	0.587	6.840**
w Single Parent	(0.152)	(0.141)	(1.514)	(0.970)	(0.794)	(1.048)	(0.965)	(0.999)	(2.779)
Female Peers	-0.130	-0.107	-0.327	0.123	0.807	-0.770	-0.390	2.144**	0.248
w Single Parent	(0.144)	(0.132)	(1.154)	(0.876)	(0.855)	(0.843)	(1.018)	(0.941)	(3.368)
Observations	3,575	3,576	3,537	3,459	3,459	3,459	3,422	3,459	3,593
R^2	0.163	0.150	0.061	0.530	0.709	0.606	0.483	0.560	0.960

Notes: I restrict the sample to students who do not have experienced parental divorce, separation or death. Each column represents results from a separate regression, including indicators for a male student with a single parent and for a male student, and school fixed effects. The independent variables are the proportion of classroom peers with a single parent. In each column, dependent variables are indicators for having a dad with a BA or higher degree, for having a mom with a BA or higher degree, and household income (in millions of KRW), and indicators for homeroom teacher gender, whether the homeroom teacher is older than 40 years, whether the homeroom teacher has less than 5 years of teaching experience, whether the homeroom teacher graduated from a teachers college, whether the homeroom teacher has a postgraduate degree, and class size. Standard errors, shown in parentheses, are clustered at the classroom level.
*p<0.10, ** p<0.05, *** p<0.01.

3.4 Empirical Results

3.4.1 Main Effects

I present the results of misbehaving peers on students' MCIs in Table 3.6. Each column includes school fixed effects along with student- and classroom-level observed characteristics. Robust standard errors are clustered at the classroom level, which allows students to correlate with each other in their classroom. The outcome variable for each column is the MCI.

Table 3.6: Results - Peer Effects on a Student's Own Misbehavior Composite Index

	(1)	(2)	(3)
Male Peers w Single Parent	0.963**	1.079**	1.160**
	(0.478)	(0.471)	(0.463)
Female Peers w Single Parent	-0.666	-0.703	-0.407
	(0.428)	(0.439)	(0.440)
Observations	3,539	3,465	3,298
R^2	0.056	0.059	0.062
School fixed effects	Y	Y	Y
Student-level controls	N	Y	Y
Classroom-level controls	N	N	Y

Notes: I restrict the sample to students who do not have experienced parental divorce, separation or death. Each column represents results from a separate regression. In each column, the dependent variables are the composite index of a students' own misbehavior. Student-level controls include indicators for a dad having a BA degree or higher, for a mom having a BA degree or higher, and household income (in millions of KRW). Classroom-level controls include peer characteristics that are leave-out means of an indicator for a male student and student-level controls, indicators for homeroom teacher gender, whether the homeroom teacher is older than 40 years, whether the homeroom teacher has less than 5 years of teaching experience, whether the homeroom teacher graduated from a teachers college, whether the homeroom teacher has a Master degree or higher, and class size. Standard errors, shown in parentheses, are clustered at the classroom level.
*p<0.10, ** p<0.05, *** p<0.01.

Columns (1)–(3) present the reduced-form estimates by replacing classroom peers' MCI with both a fraction of male peers with a single parent and a fraction of female peers with a single parent. Column (1) shows that the coefficient on male peers with a single parent (the variable of interest) is 0.963. This estimate is just statistically significant at the 5% level, but it suggests that adding one misbehaving male peer, proxied by male peer with a single parent, to a classroom of 30 increases a student's MCI by 3.2% (0.963/30) of a standard deviation. In contrast, female peers with a single parent negatively impact a student's MCI, but this impact is not statistically significant. In Columns (2)-(3), including student- and classroom-level controls does not change these estimates much, providing evidence of random assignment.

In addition, I examine the non-linear peer effects by student gender. In Table 3.7, I present the results by allowing the interaction of the main parameters, as shown in Table 3.6, with student gender. The specification for the RF estimation is based on the following equation:

$$\begin{aligned} MCI_{ics} = \alpha + \frac{1}{n_c - 1} \sum_{i \neq j} (\beta_{00} MaleSingleParent_{jcs} \times OwnMale_{ics} + \\ \beta_{01} MaleSingleParent_{jcs} \times OwnFemale_{ics} + \beta_{10} FemaleSingleParent_{jcs} \times OwnMale_{ics} + \\ \beta_{11} FemaleSingleParent_{jcs} \times OwnFemale_{ics} + \gamma X_{jcs}) \\ + \delta X_{ics} + \eta C_{cs} + \theta_s + \varepsilon_{ics} \end{aligned} \tag{3.3}$$

Columns (1)–(3) display the RF estimates by estimating Equation (3.3). The results show that the impacts of male peers with a single parent are noticeably revealed for a male student: adding one male peer with a single parent, to a classroom of 30 increases a male student's MCI by 5.3% (1.591/30) of a standard deviation. These peers also have a positive effect on female students, but the estimate is not precisely estimated. These results do not vary much with the addition of controls in Columns (2) and (3), providing evidence of the robustness of the results.

Taken together, there is suggestive evidence that the impact of misbehaving male peers (boys) is the influential driver to generate behavioral spillovers and that such impact is mainly driven by the spread on a male-to-male basis.

Table 3.7: Result - Peer Effects Interacted with Student Gender on a Student's Own Misbehavior Composite Index

	(1)	(2)	(3)
Male Peers w Single Parent	1.591**	1.720***	1.830***
× Own Male	(0.621)	(0.636)	(0.636)
Female Peers w Single Parent	-0.508	-0.572	-0.036
× Own Male	(0.792)	(0.783)	(0.811)
Male Peers w Single Parent	0.404	0.528	0.576
× Own Female	(0.565)	(0.549)	(0.539)
Female Peers w Single Parent	-0.804	-0.824	-0.667
× Own Female	0.558)	(0.592)	(0.588)
Observations	3,539	3,465	3,298
R^2	0.057	0.060	0.063
School fixed effects	Y	Y	Y
Student-level controls	N	Y	Y
Classroom-level controls	N	N	Y

Notes: I restrict the sample to students who do not have experienced parental divorce, separation or death. Each column represents results from a separate regression. In each column, the dependent variables are the composite index of a students' own misbehavior. Student-level controls include indicators for a dad having a BA degree or higher, for a mom having a BA degree or higher, and household income (in millions of KRW). Classroom-level controls include peer characteristics that are leave-out means of an indicator for a male student and student-level controls, indicators for homeroom teacher gender, whether the homeroom teacher is older than 40 years, whether the homeroom teacher has less than 5 years of teaching experience, whether the homeroom teacher graduated from a teachers college, whether the homeroom teacher has a Master degree or higher, and class size. Standard errors, shown in parentheses, are clustered at the classroom level.
*p<0.10, ** p<0.05, *** p<0.01.

3.4.2 Additional Heterogeneous Effects

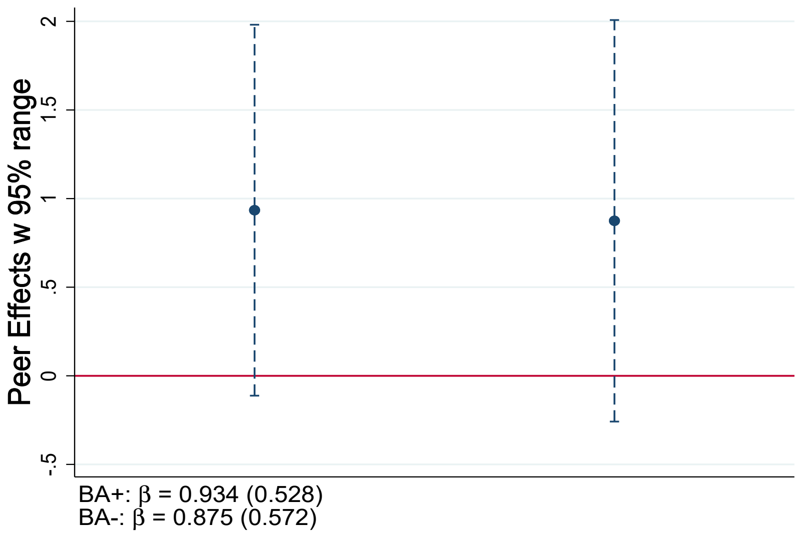
I further explore the extent of which students are most affected by the exposure to misbehaving peers as proxied by peers with a single parent. Specifically, I allow the effect to be heterogeneous by parental education, household income, and class size. In these analyses, I mostly focus on the estimated coefficients of misbehaving male peer effects and these effects interacted with the indicator for a male student. The results are presented in Figures 3.2–3.4. The 95% confidence interval of the coefficients are shown in the figures.

Figure 3.2 displays the point estimates of misbehaving male peer effects separately based on parental education. I perform the subgroup analysis by interacting peer measures with indicators for parental education. I group students into two groups, including those having parents with a higher school certificate or under and those with a Bachelor's degree or higher. Panel A shows that the estimated coefficient for those having parents with a higher school certificate or under is smaller than that for those with a Bachelor's degree or higher. However, two estimates are statistically insignificant at conventional levels. In Panel B, the between-group difference in the effect is reversed for male students: the effect in a group with less educated parents is larger than that in a group with more highly educated parents. The corresponding estimates are statistically significant at conventional levels.

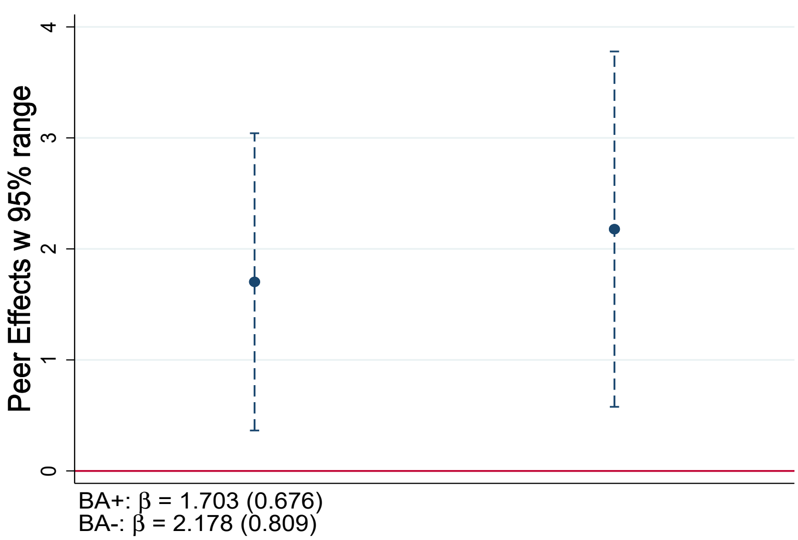
Figure 3.3 shows the coefficients estimated from the subgroup analysis by interacting peer measures with indicators for the household income level. For doing so, I divide students into two groups – low-income and high-income groups – according to whether their household is below the poverty line. Panel A shows that misbehaving male peers are more likely to affect students in low-income households than in high-income households. The point estimate suggests that one misbehaving male peer among 30 students would increase MCI for students in low-income households by 0.04 (1.147/30) standard deviations. However, this estimate is not statistically significant. In Panel B, the estimated coefficients are much larger for male students. Misbehaving male peers still have a large and positive effect on the change in MCI for male students in low-income households. The estimated coefficient for this effect is statistically significant at the 5% level.

Figure 3.4 depicts the results of the subgroup analysis by interacting peer measures with indicators for the class size. I categorize student groups as either having an above- or a below-median class size (33). In Panel A, misbehaving male peers have a larger impact on others' MCI in larger classrooms, while none of estimated coefficients are significantly different from zero. In Panel B, I find a larger estimate in larger classrooms for male students. However, this effect is statistically insignificant, whereas the effect in classrooms with a below-median size is statistically significant.

Figure 3.2: Effects of Misbehaving Male Peers by Parental Education



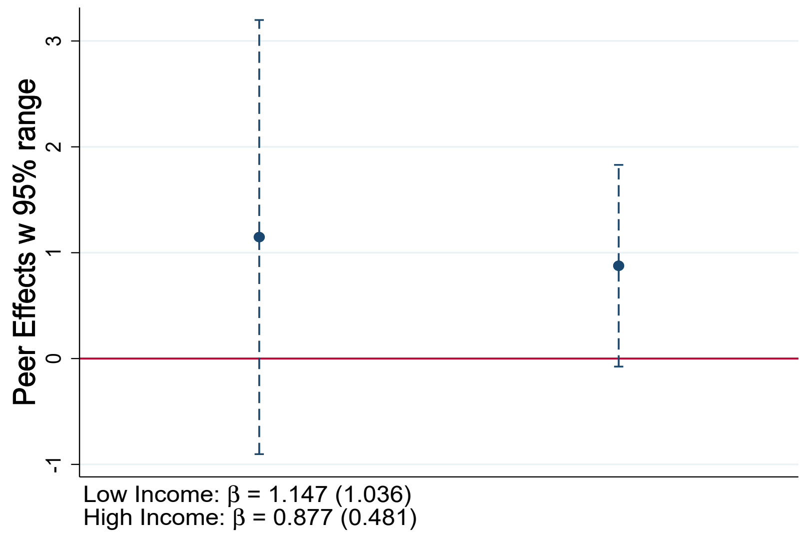
(a) Panel A: Baseline



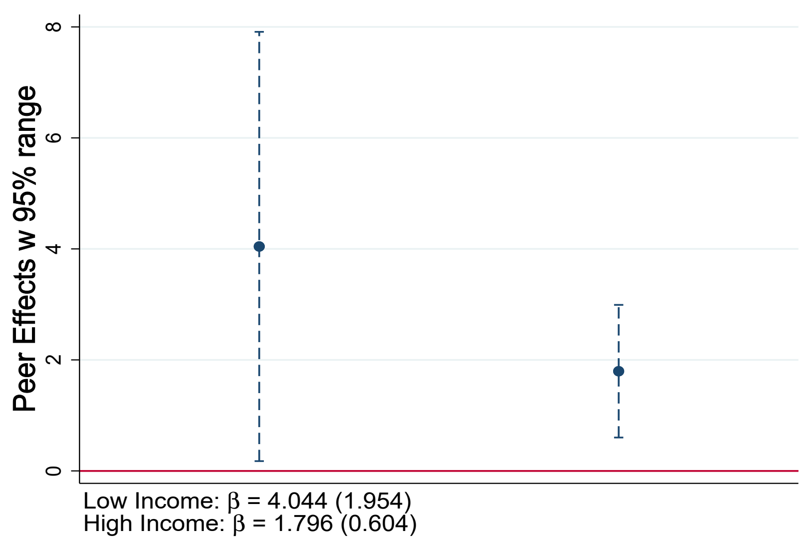
(b) Panel B: Misbehaving Male Peers to Male Students

Notes: I restrict the sample to students who do not have experienced parental divorce, separation or death. This figure presents the estimated coefficients on misbehaving male peers (male peers with a single parent) from specification (1) in Tables 3.6 and 3.7 by interacting peer measures with indicators for whether at least one parent has a BA degree or higher. Panel A shows the estimated effects of misbehaving male peers. Panel B shows the estimated effects of misbehaving male peers interacted with the indicator for a male student. In both Panels, β is the estimated coefficient while its standard error is shown in parentheses. 95% confidence intervals are shown with dashed lines.

Figure 3.3: Effects of Misbehaving Male Peers by Household Income



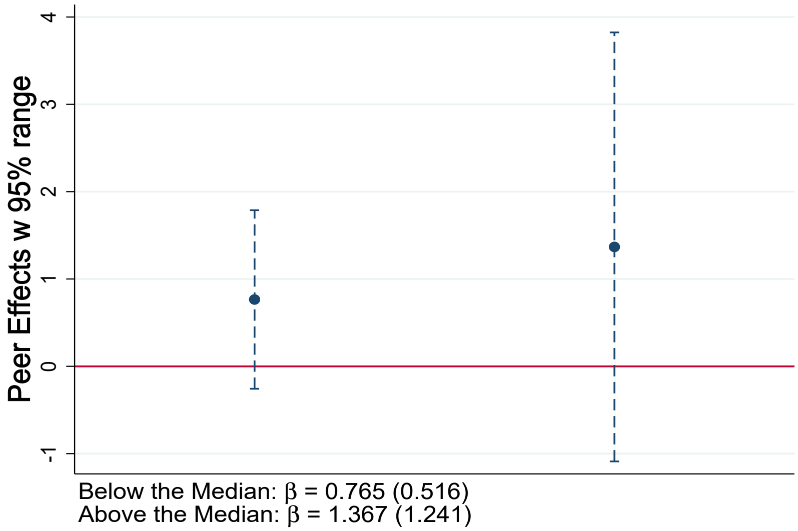
(a) Panel A: Baseline



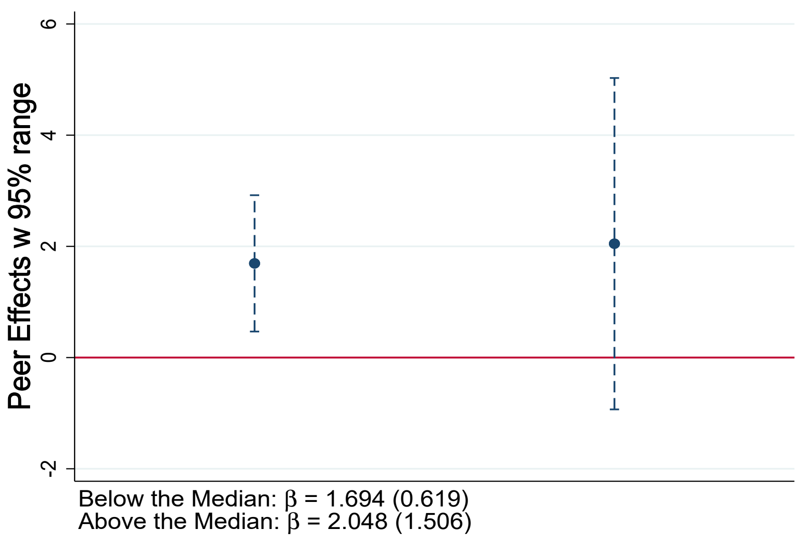
(b) Panel B: Misbehaving Male Peers to Male Students

Notes: I restrict the sample to students who do not have experienced parental divorce, separation or death. This figure presents the estimated coefficients on misbehaving male peers (male peers with a single parent) from specification (1) in Tables 3.6 and 3.7 by interacting peer measures with indicators for whether a student's household is below the poverty line. Panel A shows the estimated effects of misbehaving male peers. Panel B shows the estimated effects of misbehaving male peers interacted with the indicator for a male student. In both Panels, β is the estimated coefficient while its standard error is shown in parentheses. 95% confidence intervals are shown with dashed lines.

Figure 3.4: Effects of Misbehaving Male Peers by the Class Size



(a) Panel A: Baseline



(b) Panel B: Misbehaving Male Peers to Male Students

Notes: I restrict the sample to students who do not have experienced parental divorce, separation or death. This figure presents the estimated coefficients on misbehaving male peers (male peers with a single parent) from specification (1) in Tables 3.6 and 3.7 by interacting peer measures with indicators for whether the class size is below the median (35). Panel A shows the estimated effects of misbehaving male peers. Panel B shows the estimated effects of misbehaving male peers interacted with the indicator for a male student. In both Panels, β is the estimated coefficient while its standard error is shown in parentheses. 95% confidence intervals are shown with dashed lines.

In general, these heterogeneous effects have to be interpreted with some caution due to the lack of the statistical power. Nonetheless, my analyses indicate that class misbehaviors appear to spread on a male-to-male basis for students who live in low-income households.

3.5 Conclusion

This paper estimates the impact of misbehaving peers on the intensity of a student's own misbehavior during secondary school in South Korea. To distinguish peer effects from confounding factors, I use a random assignment setting within schools while instrumenting for misbehaving peers with peers from single-parent households.

Results indicate that misbehaving male peers statistically significantly increase the intensity of others' misbehaviors in the classroom, while misbehaving female peers do not. The RF result shows that the increase of one single misbehaving male peer among 30 students in the classroom leads to the increase of other students' MCI by around 0.03 standard deviations. The inclusion of student, peer, and teacher characteristics does not change these results much, so random assignment truly holds. Those effects are more salient for male students. Additional heterogeneous analyses suggests that misbehaving peer effects are pronounced for interactions between male students who live in low-income households.

These findings on the peer effects provide robust evidence that misbehaving boys increase the behavioral problems for other pupils in the classroom. This study therefore expands the literature on the heterogeneous effects of school peers and provides policymakers with some insights regarding changes in peer group composition. Finally, I suggest that educational policy intervention to address student misbehavior in the classroom might be effective if it mainly targets misbehaving boys.

4. RETIREMENT AND HEALTHCARE UTILIZATION: EVIDENCE FROM PENSION ELIGIBILITY AGES IN SOUTH KOREA

4.1 Introduction

In most countries, eligibility for the public pension benefit is determined by a specific cutoff age. A reform in this cutoff age generates changes in the labor activities of elderly workers as well as social and economic effects for them (5). For instance, an increase in the eligibility age for pension benefits would directly incentivize elderly workers to retire later. It may also alter their healthcare utilization as the unintended second-order effect. Combining these effects also helps in understanding the potential trade-offs between retirement and healthcare utilization.

Recent studies have examined the treatment effects of retirement on healthcare utilization.[44] Empirical findings of these studies are mixed according to different contexts.[45] Studies that focus on developed countries document a negative or no significant impact of retirement on healthcare utilization ((74); (75); (76); (77); (5); (78); (79)). In contrast, a few studies that include developing countries or some high-income countries find a positive effect of retirement on some healthcare utilization outcomes ((80); (81); (82); (83)). These findings support that retirement-related policies should be carefully developed by referring to heterogeneous evidence among different populations. In this regard, better understanding of the effects of pension eligibility both on retirement and on healthcare utilization from a variety of populations and discussing about the relationship between these outcomes can assist policymakers to better design retirement policies that do not cause increases in healthcare expenditures among older patients.

This paper presents new evidence to complement this literature by looking at how pension eligibility affects changes in retirement and healthcare utilization outcomes in South Korea. Based on

[44]To overcome the endogeneity between retirement and healthcare utilization outcomes, the majority of these studies use a regression discontinuity design (RDD) or an instrumental variable (IV) method by resting on the age-related rule for either pension program or statutory retirement.

[45]To the best of my knowledge, the published studies that focused on a single country in the economics literature have so far looked at the retirement effects in the US, Germany, Austria, Denmark, Sweden, Netherlands, China, and Vietnam.

these impacts, I additionally provide the relationship between retirement and healthcare utilization within this context. For doing these tasks, I use longitudinal data from the Korean Longitudinal Study of Aging (KLoSA) that surveyed individuals aged over 45 in South Korea. This dataset includes information on demographic characteristics and labor market activities. I combine this information with outcomes regarding inpatient care, outpatient care, and dental care utilization.

South Korea is one of the developed Asian countries with the fastest aging population (OECD). Specifically, it has 14.2% elderly proportion of the total population in 2017, which displays the entrance into "aged society" by taking just 17 years after it became "aging society" in 2000.[46] To alleviate the fiscal burden of its aging populations, the South Korean government reformed the public pension benefit structure and raised the eligibility ages for the pension benefit: the reform of the national pension system raised not only the normal pension age from 61 in 2013 to 65 in 2033, but also the early pension age from 56 to 60 during the corresponding period.[47]

For causal identification, I rely on age thresholds to receive (normal) old-age pension benefits, providing identifying variation for a regression discontinuity design (RDD) to address selection bias. In particular, I employ the donut RDD approach that drops observations near the cutoff age to account for the presence of anticipation effects and mandatory retirement. During this analysis, I interpret the changes in retirement or healthcare utilization outcomes at turning pensionable ages as the Intent-to-Treat (ITT) effect of public pension eligibility. To further show the relationship between retirement and healthcare utilization outcomes, I focus on the Local Average Treatment Effect (LATE) parameters obtained from a fuzzy RDD and a pair of RD coefficients for the ITT impact on these outcomes for subgroups that are constructed based upon covariates.

I find that crossing the pension eligibility age increases the probability of being in retirement and inpatient care utilization, while doing so has no significant effect on outpatient care utilization. Specifically, the pension eligibility results in 5.1pp increases in retirement and 7.8% of a standard

[46]Following the definition of the World Health Organization (WHO) and the United Nations (UN), an "aging society" is a society in which more than 7% of the population is 65 years or older, an "aged society" is a society in which more than 14% of the population is 65 years or older, and a "super-aged society" is a society in which more than 21% of the population is 65 years or older.

[47]In South Korea, public pension is named as the national pension. Before 2013, the pension benefits were available at the normal (early) pension age 60 (55).

deviation increases in the inpatient care index. It reduces the inpatient care index by 3.5% of a standard deviation, but the corresponding RD estimate is not statistically significant. These findings are robust to the inclusion of covariates or different specifications assuming a non-linearity of the running variable. In addition, the effect of crossing pensionable ages on healthcare utilization may not be significantly driven by the effect on health status and the selection into pension receipt. As the first pass driving these results, I also find that crossing pensionable ages has significant and positive effects both on the likelihood that the elderly receive the pension benefit and the annual amount of pension benefits. The heterogeneous analyses suggest that these effects are more likely to be pronounced for those who are men, married, or low-educated.

By connecting the RD estimates for retirement with the RD estimates for healthcare utilization, I explore the relationship between retirement and healthcare utilization. The LATE parameter from a fuzzy RDD – equivalently, scaling the ITT parameter for healthcare utilization by the ITT parameter for retirement (the first-stage parameter) – suggests that inpatient care utilization increases at retirement, whereas outpatient care utilization decreases at retirement. These effects are substantial, but likely to be driven by the low explanatory power of the first-stage estimate in the calculation of the LATE parameters. In addition, the LATE parameter for outpatient care utilization is statistically insignificant. The additional examination of the relationship by depending on the ITT parameters for covariates-based subgroups suggests that retirement might not be a strong driver for the changes in inpatient care utilization across these subgroups.

My paper is related to a growing body of research examining the impact of retirement on healthcare utilization. (76) relies on a fuzzy RDD that uses an exogenous rule of the German pensionable age and finds a reduction in outpatient care utilization. (84) use an instrumental variable (IV) approach based on an increase in the early retirement age of women in Hungary and suggest that retirement reduces inpatient and outpatient healthcare utilization. (81) focus on ten European countries and use an IV strategy. They find retirement increases the number of doctor's visits and the probability of visiting a doctor increase more than four times. (85) rest on an IV strategy that uses the US social security retirement age. They show the negative effect of retirement on the probability of hospital stay. (82) exploit a fuzzy nonparametric RDD that uses a statutory retirement age of urban populations in China and find the positive effect of retirement on healthcare utilization. (79) examines the causal effect of retirement in Denmark by applying a fuzzy RDD that relies on a discontinuity in the early retirement pensionable age and reports a reduction in inpatient and outpatient care utilization at retirement. (5) exploit exogenous variation in the early retirement age in the Austrian pension system and find a significant negative effect of retirement on healthcare utilization. (80) relies on an IV approach to estimate the treatment effect of retirement on inpatient and outpatient care utilization in urban Vietnam. He finds that retirement leads to the increased utilization for outpatient services but has no statistically insignificant effects on utilization outcomes for inpatient services. (83) complement the work of (82) by leveraging administrative data. By using a fuzzy RDD, they find the positive effect of retirement on outpatient care utilization, whereas it does not significantly affect inpatient care utilization.

The contributions of this paper are twofold. First, to the best of my knowledge, this paper is the first to explore the relationship between retirement and healthcare utilization in an Asian developed country. While there is uncertainty in interpreting the size of the average causal effect (LATE) due to the low first-stage estimate, my evidence is consistent with the finding of (82) and (81) documenting the positive relationship between retirement and inpatient care utilization in the developed country or some high-income countries.

The second contribution of this paper to the literature discusses the role of retirement in healthcare utilization by resting on the ITT parameters for covariates-based subgroups. Exploiting the ITT estimates for retirement or healthcare utilization for each subgroup, I examine the relationship between retirement and inpatient care utilization, outpatient care utilization, and dental care utilization.[48] It remains uncertain whether my finding through these ITT parameters for each subgroup can extend to other contexts. With this caveat, however, my finding suggests that the average causal effect of retirement on healthcare utilization, at least in the context similar this, should be

[48] My approach for doing this work is most similar to (86), who show that the age rule for the pension benefit affects mortality by resting on the ITT effects-based analysis. However, my paper differs from this paper in that I focus on the effects of the pension age rule on healthcare utilization for the main analysis, rather than health itself.

cautiously interpreted according to whether it is driven by the relationship across a handful of subpopulations.

4.2 Background and Data

4.2.1 The South Korean Pension System

The pension system in South Korea is currently based on the multi-pillar system of old-age income security with public-private pension role division. As public pensions, the basic old-age pension (zero pillar) protects those aged 65 or over who are below the income eligibility threshold, and the national pension (first pillar) covers all citizens except government employees, private teachers, and military personnel. In addition, the corporate pension (second pillar) provides retired employees with two forms of benefits, the defined benefit (DB) and the defined contribution (DC) plan, and finally any person can buy voluntary personal pension saving accounts (third pillar) from a financial company for old-age income protection.

These pension systems present an average gross pension replacement rate of 39.3%. Compared to the total OECD average of 52.9% (2019), South Korea is thus one of the OECD countries that have the least generous pension systems. As the most influential pension scheme, the national pension system guarantees a statutory replacement rate of 40.0% for 40 years of contribution.

To qualify for the national pension benefit, a person must have at least 10 years of contribution.[49] Only a person who satisfies these contribution years can apply for and be available for a normal pension benefit at age 60 or receive an early pension benefit at age 55 with 6 percentage points reductions for every year below the normal pension age in benefit before 2013. In 2013, the policy change regarding eligible pension ages began and these ages increased from 60 (55) to 61 (56) for the birth cohort 1953–1956, 62 for birth cohort 1957–1960, 63 for the birth cohort 1961–1964, 64 for the birth cohort 1965–1968, and 65 for the birth cohort 1969 or younger. According to the continued schedule of the pension reform regarding pension age changes, the normal (early) pension age has been incrementally increased by one year every five years and thus raised from 60 (55) to 62 (57) by 2018, and it will be 65 (60) until 2033.

[49]This is the same requirement as the US social security for pension benefit receipt.

4.2.2 The South Korean Healthcare System

South Korea is one of the countries that provide the entire population with universal access to healthcare services. The National Health Insurance (NHI) system has the biggest role to facilitate this goal and specifically covered over 50 million (97.2%) of the total population in 2017.[50] The insured are divided into two groups: employee insured and self-employee insured. The employee insured can include their spouse and children for the support and pay 6.67% (in 2020) of their average salary in contribution, though the payment rate changes every year. The self-employee insured is a person excluded from the category of insured employee. Their contribution amount usually relies on the insurance contribution points with respect to their income, property and living standard.

The patient cost-sharing structure is important for affecting the utilization of healthcare services. The amount of payments for healthcare utilization is consisted of reimbursement and non-reimbursement: the NHI covers only the amount of reimbursements; the amount of non-reimbursements should be paid by individual patients themselves or the purchase of private health insurance. Unlike the US health insurance plan, there is no deductible in the NHI. However, co-payments are required according to the type of healthcare institution, though the NHI program is available. The amount of co-payment that insured persons have to pay varies according to where the care is delivered and the type of care. Generally, in 2020, co-payment amounts for adults are 20% of inpatient hospital care, 30–60% of outpatient hospital care, and 20–40% of pharmacy bills.[51] This co-payment system has no difference between employees and retirees. That is, after retirement, insured persons can continue to have unlimited access to healthcare services with the lower out-of-pocket expense.

In addition, South Korean patients have unrestricted access to healthcare providers. That is, there is no restriction for a patient's choices of healthcare services, whether such providers are

[50]The remaining 2.8%, which is classified as low-income households, is supported by the Medical Aid (MA) program.

[51]Regarding co-payment rates for outpatient hospital care, the use of clinic, hospital, general hospital, and tertiary hospital is 30%, 40%, 50%, and 60%, respectively.

hospitals or physicians' offices (or clinics). Patients also have no limit on the number of visits. They can directly access specialist care without going through a gatekeeper or a referral system.

4.2.3 Data and Sample Statistics

The dataset for the analysis is the Korean Longitudinal Study of Aging (KLoSA), a nationally representative panel survey that collects data from 10,254 stratified randomly chosen individuals aged 45 years and older in 2006. Then, 920 individuals who were born from 1962 through 1963 were additionally surveyed in 2014. Because this survey is conducted every two years, I obtain longitudinal individual-level data for seven rounds of surveys between 2006 and 2018.

I restrict the sample to observations for those aged 45–75 in all seven waves of the survey. I exclude those who neither work nor retire. I further drop the observations with missing information on retirement status. Then, I construct the stacked dataset with 42,213 individual-year observations for 8,994 individuals. 78.6% of these individuals are observed more than twice in the dataset. Because I rely on a donut RDD approach, I additionally exclude 5,928 observations within a bandwidth of four bins from the pensionable age. Hence, my final sample for the analysis is consisted of 36,285 observations.

The KloSA includes several measures of individual healthcare and health. I use three healthcare utilization indexes as the outcome variables in the main analysis while exploiting some health-related measures to uncover the potential threats to the main results. The healthcare measures are the following: 1) *Inpatient care*: the incidence, the number of times the individuals received inpatient care, and the total amount of out-of-pocket expenditure for the care; 2) *Outpatient care*: the incidence, the number of times the individuals received outpatient care, and the total amount of out-of-pocket expenditure for the care; and 3) *Dental care*: the incidence, the number of times the individuals received dental care, and the total amount of out-of-pocket expenditure for the care.[52] By exploiting these measures, I construct the composite indices of inpatient, outpatient,

[52]These healthcare measures are recorded, based on changes after the last interview date. Hence, the majority of these measures are aggregated within the past two years. Exceptionally, information on these measures in the first wave is surveyed, based on the previous year. In order to use annual measures for my analysis, I divide the number of times the individuals receive healthcare and the total amount of out-of-pocket for healthcare by two after the second wave. However, it is not possible to apply this way to measure the annual incidence for inpatient, outpatient, and dental

and dental care. To calculate these three indices, I rest on the method of (4). I describe the detail of this method in Appendix Section A.2. This aggregation might improve statistical power to detect effects that go in the same direction for my analysis. The measures for individuals' health status are the followings: 1) *Physical health*: high blood pressure, diabetes, cancer, lung disease, live disease, heart disease, brain disease, and bone disease; and 2) *Mental health*: psychiatric illness.

In addition, the KloSA includes information on retirement status and other labor market activities of individuals. As the treatment, I define individuals as "retired" if they report in (complete or partial) retirement and their employment status is "not working."[53] In other words, I do not consider individuals as "retired" if they report being retired, but also report that they are either working full- or part-time. Hence, retirement is considered as a change from working to non-working in this paper.

Table 4.1 presents summary statistics for 36,285 observations. Average age of these observations is 62.20 years. 44% of observations are men. 77% of them live in urban areas. 11% of the sample is those who do not have a BA degree. Those who are in retirement is 27% of the sample. 21% of the sample received the pension benefit last year. The average amount of pension benefits of the sample is 69.1 (10,000 KRW).[54] The likelihood of having used inpatient care last year or within the past two years is 11%, while the likelihood of having used outpatient care last year or within the past two years is 70%. Also, the likelihood of having used dental care last year or within the past two years is 25%. Individuals stay in hospital 0.09 times per year, while they visit a doctor 5.39 times per year. They also visit a dental hospital or clinic 1.01 times per year. Average out-of-pocket annual expenditure on inpatient care, outpatient care, and dental care is 8.16 (10,000 KRW), 28.53 (10,000 KRW), and 13.76 (10,000 KRW), respectively.[55]

care: thus, I use the original healthcare incidence reported by individuals for my analysis. Compared to the measures based on administrative data or a couple of the past months, it is noted that my measures on healthcare utilization are more likely to have recall bias and other measurement errors.

[53] My definition of the retirement status of individuals is similar to (76) and (82).

[54] The average amount of pension benefits for those who received pension last year is 328.2 (10,000 KRW).

[55] By limiting those who have used each healthcare service, the corresponding expenditure is 74.65 (10,000 KRW), 40.58 (10,000 KRW), and 55.05 (10,000 KRW), respectively.

Table 4.1: Summary Statistics

	Mean	Std. Dev	Min	Max	N
A. Healthcare Utilization Outcomes					
Inpatient Care Index	0.00	1.00	-27.98	43.84	36,093
Outpatient Care Index	0.00	1.00	-1.05	48.43	35,985
Dental Care Index	0.00	1.00	-0.44	30.84	36,170
Inpatient Care Incidence	0.11	0.32	0.00	1.00	36,273
Hospital/Clinic Stays	0.09	0.39	0.00	12.00	36,273
Inpatient Care Cost (10,000 KRW)	8.16	61.93	0.00	4,950	36,093
Outpatient Care Incidence	0.70	0.46	0.00	1.00	36,248
Hospital/Clinic Visits	5.69	11.70	0.00	360.00	36,248
Outpatient Care Cost (10,000 KRW)	28.53	116.00	0.00	10,500	36,009
Dental Care Incidence	0.25	0.43	0.00	1.00	36,277
Dental Hospital/Clinic Visits	1.01	4.12	0.00	310.00	36,277
Dental Care Cost (10,000 KRW)	13.76	73.37	0.00	3,600	36,170
B. National Pension-Related Outcomes					
Pension Received	0.21	0.41	0.00	1.00	36,267
Pension Benefit (10,000 KRW)	69.10	189.90	0.00	3,600	36,267
C. Dummy for the Age Cutoff and Retirement					
1[Age $\geq$ Cutoff]	0.60	0.49	0.00	1.00	36,285
Retirement	0.27	0.44	0.00	1.00	36,284
D. Individual Characteristics (including Covariates)					
Age	62.29	8.50	45.00	75.00	36,285
Male	0.44	0.50	0.00	1.00	36,285
Spouse	0.82	0.38	0.00	1.00	36,285
Offspring	0.98	0.15	0.00	1.00	36,274
Low Education (BA-)	0.11	0.32	0.00	1.00	36,275
Urban Area	0.77	0.42	0.00	1.00	36,285

Notes: Inpatient/Outpatient/Dental care incidence represents the probability of having inpatient/outpatient/dental care services last year or within the past two years. (Dental) Hospital stays or visits represent the number of times individuals received healthcare services within the past one year. Inpatient/outpatient/dental care cost is the out-of-pocket expenditure for inpatient/outpatient/dental care service use within the past one year. Pension received is the indicator for whether individuals received a pension in the last year. Retirement is the indicator for whether individuals are in retirement.

Appendix Table F11 summarizes the difference in means between individuals above and below the threshold age. Individuals above the threshold age have a higher probability of receiving a pension or being in retirement than those below the threshold age. In addition, individuals above the threshold age have a higher pension benefits, inpatient care index, and outpatient care index than those below the threshold age. The difference in the dental care index is smaller between individuals above and below the threshold age compared to inpatient and outpatient care indices. All differences are statistically significant at the 1% level (Column (3)).

4.3 Empirical Strategy

4.3.1 Regression Discontinuity Design (RDD)

To get the causal effects of the pension eligibility on retirement and healthcare utilization, it is necessary to resolve its endogeneity. The most likely problem regarding the endogeneity is reverse causality. For example, an individual who plans to retire but needs post-retirement income is more likely to pursue obtaining the eligibility for the pension benefit. Moreover, the elderly who regularly use medical services may prefer the pension benefit to pay for them. In order to avoid these concerns, I therefore use the exogenous rule (pension eligibility ages) for the pension benefit to compare the outcomes of those who are eligible to those who are not eligible.

I perform a regression discontinuity design (RDD) to estimate the causal impact of the normal pension age on the outcome of interest, including retirement status and healthcare utilization ((87); (88)). The assignment variable is one's age in a given year of my stacked dataset. The running variable is defined by subtracting the pension age cutoff for that year from the assignment variable. I estimate the following reduced-form equation:

$$Y_{it} = \alpha_0 + \alpha_1 1[Age_{it} \geq C_{bt}] + \alpha_2(Age_{it} - C_{bt}) + \alpha_3 1[Age_{it} \geq C_{bt}] \times (Age_{it} - C_{bt}) + \delta X_{it} + \mu_{it} \tag{4.1}$$

where Y_{it} is the outcome variable for individual i in year t. $1[Age_{it} \geq C_{bt}]$ is the indicator function that is "1" if age (Age_{it}) of individual i in year t is greater than or equal to the pension age cutoff C_{bt} for birth cohort b in year t, and otherwise "0".[56] By interacting the running variable ($Age_{it} - C_{bt}$) with the treatment variable ($1[Age_{it} \geq C_{bt}]$), I allow the slopes of our fitted lines to differ on either side of the age cutoff (C_{bt}). X_{it} is a set of covariates, including the indicators for whether an individual is male, for whether an individual has a spouse, for whether an individual has at least one offspring, for whether an individual has a BA degree or higher, and for whether an

[56]The age cutoff (C_{bt}) is defined as 60 for all birth cohorts before 2013. After 2013, it is defined as 61 for the birth cohort 1953–1956, 62 for the birth cohort 1957–1960, and 63 for the birth cohort 1961–1964.

individual lives in an urban area. Controlling for these covariates is not required for identification. However, doing so could alter the precision of my estimates, but should not significantly change the estimated coefficient of interest if my identifying assumption holds. μ_{it} is the idiosyncratic error term. Because each individual repeatedly serves as a separate observation for each year, I cluster the standard errors at the individual level in all specifications.

The estimator of interest is α_1 which represents the Intent-to-Treat (ITT) effect. The model above is based on a local linear regression model with a rectangular kernel.[57] To circumvent concern about the anticipation effect and mandatory retirement, I drop observations within donut hole sizes of two ages. For example, because the normal pension age for the pension benefit is well-informed, a person who is aged just below the pension age but unhealthy might decide to retire early. Or, people who are more likely to care about their health might accelerate the healthcare service use when they are close to the normal pension age. Moreover, some Korean firms force employees to retire at age 58–62 according to their own retirement rules. Hence, my approach to estimate the normal pension age effects on retirement and healthcare utilization is, by nature, the donut RDD

4.3.2 Tests for Validity of the RDD

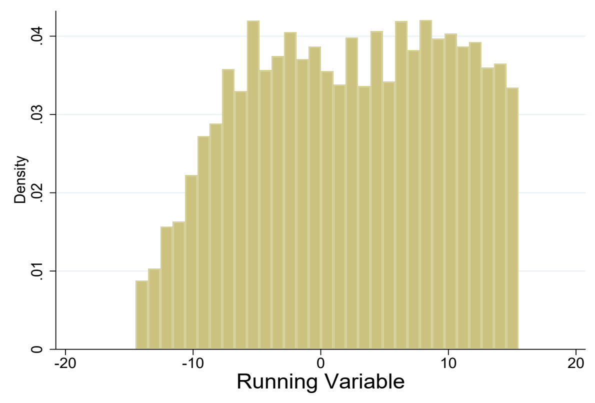
The main identifying assumption of my RDD is no existence of manipulation of age around the age cutoff. Figure 4.1 shows the frequency of the running variable (the normalized age). Panel A plots the histogram using one-age bins. There is little evidence of large heaps at the running variable. Panel B shows the binscattered plot along with the quadratic fitted lines. This panel also shows that the mean frequency of the running variable seems to be smooth across the age cutoff. To formally test for possible manipulation of age around the age cutoff, I compute the T-stat of -0.3473 with a p-value of 0.7284 obtained by robust bias-correction.[58] Because the p-value for the

[57]This approach follows (89). To avoid the overfitting problem of higher-order polynomials, I use a local linear regression model as the basic specification (90). This approach also gives more weights to the eligibility age thresholds and does not require any strong functional assumptions on the data. However, nonlinearities in the running variable may generate false discontinuities. To check this possibility, I additionally look at the RD parameters by estimating a global quadratic model in Appendix C.

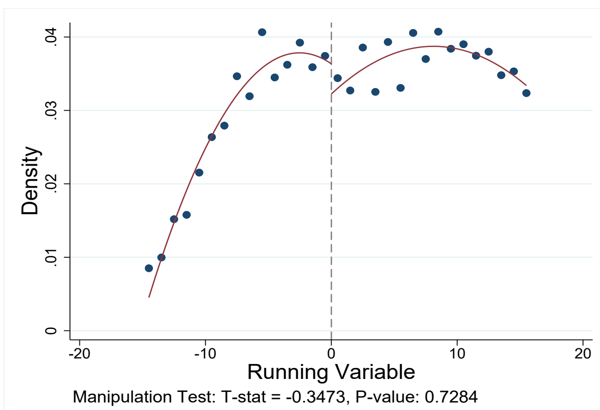
[58]This manipulation test is a hypothesis test on the continuity of the density of the running variable at the cutoff point. The density of the running variable is estimated by using a second-order local-polynomial density estimator

manipulation test is greater than 0.05, I conclude that there is no statistical evidence of systematic manipulation of age around the age cutoff. As I noted earlier, I focus on a donut specification, excluding two ages around the age cutoff. While I find no evidence of manipulation age near the age cutoff, removing this donut helps to address this concern.

Figure 4.1: Distribution of Running Variable



(a) Panel A: Histogram



(b) Panel B: Binscattered Plot

Notes: This figure depicts the distribution of running variable, as computed by distracting the cutoff from ages. Panel A shows the histogram by running variable. Panel B shows the bin-scattered plot with the quadratic fitted lines. T-statistics and P-value for the manipulation test are obtained by the robust bias-correction (91).

based on the c.d.f. of my sample data. To perform this test, I refer to (91).

To further test the validity of the RDD approach, I examine whether the observable predetermined characteristics (X_{it}) defined in Equation (4.1) are similar around the cutoff age. If there are discontinuities in these characteristics, then one would be concerned that the results could be driven by an unobserved confounder. In Columns (1)–(5) of Table 4.2, I summarize the RD estimates for these characteristics (covariates). The majority of RD estimates are small and statistically insignificant, indicating that there is little evidence that those characteristics vary discontinuously across the age cutoff. Moreover, I create the indexes of predicted inpatient care, outpatient care, and dental care utilization. Columns (6)–(8) of Table 4.2 show that the RD estimates are close to zero, and no estimates are statistically significant at conventional levels. Hence, those indexes are also smooth across the age cutoff. Consistent with these results, I find no significant differences around the age cutoff in covariates or predicted indexes in Appendix Figure E5.

Table 4.2: RD Estimates for Baseline Covariates and Predicted Indexes

	Covariates					**Indexes**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Male	Spouse	Offspring	Low Edu	Urban Area	Inpatient Care	Outpatient Care	Dental Care
$1[Age_{it} \geq C_{bt}]$	0.011	0.012	0.006	0.004	-0.020*	0.001	-0.001	-0.000
	(0.014)	(0.010)	(0.005)	(0.009)	(0.011)	(0.001)	(0.002)	(0.001)
Observations	36,285	36,285	36,274	36,275	36,285	36,072	35,964	36,149
R^2	0.000	0.037	0.004	0.024	0.018	0.059	0.021	0.016

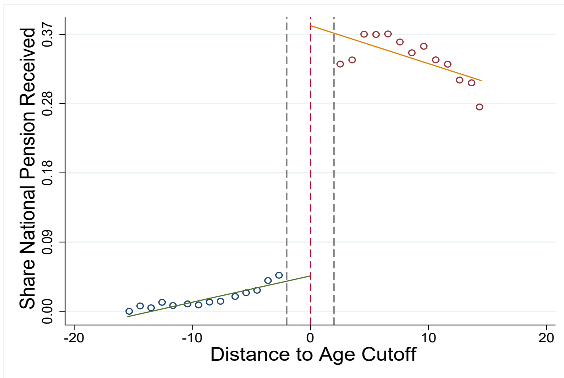
Notes: Each column represents results from a local linear regression. The dependent variables are the indicators for (1) whether an individual is male, for (2) whether an individual has a spouse, for (3) whether an individual has at least one offspring, (4) whether an individual has a BA degree or higher, and (5) whether an individual lives in an urban area. The dependent variables in Columns (6)–(8) are the predicted values of (6) inpatient care index, (7) outpatient care index, and (8) dental care index from a separate regression of each index on covariates in Columns (1)∼(5). Standard errors, shown in parentheses, are clustered at the individual level.
*p<0.10, ** p<0.05, *** p<0.01.

4.4 Empirical Results

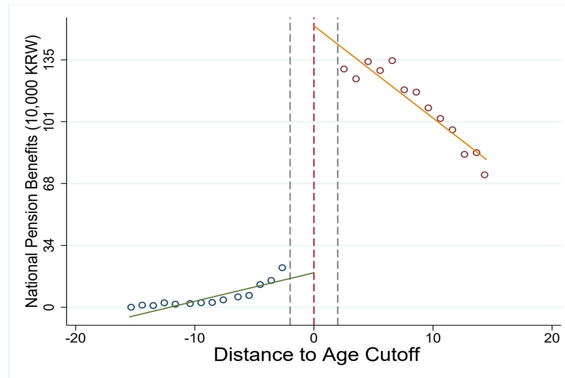
4.4.1 Effects on Pension-related Changes

As the first pass at the effect of crossing pension eligibility, I begin with the RD estimates looking at the effects of the age cutoff on pension-related outcomes. Among these outcomes, a discontinuously increased pension benefits at the age cutoff might be the important mechanism driving the effect of pension eligibility both on retirement and healthcare utilization in the rest of this paper. Figure 4.2 shows graphical evidence of the effect of pension eligibility on these outcomes. Panels A, C, and E plot the likelihood that elderly people receive the pension benefits while Panels B, D and F plot the log of the annual amount of pension benefits. The age cutoffs are normalized at zero. The linear lines on both the left- and right-hand sides graphically represent the linear fitted values of the estimated bins. The dotted lines on both sides also represent the donut hole size (four bins near the age cutoff). The sample on the left-hand side shows the outcomes for non-affected people who were not eligible for the (normal) pension benefit. The sample of the right-hand side shows affected people who were eligible for the (normal) pension benefit. Overall, there are clear and noticeable discontinuities in the outcomes around the age cutoff for all individuals (Panels A and B), men (Panels C and D), and women (Panels E and F). This suggests that affected people (treatment group) are more likely to be eligible for the pension benefit or receive the pension benefit than unaffected people (control group) at the age cutoff.

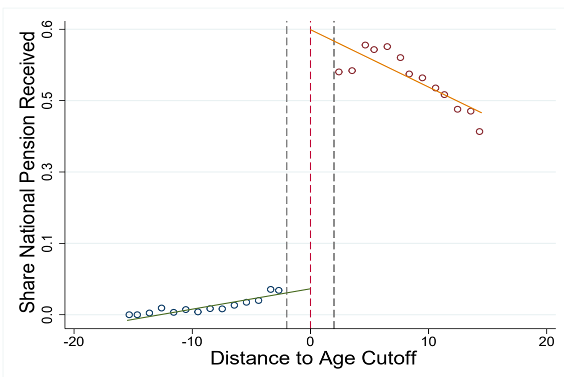
Figure 4.2: Discontinuity in Pension-related Outcomes



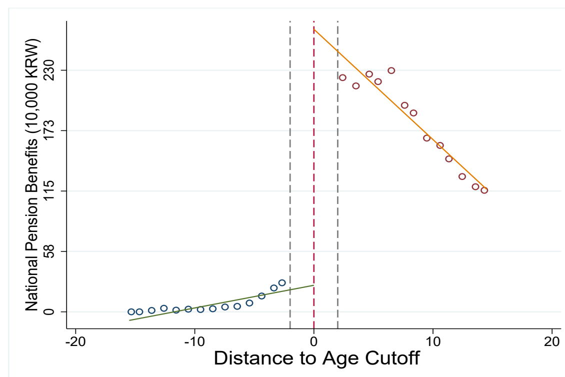
(a) Panel A: Pension Received (All)



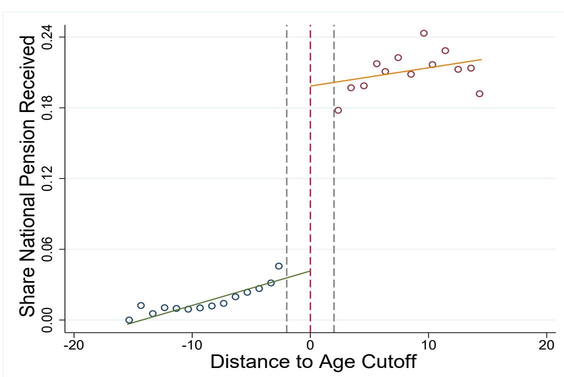
(b) Panel B: Pension Benefit (All)



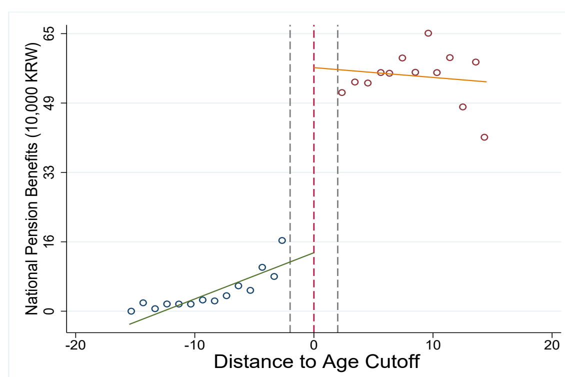
(c) Panel C: Pension Received (Men)



(d) Panel D: Pension Benefit (Men)



(e) Panel E: Pension Received (Women)



(f) Panel F: Pension Benefit (Women)

Notes: This figure depicts the discontinuity in Public Pension-related outcomes. Each panel shows the discontinuity at the cutoff (0 on the x-axis) for the fraction of individuals who received pension benefits or the mean of pension benefits that they received for all samples ((a) and (b)), men ((c) and (d)), and women ((e) and (f)). The outcome on the y-axis reports the fraction or the mean by the distance from the cutoff (red dotted line) with 1 width of bins. Two gray dotted lines represent the donut hole size, excluding 4 bins around the age cutoff. The best-fit solid lines are linear fitted on each side of the cutoff.

Table 4.3 shows the corresponding RD estimates for the impact of the pension eligibility ages on pension-related outcomes. All estimates are the local linear RD estimates. The outcomes in Columns (1)-(3) are the likelihood that people receive the pension benefit while the outcomes in Columns (4)-(6) are the log of the annual amount of pension benefits. In Panel A, all RD estimates are positive and statistically significant at the 1% level. These RD estimates are robust to the inclusion of covariates (Panel B). In Panel A, Column (1), the magnitude of the estimated likelihood of receiving the pension benefit for all individuals is 33.0pp. In Panel A, Column (4), the magnitude of the estimated (annual) pension benefits for those individuals is 131.49 (10,000 KRW). In Columns (2)-(3) and (5)-(6), the estimated RD parameters vary across genders: that is, the estimated coefficients for men are much greater than those for women. Hence, the figure and table show that crossing pension eligibility ages positively affects both the likelihood that the elderly receives the pension benefit and the annual amount of pension benefits.

Table 4.3: RD Estimates on Pension-related Outcomes

	Pension Received			**Pension Benefit**		
	(1)	(2)	(3)	(4)	(5)	(6)
	All	Men	Women	All	Men	Women
A. Without Covariates						
$1[Age_{it} \geq C_{bt}]$	0.330***	0.537***	0.156***	131.49***	237.43***	42.70***
	(0.011)	(0.017)	(0.013)	(5.822)	(10.82)	(4.623)
Observations	36,154	15,875	20,279	36,154	15,875	20,279
B. With Covariates						
$1[Age_{it} \geq C_{bt}]$	0.330***	0.535***	0.159***	131.1***	237.6***	43.52***
	(0.011)	(0.017)	(0.012)	(5.772)	(10.85)	(4.557)
Observations	36,132	15,864	20,268	36,132	15,864	20,268

Notes: Each cell represents results from a local linear regression. The dependent variables are the probability of receiving pension benefits or the annual amount of pension benefits (10,000 KRW). The covariates include the indicators for whether an individual is male, for whether an individual has a spouse, for whether an individual has at least one offspring, whether an individual has a BA or higher degree, and whether an individual lives in a single detached house. Standard errors, shown in parentheses, are clustered at the individual level.
*p<0.10, ** p<0.05, *** p<0.01.

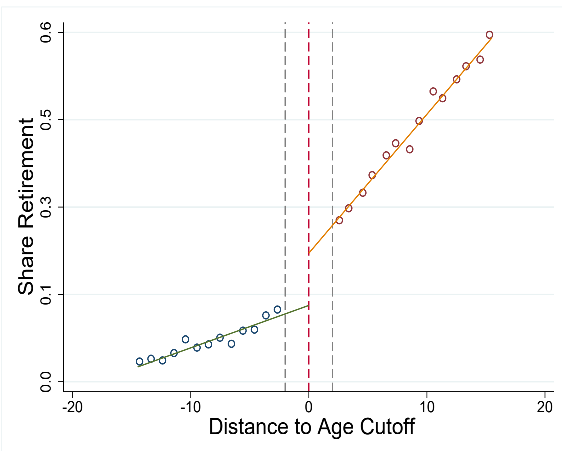
4.4.2 Effects on Retirement and Healthcare Utilization

I turn to the impact of pension eligibility on retirement or healthcare utilization. Specifically, I investigate whether crossing the age cutoff is associated with increases in retirement, and then ask whether crossing the age cutoff changes healthcare utilization.

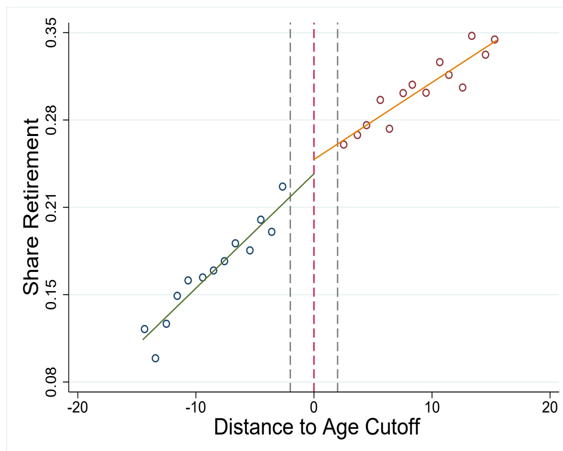
Figure 4.3 shows the likelihood of being in retirement for those just above and just below the eligibility threshold at the age cutoff. The age cutoff is normalized at zero. The sample on the left-hand side is an individual before pension eligibility ages, while the sample on the right-hand side is an individual after pension eligibility ages. There is a graphically clear jump around the age cutoff for all individuals (Panel A). The graphical jump is revealed for men at the age cutoff (Panel B), whereas there is no significant jump for women at the age cutoff (Panel C).

Figure 4.3: Discontinuity in Retirement

(a) Panel A: All



(b) Panel B: Men



(c) Panel C: Women

Notes: This figure depicts the discontinuity in retirement. Each panel shows the discontinuity at the cutoff (0 on x-axis) for (a) all samples, (b) men, and (c) women, respectively. The share of retirement on the y-axis reports retirement rate by the distance from the cutoff (dotted line) with 1 width of bins. Two gray dotted lines represent the donut hole size, excluding 4 bins around the age cutoff. The best-fit solid lines are linear fitted on each side of the cutoff.

Table 4.4 reports the corresponding RD estimates for the impact of crossing pensionable ages on retirement. All estimated parameters are local linear estimates. In Panel A, the RD estimates are positive, indicating that threshold-crossing leads to increases in retirement. The magnitude of the difference in retirement suggests that age cutoff-crossing increases the probability of being in retirement by 5.1pp (Column (1)). In addition, the RD parameter for this effect is statistically significant at the 1% level. Based on the standard deviation of the retirement rate (42%), the effect size for retirement is small.[59] Columns (2) and (3) show that the effect is larger and still statistically significant for men, whereas the effect is smaller and statistically insignificant for women. This finding suggests that the change in retirement near the eligibility at the age cutoff is particularly revealed for men. Adding covariates does not significantly change these results (Panel B).

Table 4.4: RD Estimates on Retirement

	(1) All	(2) Men	(3) Women
A. Without Covariates			
$1[Age_{it} \geq C_{bt}]$	0.051***	0.098***	0.010
	(0.012)	(0.016)	(0.016)
Observations	36,284	15,927	20,357
B. With Covariates			
$1[Age_{it} \geq C_{bt}]$	0.053***	0.101***	0.011
	(0.012)	(0.016)	(0.016)
Observations	36,263	15,916	20,347

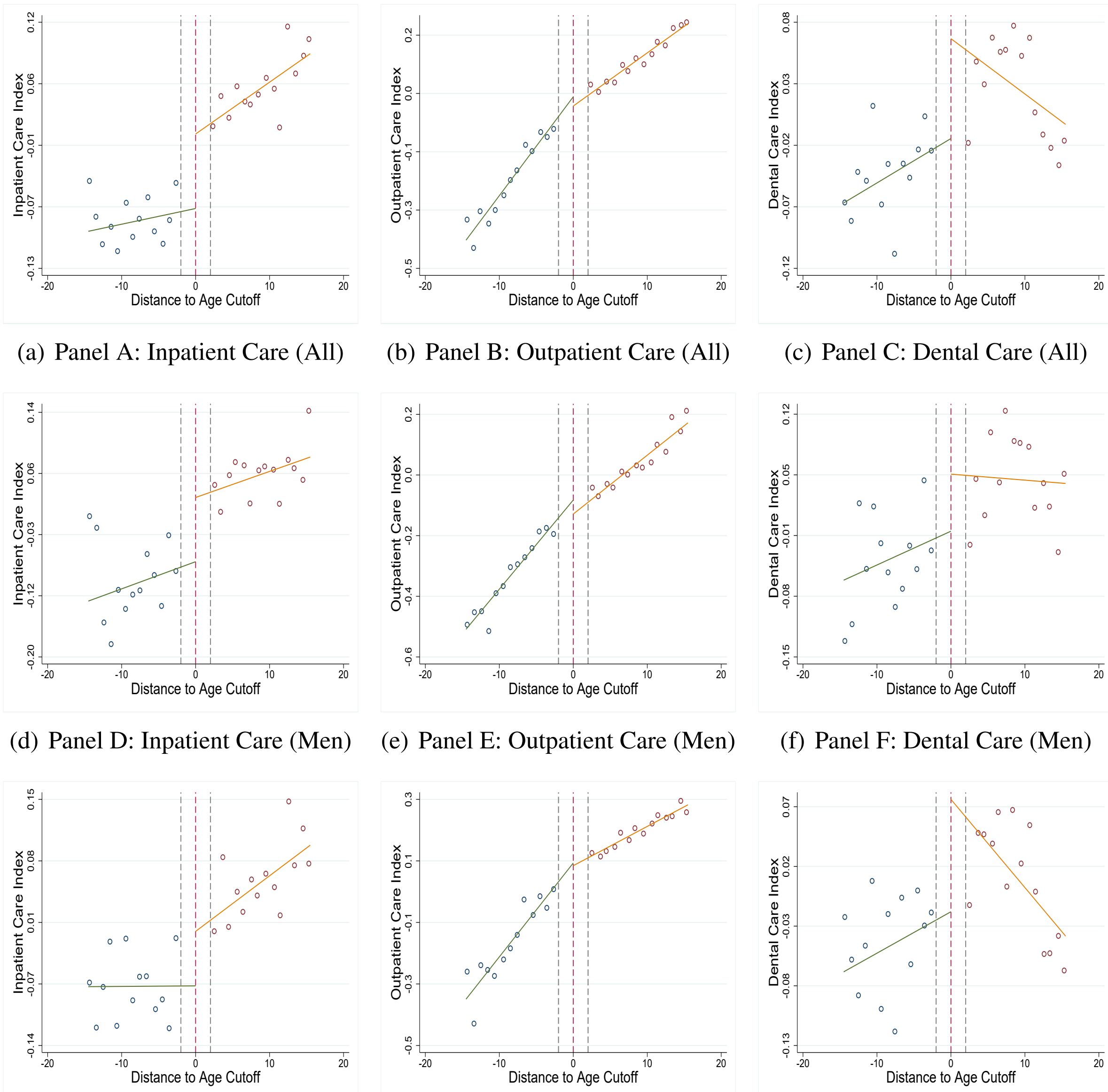
Notes: Each cell represents results from a local linear regression. The dependent variable is the dummy for retirement. The covariates include the indicators for whether an individual is male, for whether an individual has a spouse, for whether an individual has at least one offspring, whether an individual has a BA or higher degree, and whether an individual lives in a single detached house. Standard errors, shown in parentheses, are clustered at the individual level.
*p<0.10, ** p<0.05, *** p<0.01.

Next, the main outcome of interest is healthcare utilization of the elderly. As I noted earlier, I use three indices of inpatient, outpatient, and dental care as healthcare utilization outcomes. Figure 4.4 provides a graphical representation of the RD estimates for the effect of crossing pensionable ages on those three outcomes. Panels A, B, and C show that crossing the age cutoff leads to a

[59] A 5.1pp increased discontinuity in retirement accounts for 10.4% (5.1/42) of a standard deviation of the retirement rate. Putting the effect size in the context also suggests that the estimate (5.1pp) is less than most in the existing literature. For example, the estimated pension eligibility effects on retirement are 14.8pp (85) and 15pp ((76); (5)).

graphical jump up for both inpatient and dental care or a graphical jump down for outpatient care. The similar pattern in the effect on three outcomes is graphically shown across gender, except for no clear jump in the outpatient care index for women (Panels D, E, F, G, H, and I)

Figure 4.4: Discontinuity in Healthcare Utilization



(a) Panel A: Inpatient Care (All) (b) Panel B: Outpatient Care (All) (c) Panel C: Dental Care (All)

(d) Panel D: Inpatient Care (Men) (e) Panel E: Outpatient Care (Men) (f) Panel F: Dental Care (Men)

(g) Panel G: Inpatient Care (Women) (h) Panel H: Outpatient Care (Women) (i) Panel I: Dental Care (Women)

Notes: This figure depicts the discontinuity in healthcare utilization outcomes. In each panel, the value on the y-axis reports the mean outcome by the distance from the cutoff (dotted line) with one width of bins. Two gray dotted lines represent the donut hole size, excluding 4 bins around the age cutoff. The best-fit solid lines are linear fitted on each side of the cutoff.

The corresponding RD estimates in Panel A of Table 4.5 show that crossing pensionable ages positively affects both inpatient and dental care indices for all individuals (Columns (1) and (3)). Specifically, the magnitude of the estimated effect indicates that doing so increases the inpatient care index by 0.078 standard deviations and the dental care index by 0.077 standard deviations. However, pension eligibility ages-crossing negatively affects the outpatient care index. The corresponding estimate (-0.035) is economically small and statistically insignificant. Hence, I conclude that turning to pensionable ages has no significant effect on outpatient care utilization. Panels B and C break out the estimates by gender. Doing so leads to reduced statistical power due to the decreased sample size. Hence, the estimates are less precise than for all individuals. The sign of all estimates for both men and women remains unchanged. The effect size for RD estimates is not significantly different from the result in Panel A. However, the RD parameters for inpatient and outpatient care indexes for men are greater than for women, whereas the RD parameter for dental care index for men is less than that for women. This finding suggests that the effect on inpatient care utilization and outpatient care utilization is driven by the effect for men rather than that for women. The inclusion of covariates does not affect point estimates in a meaningful way, consistent with the identifying assumption (Appendix Table F12).

Table 4.5: RD Estimates on Healthcare Utilization (ITT)

	(1) Inpatient Care Index	(2) Outpatient Care Index	(3) Dental Care Index
A. Without Covariates, All			
$1[Age_{it} \geq C_{bt}]$	0.078***	-0.035	0.077***
	(0.026)	(0.026)	(0.028)
Observations	36,093	35,985	36,170
B. Without Covariates, Men			
$1[Age_{it} \geq C_{bt}]$	0.089**	-0.051	0.061
	(0.044)	(0.036)	(0.044)
Observations	15,834	15,801	15,878
C. Without Covariates, Women			
$1[Age_{it} \geq C_{bt}]$	0.067**	-0.018	0.089**
	(0.032)	(0.036)	(0.036)
Observations	20,259	20,184	20,292

Notes: Each cell represents results from a local linear regression. The dependent variables in Columns (1)–(3) are the indices for inpatient, outpatient, or dental cares. The process to construct these indexes is outlined by (4). Standard errors, shown in parentheses, are clustered at the individual level.
*p<0.10, ** p<0.05, *** p<0.01.

To investigate the additional robustness of the results in Table 4.5, I estimate Equation (4.1) using a variety of donut holes or bandwidths. Appendix Figure E6 shows how the local linear RD parameter for healthcare utilization varies by donut size. In each panel, the size of the donut hole is larger when moving from left to right. Panels A, B, and C show the RD estimates for all individuals' inpatient, outpatient, and dental care indices across donut size. Using larger donut holes leads to the general difference in estimates, though using very smaller donut holes provides similar estimates for the inpatient care index (Panel A). However, most of the RD parameters using larger donut holes are imprecisely estimated due to the decreased sample size. These trends in the estimates across the donut size are similarly shown across genders (Panels D, E, F, G, H, and I).

On the other hand, Appendix Figure E7 shows the local linear RD parameter with respect to the choice of the bandwidth. In each panel, the bandwidth size is smaller, as moving from the right and the left. The estimates for the inpatient care index seem to be robust to different bandwidth sizes (Panel A). The RD parameters for both outpatient and dental care indexes are smaller or even flip over with smaller bandwidth sizes (Panels B and C).[60] In particular, the changed sign of the effect on the dental care index with the narrowest bandwidth (8) does not allow me to make a strong conclusion for the positive estimate as shown in Panel A, Column (3) of Table 4.5, although this estimate for the largest bandwidth (16) is statistically significant at the 1% level. Overall, the estimates for all RD parameters generally become less precise with smaller bandwidths because of smaller sample sizes (Panels A, B, and C). The RD parameters for men or women have similar trends in the results across the bandwidth size (Panels D, E, F, G, H, and I). In addition, the global quadratic RD estimates according to the choice of the bandwidth are provided in Appendix Figure E8.[61] The estimated RD parameters across bandwidths are similar to the results in Appendix Figure E7. This also suggests that the main results obtained from a local linear regression are robust to the different specification that assumes a non-linearity of the running variable.

Table 4.6 also explores heterogeneity in effects of pension eligibility by covariates other than gender. Panel A shows that the effect on retirement appears to be driven by those who have at least one offspring, are low educated, or live in an urban area. Similarly, this pattern for heterogeneous effects are shown in the effect on the inpatient care index (Panel B) and is also pronounced in married individuals. Panel C shows that the effect on the outpatient care index is noticeable for those who are low-educated, or live in a rural area. Most of differential effects on the dental care index are similar to the effects on the inpatient care index, though the effect for those who are highly-educated is larger than for those who are low-educated.

[60] In Figure 4.4, I observe some bins that are likely to be far from the fit line (red line) near the age cutoff. As the bandwidth is narrower, this may lead to the changed sign of RD estimates obtained by estimating a local linear regression.

[61] To estimate a global quadratic regression, I add quadratic terms in the running variable in the specification.

Table 4.6: Heterogeneous RD estimates both on Retirement and Healthcare Utilization (ITT)

	(1) All	(2) Married	(3) Single	(4) Offspring	(5) No Offspring	(6) Low Education	(7) High Education	(8) Urban Area	(9) Rural Area
A. Retirement									
$1[Age_{it} \geq C_{bt}]$	0.051***	0.043***	0.057	0.054***	-0.027	0.197***	0.031**	0.055***	0.035
	(0.012)	(0.013)	(0.033)	(0.012)	(0.077)	(0.037)	(0.012)	(0.014)	(0.022)
Observations	36,284	29,931	6,353	35,381	892	4,160	32,114	27,791	8,493
B. Inpatient Care Index									
$1[Age_{it} \geq C_{bt}]$	0.078***	0.081***	0.026	0.077***	0.194	0.165**	0.062**	0.080***	0.073
	(0.027)	(0.028)	(0.088)	(0.027)	(0.235)	(0.074)	(0.029)	(0.029)	(0.065)
Observations	36,093	29,801	6,292	35,197	885	4,150	31,933	27,656	8,437
C. Outpatient Care Index									
$1[Age_{it} \geq C_{bt}]$	-0.035	-0.042	-0.004	-0.039	0.181	-0.153**	-0.026	-0.018	-0.137**
	(0.026)	(0.027)	(0.081)	(0.026)	(0.161)	(0.067)	(0.028)	(0.030)	(0.058)
Observations	35,985	29,740	6,245	35,111	863	4,147	31,828	27,559	8,426
D. Dental Care Index									
$1[Age_{it} \geq C_{bt}]$	0.077***	0.074**	0.095	0.076***	0.206	-0.061	0.093***	0.094***	-0.005
	(0.028)	(0.030)	(0.076)	(0.028)	(0.174)	(0.082)	(0.030)	(0.032)	(0.060)
Observations	36,170	29,851	6,319	35,274	885	4,150	32,010	27,703	8,467

Notes: Each cell represents results from a local linear regression. The dependent variables in Columns (1)–(3) are the indices for inpatient, outpatient, or dental cares. The process to construct these indexes is outlined by (4). Standard errors, shown in parentheses, are clustered at the individual level.
*p<0.10, ** p<0.05, *** p<0.01.

In addition, I examine how threshold-crossing affects individual outcomes that are used to calculate each healthcare utilization index. The RD estimates for these outcomes are provided in Appendix Table F13. Panel A of this table reports RD estimates for all individuals. I find positive effects for inpatient care outcomes: the likelihood of using inpatient care services increased by 1.6pp (Column (1)), the number of times the individuals hospitalize raised by 0.035 stays (Column (2)), and the total amount of expenditure for inpatient care increased by 2.657 (10,000 KRW) (Column (3)). In contrast, crossing the age cutoff negatively affects two of outpatient care outcomes: the likelihood of using outpatient care services decreased by 4.6pp (Column (4)) and the total amount of out-of-pocket expenditure for the outpatient care decreased by 0.243 (10,000 KRW) (Column (6)). Interestingly, it increases the number of times the individuals visit a hospital or a doctor's office by 0.241 visits. This may be due to the increased visits for those who use outpatient care at the pensionable age.[62] The effects for dental care outcomes are all positive: the likelihood of using dental care services increased by 2.8pp (Column (7)), the number of times the individuals visit a dental hospital or a dentist's office increased by 0.374 visits (Column (8)), and the total amount of out-of-pocket expenditure for the dental care increased by 1.657 (10,000 KRW) (Column (9)). All RD parameters for the incidence are statistically significant at conventional levels (Columns (1), (4), and (7)). Most of the magnitudes for RD parameters are not substantial. For example, the magnitude for the outpatient care incidence is 4.6pp, which is about 10.0% of a standard deviation (46%) of the outpatient care incidence. Panels B and C of Appendix Table C3 present the RD results by gender, which has sometimes a lower statistical power, due to reduced sample sizes. Four among RD parameters for men are greater than those for women, though these parameters for men and women are not significantly different from each other (Columns (1), (2), (3), and (7))

[62] After I limit individuals to those who report the use of inpatient care, I earn the RD estimate (0.849) for hospital or clinic visits and greater than otherwise the original RD estimate (0.292). This estimate is statistically significant at the 5% level.

4.4.2.1 Mechanisms behind the Effects on Healthcare Utilization

Regarding the effect on individual outcomes of healthcare utilization, my results indicate that crossing pension eligibility ages would lead individuals to increase their use of inpatient care services and reduce their use of outpatient care services. These results also provide evidence that higher pension income raises demand for inpatient care services, whereas it decreases demand for outpatient care services.[63] The natural question is why does the effect of pension eligibility both on inpatient and outpatient service use show heterogeneity in its direction. One of the possible explanations is that pension eligibility leads to substitution away from outpatient care toward inpatient care. However, my RD estimates do not clearly speak to whether individuals substitute inpatient care for outpatient care.

To address this possibility, I group individuals into one of four categories: (1) no inpatient and outpatient care uses; (2) only inpatient care use; (3) only outpatient care use; or (4) both inpatient and outpatient uses. Using this category variable as the outcome in Equation (1), I perform a multinomial logit to confirm the substitution.

Table 4.7 shows estimated coefficients for the effect of pension eligibility on the predicted probabilities of being in each category. "No inpatient and outpatient uses" is the base category. The results provide suggestive evidence that the increase in the use of inpatient care due to pensionable ages-crossing is resulted from individuals substituting away from outpatient care. Estimated coefficients indicate that crossing pensionable ages increases the probability of choosing inpatient care and no outpatient care by 18.1pp, whereas it decreases the probability of choosing outpatient care and no inpatient care by 19.3pp (Column (1)). Columns (2) and (3) show that the substitution effect for both men and women is similar in the direction to this result. Overall, this result suggests that crossing pensionable ages leads some individuals to substitute inpatient care for outpatient

[63] As I noted, the estimated RD parameter indicates that moving across the age cutoff increases pension income by 131.49 (10,000 KRW) (Column (4) of Table 4.5). This amount is greater than the average out-of-pocket expenditure (74.65) for inpatient care of those who used inpatient care services. This amount is much larger than a discontinuously increased out-of-pocket expenditure (2.657) for inpatient care at the pensionable age (Column (3) of Appendix Table F13). These comparisons suggest that the increased spending for inpatient care at the pensionable age may be affordable with the increased pension income. This evidence might be consistent with a theoretical prediction that a higher income enables individuals to increase their healthcare spending (92).

care.

Table 4.7: Effect of Pension Eligibility on Type of Healthcare Use

	(1) All	(2) Men	(3) Women
Inpatient Care Use Only	0.180	0.214	0.125
	(0.169)	(0.241)	(0.237)
Outpatient Care Use Only	-0.193***	-0.200**	-0.189**
	(0.066)	(0.096)	(0.091)
Both Inpatient Care and Outpatient Care	0.056	0.159	-0.004
	(0.125)	(0.207)	(0.159)
Observations	36,237	15,912	20,325

Notes: Each column represents results from a multinomial logit regression. All regressions include independent variables in Equation (1). Omitted category is "No Healthcare Use". Standard errors, shown in parentheses, are clustered at the individual level.
*p<0.10, ** p<0.05, *** p<0.01.

Next, the changed price of healthcare at the pensionable age might affect the changes in healthcare utilization. A lower cost sharing among elderly populations could increase healthcare utilization (83). However, there is no difference in cost-sharing rates for healthcare between those who are over pensionable ages and those who are under in South Korea; thus, it is unlikely that the changes in healthcare utilization at the pensionable age might be driven by the changed price of healthcare through the cost-sharing system of the national health insurance (NHI).

Results in Appendix Table F13 indirectly support this by showing that out-of-pocket spending for inpatient care or dental care (Columns (3) and (9)) increases while the use of these cares increases (Columns (1) and (7)). Also, out-of-pocket spending for outpatient care (Column (6)) increases while the use of this care decreases (Column (4)). These results suggest little evidence that individuals who turn to the pensionable age receive the lower (or higher) cost-sharing rates for the healthcare use and hence decrease (or increase) out-of-pocket payment for healthcare.

4.4.3 Threats to Identification

A potential concern is that my treatment (age cutoff-crossing) is confounded by other treatments that change discontinuously across the age cutoff. To best, I am not aware of such treatments in the South Korean context. Most likely treatments near the age cutoff are the generosity of outpatient care or basic old-age pension benefits. Specifically, those over 65 are able to pay for outpatient care in a doctor's office at a flat rate. Those who turn 65 years and are under an amount of base income for eligibility are able to receive a basic old-age pension that gives them the flat amount of pension benefits. Due to a lack of the data availability, I cannot directly test for whether the likelihoods of receiving benefits for outpatient service use or basic old-age pension changes significantly at the age cutoff. However, I do emphasize on the fact that my age cutoffs do not include 65, suggesting that changes in such generosities do not happen to take place exactly at the age cutoff.

Despite this consideration, I perform the exercise to compare my main RD estimates for healthcare utilization (Table 4.5) to a distribution of placebo RD estimates at the falsified age cutoffs from [-11,-1] and [1,11] in the running variable in an increment of one. All placebo estimates are obtained by the donut RDD approach dropping observations with a bandwidth of four bins around the age cutoff. If my estimates are located in the tail of the distribution of placebo estimates, my results seem to be systematic. Conversely, if my estimated discontinuity in healthcare utilization at the age cutoff is occurred by chance, then these estimates might be located in the middle of the distribution of placebo estimates.

Results are shown in Appendix Figure E9. In each figure, I display my main estimates from Table 4.5 using a red dotted line to compare the density of placebo estimates to my main estimates. For each outcome, I also report the fraction of times the placebo estimates are smaller than my RD estimate at the bottom of each figure. Placebo estimates for the inpatient care index for all individuals are only larger than my original estimate (0.078) by 5% times (Panel A). This result suggests that my estimated effect on inpatient care utilization is located at the tail of the distribution of placebo effects; thus, the effect is likely systematic. However, the statistically insignificantly estimated effect on the outpatient care index is located at the middle of the placebo effect distribution. Also, the fragile effect on the dental care index with narrower bandwidths is contained in the center of the distribution of placebo estimates. These results also support that my results are not abnormal but systematic. Additional falsification exercises for men (Panels D, E, and F) or women (Panels G, H, and I) show the similar results.

In addition, one may be concerned that the impact of crossing pensionable ages on the pattern of healthcare utilization mostly comes from the correlation between age cutoff-crossing and individuals' health status.[64] If doing so, then my results just reflect the impact on health status rather than healthcare utilization. To check this possibility, I first estimate Equation (4.1) while using outcome variables as nine health outcomes. Results from this task show that most of the point estimates for all individuals, men, or women are economically small or statistically insignificant (Appendix Table F14). This finding is consistent with the view that health is a stock variable that is not likely to change immediately upon the treatment.[65] Then I re-estimate Equation (4.1) by including those health outcomes as additional controls. Appendix Table F15 shows that my RD parameters are similar to my baseline estimates. These findings suggest that the effect of crossing pensionable ages on healthcare utilization is not significantly mediated by the change in individuals' health status.

Finally, I perform the bounding analysis to assess the degree to which the selection into pension receipt could affect my results. The intuition behind this analysis is that individuals who expect a higher (lower) pension are more (less) likely to apply to receive the pension. Similar to (95), I trim my sample to obtain lower and upper bound estimates.[66] Appendix Table F16 shows my original

[64]Individuals' health and healthcare spending can be mutually correlated: health status affects healthcare spending; healthcare spending may be an input into health outcomes (85). Improved health at the pensionable age may reduce health care utilization. If the improved health is the result of increased healthcare spending, then healthcare utilization could increase.

[65]Given that there is a discontinuous jump in the amount of pension benefits at the age cutoff (Panel B of Figure 4.2), this result is consistent with the evidence that the causal link between income or wealth and health is weak or modest ((93); (94)).

[66]To find a lower bound for the estimated impact of pensionable ages on healthcare utilization, I assume that only individuals who have the highest amount of pension select into pension receipt. By dropping observations with the highest pension received in each bin above the age cutoff, I obtain lower bound estimates. Similarly, I drop observations with the lowest pension received in each bin above the age cutoff to obtain upper bound estimates.

estimates (Panel A) still exist in the range from the lower bound (Panel B) to the upper bound (Panel C). Hence, this result suggests that the selection into pension receipt does not significantly affect my results.

4.5 The Relationship between Retirement and Healthcare Utilization

My empirical findings above suggest that crossing pensionable ages has a positive effect on the probability of being in retired and the utilization for inpatient care services, but no significant effect on the utilization for outpatient care services. While the effect on dental care utilization is positive and precisely estimated with the largest bandwidth, this estimated positive effect is not robust with the narrower bandwidths. In this section, I investigate the direct relationship between retirement and healthcare utilization by using LATE and ITT parameters.

4.5.1 The Treatment Effect of Retirement: LATE Parameters

I estimate the treatment effect of retirement on healthcare utilization at the pensionable age. Retirement decisions are endogenous; hence, correlations between retirement and health-related outcomes do not indicate a causal effect of retirement (85). To obtain the causal effect, I use a fuzzy regression discontinuity design (RDD) that helps address the endogeneity of the retirement effect by exploiting the pensionable age as an exogenous instrument for retirement status. This instrument should not be directly correlated with healthcare utilization except through its impact on retirement. The specification I estimate is:

$$Y_{it} = \beta_0 + \beta_1 \widehat{retirement} + \beta_2(Age_{it} - C_{bt}) + \beta_3 1[Age_{it} \geq C_{bt}] \times (Age_{it} - C_{bt}) + \eta_{it} \tag{4.2}$$

where $\widehat{retirement}$ is the predicted value by estimating the first-stage regression of the retirement dummy on $1[Age_{it} \geq C_{bt}]$, $Age_{it} - C_{bt}$, and $1[Age_{it} \geq C_{bt}] \times Age_{it} - C_{bt}$ that I defined previously. I still cluster the standard errors at the individual level in this specification.

β_1 represents the local average treatment effect (LATE) of retirement on healthcare utilization outcomes, which is nearly equivalent to the effect by scaling my ITT parameters for healthcare utilization outcomes by the first-stage parameters. Here, the effect of pensionable ages on retirement is used as the associated "first-stage" change in retirement.

Table 4.8 summarizes the LATE parameters for retirement effects on healthcare utilization. All parameters are the same in the sign as ITT parameters as shown in Table 4.5: retirement increases both inpatient and dental care utilization, but decreases outpatient care utilization. However, the sizes of these parameters are substantial. For example, the LATE parameters for inpatient care index of all individuals suggest that retirement increases inpatient care index by 1.481 standard deviations (Panel A, Column (1)). This parameter is statistically significant at the 5% level. In addition, the LATE parameter for outpatient care index of these individuals shows that the discontinuous change in retirement results in 0.671 standard deviations decreases in outpatient care index (Panel A, Column (2)). The corresponding estimate for this effect is still statistically insignificant. The estimate for the dental care index indicates that retirement raises dental care index of all individuals by 1.497 standard deviations (Panel A, Column (3)), although it is not statistically significant at the 1% level. Panel C shows that LATE parameters for women are extremely substantial. However, crossing pensionable ages does not generate a relevant first-stage effect on the probability of being in women's retirement (Column (3) of Table 4.4), making it impossible to make a strong conclusion about the connection between retirement and healthcare utilization of women.

I also test the hypotheses of weak instrument by comparing the value of the Kleibergen-Paap F-statistic with the critical values proposed by (96). I report this statistic at the bottom of each panel in Table 4.8. My statistic for all individuals (Panel A) or men (Panel B) is greater than 10 and exceeds the critical value (16.38) for a 10% maximum IV size, whereas women have the statistic far below this critical value. Hence, my instrument, $1[Age_{it} \geq C_{bt}]$, in the first-stage is not weak for all individuals or men.

While my LATE parameters show a causal and sizable relationship between retirement and healthcare utilization, I note that these parameters are earned by scaling the ITT parameters for healthcare utilization by the small first-stage parameters (0.051$\sim$0.098) for retirement. In this regard, there is the possibility that the large correlation between retirement and healthcare utilization through LATE parameters might be mechanically driven by small first-stage effects.

Table 4.8: RD Estimates on Healthcare Utilization (LATE)

	(1) Inpatient Care Index	(2) Outpatient Care Index	(3) Dental Care Index
A. Without Covariates, All			
Retirement	1.481**	-0.671	1.497**
	(0.581)	(0.534)	(0.630)
Observations	36,092	35,984	36,169
1st Stage Kleibergen-Paap F-Stat	20.00	19.52	19.35
B. Without Covariates, Men			
Retirement	0.913**	-0.512	0.622
	(0.454)	(0.375)	(0.455)
Observations	15,890	15,801	15,878
1st Stage Kleibergen-Paap F-Stat	35.67	37.53	35.97
C. Without Covariates, Women			
Retirement	5.612	-1.995	9.049
	(7.985)	(5.698)	(15.25)
Observations	20,258	20,183	20,291
1st Stage Kleibergen-Paap F-Stat	0.546	0.303	0.364

Notes: Each cell represents results from a local linear regression. The dependent variables in Columns (1)–(3) are the indices for inpatient, outpatient, or dental care. The process to construct these indexes is outlined by (4). The value of Kleibergen-Paap F-statistic for weak instrument at the bottom of each panel is reported. Standard errors, shown in parentheses, are clustered at the individual level.
*p<0.10, ** p<0.05, *** p<0.01.

4.5.1.1 Putting the Treatment Effect in Context

Taken together, my LATE parameters indicate the positive relationship between retirement and inpatient care utilization, but the negative relationship between retirement and outpatient care utilization. However, interpreting my LATE parameters needs caution. The empirical specification I used is the donut RD approach, which drops some observations near the cutoff age. Hence, I cannot restrict my analysis to all the subpopulations that are in retirement upon claiming the public pension at pensionable ages. Relying on the donut RD approach naturally brings the limitation to interpret the LATE parameters as the effects on all compliers who are in retirement at pensionable ages.

Despite this point, my LATE parameters represent the causal effects of retirement on healthcare utilization and to compare these effects with existing findings in the literature. While the parameter for the outpatient care index is not precisely estimated, my findings share empirical evidence from both developed and developing countries: previous studies find the negative effects for developed countries but the positive effects for developing countries or some high-income countries ((82); (83)). Given that my LATE parameters by using healthcare utilization indices is novel in the literature; however, I compute the LATE parameters for individual outcomes of healthcare utilization to put the effect sizes in context. Appendix Table F17 shows these LATE parameters. Among these parameters, I focus on inpatient incidence, the number of hospital stays, outpatient incidence, and the number of hospital visits that are mostly observed in the existing studies. Regarding these outcomes, I find that retirement results in 30.4pp increases in the inpatient incidence and 0.675 times increases in the number of hospital stays, whereas it leads to 89.3pp decreases in the outpatient incidence and 5.639 times decreases in the number of hospital visits.

Next, I move these four parameters in Appendix Table F18 to compare corresponding LATE parameters in previous empirical studies. In this table, I show the LATE parameters along with the standard deviation of the corresponding outcomes in the parentheses. While it is somewhat difficult to compare with the effect sizes varying according to identification strategy (Column (1)), the study population (Column (2)), and institutional factor (Column (3)), the absolute size of my parameters is larger than those of most previous studies. Except for (82), most of the studies show no significant effect on inpatient care utilization (Columns (4)–(5)). (82) find 16.3pp increases in inpatient service uses and 0.409 increases in the number of times individuals received inpatient care, which represents 48.8% and 65.8% of a standard deviation, respectively. My LATE parameters account for 95.6% and 173.5% of a standard deviation. These numbers are much greater than those of (82).

On the other hand, I observe sizable effects on outpatient care utilization in other studies. Specifically, (80) finds that retirement increases the probability of outpatient service uses by 36.1pp for men, which accounts for 74.7% of a standard deviation. My parameter for this outcome is

-0.469 and represents 99.2% of a standard deviation. (82), (5), and (83) also document that retirement leads to 0.816 increases for all individuals, 0.704 decreases for men, and 6.857 increases for men (Column (7)). These numbers represent 50.6%, 16.6%, and 37.3% changes off of a standard deviation, respectively. Regarding this outcome, the absolute values (5.639 and 1.264) of my parameter for all individuals account for 57.0% of a standard deviation (9.890). For men, the absolute value of my parameter is 1.264, which is 16.4% of a standard deviation (7.726). Hence, the effect size of my estimates for the number of times individuals (or men) received outpatient care services is similar to or smaller than that of other studies.

4.5.2 The Role of Retirement: ITT parameters

By considering the positive effect of pension eligibility on retirement, my ITT parameters for healthcare utilization can be interpreted as the intent-to-treat (ITT) effects of public pension eligibility for those on the margin of being in retirement upon turning pension eligibility ages. I further discuss about the relationship between retirement and healthcare utilization by emphasizing the ITT parameters, rather than the LATE parameters.

Following the approach of (86), I examine the linear relationship between retirement and healthcare utilization by using the pair of the ITT parameters obtained from a separate RD approach for subgroups based on covariates described in Subsection 4.3.1. Using these covariates, I create 32 distinct subgroups for all individuals.[67] For these groups, I obtain the ITT parameters for retirement and healthcare utilization by estimating the local linear RD specification (Equation (1)) and compare the ITT parameters for retirement with the ITT parameters for healthcare utilization.[68] The finding of qualitatively similar effects using this approach strengthens my causal interpretation of the retirement effects from the LATE parameters.

In Figure 4.5, I visually examine the retirement RD estimate versus the healthcare utilization RD estimate for inpatient, outpatient, and dental care services by relying on the bin–scattered

[67]Specifically, the combination of these covariates provides a possible 32 subgroups for all individuals: 2 gender $\times$ 2 spouse status $\times$ 2 offspring status $\times$ 2 education $\times$ 2 residential status.

[68]This approach has the advantage that it uses all of the across-subgroup variation in both retirement and health care utilization. However, there is the downside that the RD parameters for each group lack statistical power due to the reduced sample size.

plots. For each of the subgroups, the retirement RD coefficient is on the x-axis. The estimated RD coefficient for three healthcare utilization-related outcomes is on the y-axis. The linear relationship between the coefficients is plotted by the red line.

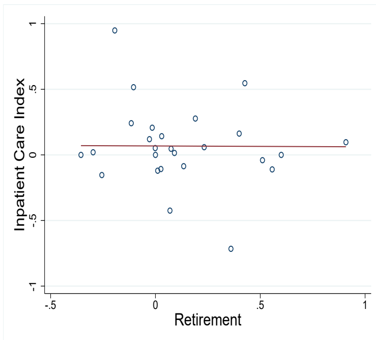
For all individuals, the relationship between retirement and health care utilization at pensionable ages appears weak visually. In Panels A–C, changes in retirement are negatively correlated with inpatient and dental care service utilization changes but positively correlated with outpatient care service utilization changes.

To evaluate the strength of these relationships, I separately regress healthcare utilization estimates on retirement estimates. Results are shown in Panel A of Table 4.9. The estimates for the inpatient care index and the outpatient care index represent economically minimal effects. These estimates are also statistically insignificant at conventional levels. The estimate for the dental care index is much more sizable than those for other two indices and statistically significant. Hence, I find no evidence that retirement plays a significant role in the changes in inpatient care or outpatient care utilization across covariates-based subgroups. While these findings from this approach are descriptive and not causal in nature, they suggest that retirement might not be a strong driver of the changes in those utilization outcomes near the pensionable age.

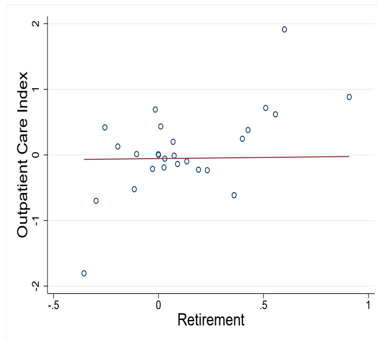
To further check if the effects in large subgroups drive the correlation, I draw another set of bin–scattered plots with weighting by the average number of observations in each subgroup to reflect their contribution to the relationship. Results are shown in Panels D–F of Figure 4.5. Interestingly, the relationship in Panels D and E indicates that retirement has a positive effect on the inpatient care index but a negative effect on the outpatient care index, which is consistent with the LATE parameters in terms of their sign. While it still appears to be visually weak, I cannot rule out the possibility that the causal relationship from the LATE parameters might be driven by the relationship across a handful of subgroups. If this is so, then investigating the relationship arising from or between specific populations, rather than understanding the "one-size-fits-all" effects from the entire population, may be useful to develop appropriate policies (83).[69]

[69]The previous studies also support this implication by showing that the retirement effect on healthcare utilization is more likely to be pronounced for people with low education (82) or blue-collar workers (5).

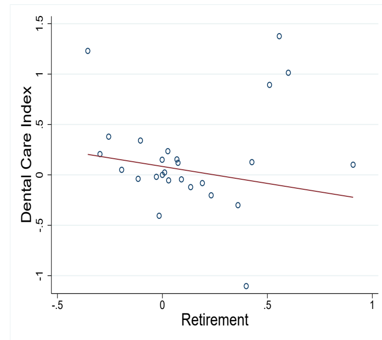
Figure 4.5: Relationship between Retirement and Healthcare Utilization



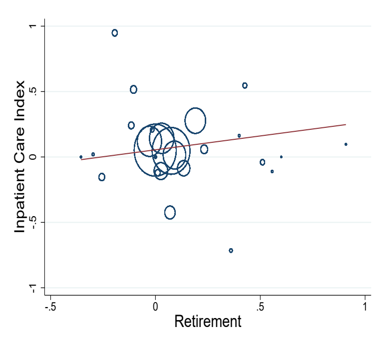
(a) Panel A: Inpatient Care (No Weight)



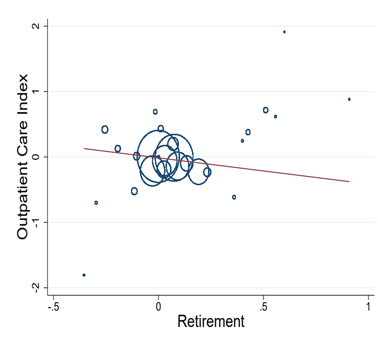
(b) Panel B: Outpatient Care (No Weight)



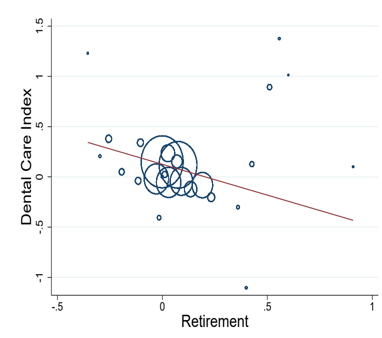
(c) Panel C: Dental Care (No Weight)



(d) Panel D: Inpatient Care (Weight)



(e) Panel E: Outpatient Care (Weight)



(f) Panel F: Dental Care (Weight)

Notes: This figure displays the relationship between retirement and healthcare utilization indices for all individuals. Each panel shows the healthcare utilization indices coefficients plotted for 32 distinct covariate subgroups on the y-axis and the estimated changes in retirement for the same 32 subgroups plotted on the x-axis. These coefficients are generated by the local linear RD regression with a donut specification. The combination of all covariates provides a possible 32 subgroups for all individuals: 2 gender $\times$ 2 spouse status $\times$ 2 offspring status $\times$ 2 education $\times$ 2 residential status. The linear relationship between these coefficients is plotted by the red line. In panels D,E, and F, bin size is weighed by the number of observations in each subgroup.

4.6 Conclusion

In this paper, I examine the causal effect of pension eligibility both on retirement and on healthcare utilization by using a regression discontinuity design (RDD). I find a positive but small effect of crossing pensionable ages on the probability of being in retirement. In addition, I find the positive effect on inpatient and dental care utilization and no significant effect on outpatient care utilization at the pensionable age. By applying a fuzzy RDD, the reported LATE estimate shows the positive effect of retirement on inpatient care utilization, whereas it shows the negative and statistically insignificant effect on outpatient care utilization. The former is consistent with the existing finding from developing countries or some high-income countries. However, an additional analysis for discussing the role of retirement suggests that retirement might not be a strong or credible driver for the changes in inpatient care utilization across all covariates-based subgroups.

In addition, I complement the growing literature documenting the relationship between retirement and healthcare utilization by taking advantage of a discontinuity at age-based eligibility thresholds for pension claiming in South Korea. My findings may provide useful evidence for policy makers in any country with a rapidly ageing population, no difference in pensionable ages across gender, public health insurance, difference in cost-sharing rates between inpatient and outpatient care, and unrestricted access to physicians. Understanding the causal nature of the relationship between retirement and healthcare utilization for the elderly is important for any public policy related to healthcare system. However, the relationship implied by the combination of my ITT estimates for each covariates-based subgroup, at least in this context, suggests that the retirement-related reform could not necessarily significantly promote or impede healthcare utilization for all older adults near the pensionable age. More in-depth investigation of subpopulation-specific responses could help policy makers design effective policies.

5. SUMMARY AND CONCLUSIONS

This dissertation applies causal inference methods to study how school peers affect students' educational and non-educational outcomes or how pension eligibility has an effect on retirement and healthcare utilization.

In the first paper, I use idiosyncratic variation in the test score distribution across classes to examine persistent effects of class rank in 7^{th} grade with the South Korean longitudinal data. I find a consistently positive impact of students' class rankings in 7^{th} grade on their test scores from 8^{th} to 12^{th} grade. I also find that math rank in 7^{th} grade positively impacts the 11^{th} grade reported likelihood of choosing the math–science track, taking higher-level math courses, and being interested in obtaining a STEM degree in college. The result of postsecondary outcomes indicates a positive effect of class rank in 7^{th} grade on the probability of attending a university. The results also show that math rank in 7^{th} grade has a positive impact on the likelihood of applying to a STEM major and majoring in a STEM field while attending a university. This paper provides strong evidence that a higher class rank plays a important role in obtaining future academic gain or choosing the later career path of students.

In addition, I estimate the impact of misbehaving male peers on the intensity of a student's own misbehavior in the South Korean middle school classrooms. Using a random assignment setting and instrumental strategy, I overcome the problems of self-section and reflection. My finding indicates that the increase of one single misbehaving male peer among 30 students in the classroom leads to the increase of other students' MCI by around 0.03 standard deviations. These effects are more noticeable for male students. This evidence suggests that educational policy intervention to address student misbehavior might be effective if it mainly targets misbehaving boys.

In the third paper, I employ a regression discontinuity design to investigate the effect of crossing pensionable ages on retirement and healthcare utilization in South Korea. I find a positive but small effect of pensionable ages-crossing on the probability of being in retirement. My results show the positive effect on inpatient and dental care utilization and no significant effect on outpatient care utilization at the pensionable age. To obtain the local average treatment effect (LATE) of retirement on healthcare utilization, I use a fuzzy regression discontinuity design. The LATE parameter shows the positive effect of retirement on inpatient care utilization, whereas it shows the negative and statistically insignificant effect on outpatient care utilization. However, an ITT-based analysis suggests that retirement might not be a strong or credible driver for the changes in inpatient care utilization across all covariates-based subgroups. This finding indicates that the retirement-related reform could not necessarily significantly promote or impede healthcare utilization for all older adults near the pensionable age.

REFERENCES

[1] S. E. Carrell, M. Hoekstra, and E. Kuka, "The long-run effects of disruptive peers," *American Economic Review*, vol. 108, no. 11, pp. 3377–3415, 2018.

[2] R. Murphy and F. Weinhardt, "Top of the class: The importance of ordinal rank," *The Review of Economic Studies*, vol. 87, no. 6, pp. 2777–2826, 2020.

[3] R. Beaman, K. Wheldall, and C. Kemp, "Differential teacher attention to boys and girls in the classroom," *Educational review*, vol. 58, no. 3, pp. 339–366, 2006.

[4] M. L. Anderson, "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects," *Journal of the American statistical Association*, vol. 103, no. 484, pp. 1481–1495, 2008.

[5] W. Frimmel and G. J. Pruckner, "Retirement and healthcare utilization," *Journal of Public Economics*, vol. 184, p. 104146, 2020.

[6] C. M. Hoxby and G. Weingarth, "Taking race out of the equation: School reassignment and the structure of peer effects," tech. rep., Working paper, 2005.

[7] V. Lavy, M. D. Paserman, and A. Schlosser, "Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom," *The Economic Journal*, vol. 122, no. 559, pp. 208–237, 2012.

[8] C. Hoxby, "Peer effects in the classroom: Learning from gender and race variation," tech. rep., National Bureau of Economic Research, 2000.

[9] V. Lavy and A. Schlosser, "Mechanisms and impacts of gender peer effects at school," *American Economic Journal: Applied Economics*, vol. 3, no. 2, pp. 1–33, 2011.

[10] S. Lee, L. J. Turner, S. Woo, and K. Kim, "All or nothing? the impact of school and classroom gender composition on effort and academic achievement," tech. rep., National Bureau of Economic Research, 2014.

[11] D. N. Figlio, "Boys named sue: Disruptive children and their peers," *Education finance and policy*, vol. 2, no. 4, pp. 376–394, 2007.

[12] S. E. Carrell and M. L. Hoekstra, "Externalities in the classroom: How children exposed to domestic violence affect everyone's kids," *American Economic Journal: Applied Economics*, vol. 2, no. 1, pp. 211–28, 2010.

[13] A. Abdulkadiroğlu, J. Angrist, and P. Pathak, "The elite illusion: Achievement effects at boston and new york exam schools," *Econometrica*, vol. 82, no. 1, pp. 137–196, 2014.

[14] W. Dobbie and R. G. Fryer Jr, "The impact of attending a school with high-achieving peers: Evidence from the new york city exam schools," *American Economic Journal: Applied Economics*, vol. 6, no. 3, pp. 58–75, 2014.

[15] T. Dee and X. Lan, "The achievement and course-taking effects of magnet schools: Regression-discontinuity evidence from urban china," *Economics of Education Review*, vol. 47, pp. 128–142, 2015.

[16] M. Hoekstra, P. Mouganie, and Y. Wang, "Peer quality and the academic benefits to attending better schools," *Journal of Labor Economics*, vol. 36, no. 4, pp. 841–884, 2018.

[17] W. Ding and S. F. Lehrer, "Do peers affect student achievement in china's secondary schools?," *The Review of Economics and Statistics*, vol. 89, no. 2, pp. 300–312, 2007.

[18] C. K. Jackson, "Student demographics, teacher sorting, and teacher quality: Evidence from the end of school desegregation," *Journal of Labor Economics*, vol. 27, no. 2, pp. 213–256, 2009.

[19] D. Berkowitz and M. Hoekstra, "Does high school quality matter? evidence from admissions data," *Economics of Education review*, vol. 30, no. 2, pp. 280–288, 2011.

[20] C. Pop-Eleches and M. Urquiola, "Going to a better school: Effects and behavioral responses," *American Economic Review*, vol. 103, no. 4, pp. 1289–1324, 2013.

[21] J. T. Denning, R. Murphy, and F. Weinhardt, "Class rank and long-run outcomes," tech. rep., National Bureau of Economic Research, 2020.

[22] B. A. Jacob, L. Lefgren, and D. P. Sims, "The persistence of teacher-induced learning," *Journal of Human resources*, vol. 45, no. 4, pp. 915–943, 2010.

[23] B. Elsner and I. E. Isphording, "A big fish in a small pond: Ability rank and human capital

investment," *Journal of Labor Economics*, vol. 35, no. 3, pp. 787–828, 2017.

[24] B. Elsner and I. E. Isphording, "Rank, sex, drugs, and crime," *Journal of Human Resources*, vol. 53, no. 2, pp. 356–381, 2018.

[25] J. M. Delaney and P. J. Devereux, "High school rank in math and english and the gender gap in stem," *Labour Economics*, p. 101969, 2021.

[26] H. Yu, "Am I the big fish? The effect of ordinal rank on student academic performance in middle school," *Journal of Economic Behavior & Organization*, vol. 176, pp. 18–41, 2020.

[27] L. Pagani, S. Comi, and F. Origo, "The effect of school rank on personality traits," *Journal of Human Resources*, vol. 56, no. 4, pp. 1187–1225, 2021.

[28] D. Card, A. Mas, E. Moretti, and E. Saez, "Inequality at work: The effect of peer salaries on job satisfaction," *American Economic Review*, vol. 102, no. 6, pp. 2981–3003, 2012.

[29] R. Chetty, N. Hendren, P. Kline, and E. Saez, "Where is the land of opportunity? the geography of intergenerational mobility in the united states," *The Quarterly Journal of Economics*, vol. 129, no. 4, pp. 1553–1623, 2014.

[30] R. Chetty and N. Hendren, "The impacts of neighborhoods on intergenerational mobility ii: County-level estimates," *The Quarterly Journal of Economics*, vol. 133, no. 3, pp. 1163–1228, 2018.

[31] C. Kang, "Classroom peer effects and academic achievement: Quasi-randomization evidence from south korea," *Journal of Urban Economics*, vol. 61, no. 3, pp. 458–495, 2007.

[32] J. Lim and J. Meer, "Persistent effects of teacher–student gender matches," *Journal of Human Resources*, vol. 55, no. 3, pp. 809–835, 2020.

[33] P. E. Todd and K. I. Wolpin, "On the specification and estimation of the production function for cognitive achievement," *The Economic Journal*, vol. 113, no. 485, pp. F3–F33, 2003.

[34] S. Cunningham, "Causal inference," *The Mixtape*, vol. 1, 2020.

[35] J. Guryan, K. Kroft, and M. J. Notowidigdo, "Peer effects in the workplace: Evidence from random groupings in professional golf tournaments," *American Economic Journal: Applied Economics*, vol. 1, no. 4, pp. 34–68, 2009.

[36] D. J. Deming and K. L. Noray, "Stem careers and the changing skill requirements of work," tech. rep., National Bureau of Economic Research, 2018.

[37] P. H. Westfall and S. S. Young, *Resampling-based multiple testing: Examples and methods for p-value adjustment*, vol. 279. John Wiley & Sons, 1993.

[38] J. Neyman and E. L. Scott, "Consistent estimates based on partially consistent observations," *Econometrica: Journal of the Econometric Society*, pp. 1–32, 1948.

[39] T. Lancaster, "The incidental parameter problem since 1948," *Journal of econometrics*, vol. 95, no. 2, pp. 391–413, 2000.

[40] E. Pahlke, J. S. Hyde, and J. E. Mertz, "The effects of single-sex compared with coeducational schooling on mathematics and science achievement: Data from korea.," *Journal of Educational Psychology*, vol. 105, no. 2, p. 444, 2013.

[41] H. Park, J. R. Behrman, and J. Choi, "Causal effects of single-sex schools on college entrance exams and college attendance: Random assignment in seoul high schools," *Demography*, vol. 50, no. 2, pp. 447–469, 2013.

[42] C. Dustmann, H. Ku, *et al.*, "Why are single-sex schools successful?," *Labour Economics*, vol. 54, pp. 79–99, 2018.

[43] G. B. Dahl and L. Lochner, "The impact of family income on child achievement: Evidence from the earned income tax credit," *American Economic Review*, vol. 102, no. 5, pp. 1927–56, 2012.

[44] K. Milligan and M. Stabile, "Do child tax benefits affect the well-being of children? evidence from canadian child benefit expansions," *American Economic Journal: Economic Policy*, vol. 3, no. 3, pp. 175–205, 2011.

[45] J. Rothstein and N. Wozny, "Permanent income and the black-white test score gap," *Journal of Human Resources*, vol. 48, no. 3, pp. 510–544, 2013.

[46] J. Gong, Y. Lu, and H. Song, "The effect of teacher gender on students' academic and noncognitive outcomes," *Journal of Labor Economics*, vol. 36, no. 3, pp. 743–778, 2018.

[47] J. Lim and J. Meer, "The impact of teacher–student gender matches random assignment

evidence from south korea," *Journal of Human Resources*, vol. 52, no. 4, pp. 979–997, 2017.

[48] J. Kinsler and R. Pavan, "Local distortions in parental beliefs over child skill," *Journal of Political Economy*, vol. 129, no. 1, pp. 81–100, 2021.

[49] R. Dizon-Ross, "Parents' beliefs about their children's academic ability: Implications for educational investments," *American Economic Review*, vol. 109, no. 8, pp. 2728–65, 2019.

[50] J. D. Angrist and K. Lang, "Does school integration generate peer effects? evidence from boston's metco program," *American Economic Review*, vol. 94, no. 5, pp. 1613–1634, 2004.

[51] J. B. Cullen, B. A. Jacob, and S. Levitt, "The effect of school choice on participants: Evidence from randomized lotteries," *Econometrica*, vol. 74, no. 5, pp. 1191–1230, 2006.

[52] S. E. Black, P. J. Devereux, and K. G. Salvanes, "Under pressure? the effect of peers on outcomes of young adults," *Journal of Labor Economics*, vol. 31, no. 1, pp. 119–153, 2013.

[53] D. Whitmore, "Resource and peer impacts on girls' academic achievement: Evidence from a randomized experiment," *American Economic Review*, vol. 95, no. 2, pp. 199–203, 2005.

[54] B. A. Jacob, "Where the boys aren't: Non-cognitive skills, returns to school and the gender gap in higher education," *Economics of Education review*, vol. 21, no. 6, pp. 589–598, 2002.

[55] P. Orpinas, C. McNicholas, and L. Nahapetyan, "Gender differences in trajectories of relational aggression perpetration and victimization from middle to high school," *Aggressive Behavior*, vol. 41, no. 5, pp. 401–412, 2015.

[56] B. Sacerdote, "Peer effects with random assignment: Results for dartmouth roommates," *The Quarterly journal of economics*, vol. 116, no. 2, pp. 681–704, 2001.

[57] D. S. Lyle, "Estimating and interpreting peer and role model effects from randomly assigned social groups at west point," *The Review of Economics and Statistics*, vol. 89, no. 2, pp. 289–299, 2007.

[58] R. Chetty, J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan, "How does your kindergarten classroom affect your earnings? evidence from project star," *The Quarterly journal of economics*, vol. 126, no. 4, pp. 1593–1660, 2011.

[59] J. Lim and J. Meer, "How do peers influence bmi? evidence from randomly assigned classrooms in south korea," *Social Science & Medicine*, vol. 197, pp. 17–23, 2018.

[60] C. F. Manski, "Identification of endogenous social effects: The reflection problem," *The Review of Economic Studies*, vol. 60, no. 3, pp. 531–542, 1993.

[61] A. M. Pearl, B. F. French, J. E. Dumas, A. D. Moreland, and R. Prinz, "Bidirectional effects of parenting quality and child externalizing behavior in predominantly single parent, under-resourced african american families," *Journal of Child and Family Studies*, vol. 23, no. 2, pp. 177–188, 2014.

[62] I. Daryanani, J. L. Hamilton, L. Y. Abramson, and L. B. Alloy, "Single mother parenting and adolescent psychopathology," *Journal of abnormal child psychology*, vol. 44, no. 7, pp. 1411–1423, 2016.

[63] J. Comeau and M. H. Boyle, "Patterns of poverty exposure and children's trajectories of externalizing and internalizing behaviors," *SSM-population health*, vol. 4, pp. 86–94, 2018.

[64] J. Choi, H. Park, and J. R. Behrman, "Separating boys and girls and increasing weight? assessing the impacts of single-sex schools through random assignment in seoul," *Social science & medicine*, vol. 134, pp. 1–11, 2015.

[65] N. W. Papageorge, V. Ronda, and Y. Zheng, "The economic value of breaking bad: Misbehavior, schooling and the labor market," tech. rep., National Bureau of Economic Research, 2019.

[66] A. C. Cameron and P. K. Trivedi, *Microeconometrics: methods and applications*. Cambridge university press, 2005.

[67] M. Neidell and J. Waldfogel, "Cognitive and noncognitive peer effects in early education," *The Review of Economics and Statistics*, vol. 92, no. 3, pp. 562–576, 2010.

[68] S. E. Carrell, M. Hoekstra, and J. E. West, "Is poor fitness contagious?: Evidence from randomly assigned friends," *Journal of public Economics*, vol. 95, no. 7-8, pp. 657–663, 2011.

[69] J. H. G. Kristoffersen, M. V. Krægpøth, H. S. Nielsen, and M. Simonsen, "Disruptive school

peers and student outcomes," *Economics of Education Review*, vol. 45, pp. 1–13, 2015.

[70] I. Horoi and B. Ost, "Disruptive peers and the estimation of teacher value added," *Economics of Education Review*, vol. 49, pp. 180–192, 2015.

[71] A. Singh and U. Kiran, "Effect of single parent family on child delinquency," *Int J Sci Re*, vol. 3, no. 9, pp. 866–8, 2014.

[72] A. Alami, S. Khosravan, L. S. Moghadam, F. Pakravan, and F. Hosseni, "Adolescents' self-esteem in single and two-parent families," *International journal of community based nursing and midwifery*, vol. 2, no. 2, p. 69, 2014.

[73] J. D. Angrist, "The perils of peer effects," *Labour Economics*, vol. 30, pp. 98–108, 2014.

[74] M. Celidoni and V. Rebba, "Healthier lifestyles after retirement in europe? evidence from share," *The European Journal of Health Economics*, vol. 18, no. 7, pp. 805–830, 2017.

[75] N. Coe and G. Zamarro, "Does retirement impact health care utilization?," *CESR-Schaeffer Working Paper*, no. 2015-032, 2015.

[76] P. Eibich, "Understanding the effect of retirement on health: Mechanisms and heterogeneity," *Journal of health economics*, vol. 43, pp. 1–12, 2015.

[77] E. Fe and B. Hollingsworth, "Estimating the eect of retirement on mental health via panel discontinuity designs," 2012.

[78] J. Hagen, "The effects of increasing the normal retirement age on health care utilization and mortality," *Journal of Population Economics*, vol. 31, no. 1, pp. 193–234, 2018.

[79] N. F. Nielsen, "Sick of retirement?," *Journal of health economics*, vol. 65, pp. 133–152, 2019.

[80] T. Dang, "Retirement and health services utilization in a low-income country," *Economics of Transition and Institutional Change*, 2021.

[81] C. Lucifora and D. Vigani, "Health care utilization at retirement: The role of the opportunity cost of time," *Health economics*, vol. 27, no. 12, pp. 2030–2050, 2018.

[82] Y. Zhang, M. Salm, and A. van Soest, "The effect of retirement on healthcare utilization: Evidence from china," *Journal of health economics*, vol. 62, pp. 165–177, 2018.

[83] Q. Zhou, K. Eggleston, and G. G. Liu, "Healthcare utilization at retirement in china," *Health Economics*, 2021.

[84] A. Bíró and P. Elek, "How does retirement affect healthcare expenditures? evidence from a change in the retirement age," *Health Economics*, vol. 27, no. 5, pp. 803–818, 2018.

[85] A. Gorry, D. Gorry, and S. N. Slavov, "Does retirement improve health and life satisfaction?," *Health economics*, vol. 27, no. 12, pp. 2067–2086, 2018.

[86] M. D. Fitzpatrick and T. J. Moore, "The mortality effects of retirement: Evidence from social security eligibility at age 62," *Journal of Public Economics*, vol. 157, pp. 121–137, 2018.

[87] G. W. Imbens and T. Lemieux, "Regression discontinuity designs: A guide to practice," *Journal of econometrics*, vol. 142, no. 2, pp. 615–635, 2008.

[88] D. S. Lee and T. Lemieux, "Regression discontinuity designs in economics," *Journal of economic literature*, vol. 48, no. 2, pp. 281–355, 2010.

[89] J. Hahn, P. Todd, and W. Van der Klaauw, "Identification and estimation of treatment effects with a regression-discontinuity design," *Econometrica*, vol. 69, no. 1, pp. 201–209, 2001.

[90] A. Gelman and G. Imbens, "Why high-order polynomials should not be used in regression discontinuity designs," *Journal of Business & Economic Statistics*, vol. 37, no. 3, pp. 447–456, 2019.

[91] M. D. Cattaneo, M. Jansson, and X. Ma, "Manipulation testing based on density discontinuity," *The Stata Journal*, vol. 18, no. 1, pp. 234–261, 2018.

[92] M. Grossman, *1. On the Concept of Health Capital and the Demand for Health*. Columbia University Press, 2017.

[93] J. Meer, D. L. Miller, and H. S. Rosen, "Exploring the health–wealth nexus," *Journal of health economics*, vol. 22, no. 5, pp. 713–730, 2003.

[94] M. Lindahl, "Estimating the effect of income on health and mortality using lottery prizes as an exogenous source of variation in income," *Journal of Human resources*, vol. 40, no. 1, pp. 144–168, 2005.

[95] D. S. Lee, "Training, wages, and sample selection: Estimating sharp bounds on treatment effects," *The Review of Economic Studies*, vol. 76, no. 3, pp. 1071–1102, 2009.

[96] J. H. Stock and M. Yogo, "Testing for weak instruments in linear iv regression," 2002.

[97] J. Rothstein, "Student sorting and bias in value-added estimation: Selection on observables and unobservables," *Education finance and policy*, vol. 4, no. 4, pp. 537–571, 2009.

[98] R. Chetty, J. N. Friedman, and J. E. Rockoff, "Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates," *American Economic Review*, vol. 104, no. 9, pp. 2593–2632, 2014.

[99] R. Chetty, J. N. Friedman, and J. E. Rockoff, "Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood," *American economic review*, vol. 104, no. 9, pp. 2633–79, 2014.

[100] S. Gershenson, S. B. Holt, and N. W. Papageorge, "Who believes in me? the effect of student–teacher demographic match on teacher expectations," *Economics of education review*, vol. 52, pp. 209–224, 2016.

[101] S. Gershenson, C. Hart, J. Hyman, C. Lindsay, and N. W. Papageorge, "The long-run impacts of same-race teachers," tech. rep., National Bureau of Economic Research, 2018.

[102] A. Eble and F. Hu, "Child beliefs, societal beliefs, and teacher-student identity match," *Economics of Education Review*, vol. 77, p. 101994, 2020.

[103] M. Hoekstra and C. Sloan, "Does race matter for police use of force? evidence from 911 calls," *Working Paper*, 2020.

[104] M. Bertoni and R. Nisticò, "Ordinal rank and peer composition: Two sides of the same coin?," 2019.

[105] S. Ertac, "Social comparisons and optimal information revelation: Theory and experiments," *Job Market Paper, UCLA*, 2005.

[106] H. W. Marsh, M. Seaton, U. Trautwein, O. Lüdtke, K.-T. Hau, A. J. O'Mara, and R. G. Craven, "The big-fish–little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research," *Educational psychology review*, vol. 20, no. 3,

pp. 319–350, 2008.

APPENDIX A

MEASUREMENT ERROR

My identification strategy rests on the randomness of class rank conditional on the test score distribution in 7^{th} grade and classroom by subject fixed effects. Other factors in 7^{th} grade may cause the test score to be a poor measure of the student ability. However, one still may be concerned that rank parameter generated by the test score could pick up the underlying student ability. Accordingly, rank effects may reflect rank-preserving effects rather than the impact of rank itself. (2) also address this kind of measurement error problem. I replicate their approaches to explore how the problem that test scores do not capture the underlying student ability due to (1) systematic measurement error and (2) the general case of non-systematic measurement error (noise) would confound the benchmark results in Table 2.5.

A.1 Systematic Measurement Error

Both teacher and peer effects that affect the test score but not the rank can generate systematic measurement error in test scores. I describe why this kind of error would be concerned and how to address this issue in Appendix B.

A.2 Non-systematic Measurement Error

Students' test scores may reflect a noisy measure of ability, not a perfect representation. This type of non-systematic measurement error could generate spurious effects of student rank on outcomes, including test scores. For example, a mistakenly high (or low) test score of a student can cause a falsely high (or low) rank in a class. In this case, this student would experience lower (or higher) observed growth in test scores, which would bias the rank effects downward (or upward).

I illustrate the specific situation by using the traditional concept of measurement error in the variable. There are two variables, student i's true ability T_i^*, and class true ability ranking R_i^*. Because I cannot observe true ability (T_i^*) directly, I use a test score T_i which might be a noisy measure of true ability: T_i=$T(T_i^*, e_{i1})$ where e_{i1} is measurement error in test scores. Also, I observe the test scores of all others in a class T_{-i} to generate the rank of a student, R_i=$R(T_i, T_{-i})$. This rank measure also might be measured with measurement error e_{i2} such that R_i=$R^{'}(R_i^*, e_{i2})$. By using these assumptions and relations, I obtain the following equation:

$$R_i = R(T_i, T_{-i}) = R(T(T_i^*, e_{i1}), T(T_{-i}^*, e_{-i1})) = h(T_i^*, e_{i1}, T_{-i}^*, e_{-i1}) \tag{A.1}$$

In Equation (A.1), R_i also depends on e_{i1}, which means that measurement error (e_{i1}) in the test score (T_i) also causes noise in measuring true ability ranking (R_i^*).

To examine how this problem would impact the estimated coefficients of the rank, I perform a set of Monte Carlo simulations where I include additional measurement error randomly drawn from a normal distribution with mean zero and a standard deviation proportional to the standard deviation (one) in baseline test scores.

Appendix Figure E1 shows the simulated estimates of the mean with 1,000 times repetitions for each measurement error level. As measurement error increases, there is a non-linearly downward bias. However, a small additional measurement error has little impact on the results, so the main estimates would not be attenuated much.

APPENDIX B

TEACHER AND PEER EFFECTS

B.1 Teacher Effects

A student's outcomes may be affected by observable and unobservable teacher attributes, including gender, race, and value-added ((97); (98); (99); (100); (47); (101); (46); (102); (32)). If omitted teacher characteristics are positively directly correlated with the rank effects on academic outcomes, then the estimates on class rank would be biased upward. Because I rest on randomness of the rank effects produced by idiosyncratic variation in the test score distributions across classes in the main identification strategy, the omission of teacher characteristics would not give rise to such a bias. I have shown this by including teacher controls in Column (3) of Table 2.5 and Panel B of Table 2.6, indicating that doing so does not largely change the main estimates.

Unobservable teaching ability to have a temporary impact on test scores but not the rank that pick up the information about true ability rank can generate spurious rank effects. These teacher effects may be accounted for by two channels. The first one is that teachers could increase students' test scores regardless of their ability (teacher-level effects). The second one is that teachers could have different transformation to students' test scores according to their ability (teacher-transformation effects).

Similar to (2), let me describe those two effects in the teachers' production function:

$$y_{ijs} = \mu_{js} + \delta_{js} a_{ij} + e_{ijs} \tag{B.1}$$

where y_{ijs} is the test score of a student (i) in a given subject (j) and school (s). μ_{js} is the teacher-level effects. δ_{js} is the (linear) teacher-transformation effects. a_{ij} is the student ability. e_{ijs} is the error term.

The differences in the teacher's production, driven by the teacher-level effects would be captured by classroom by subject fixed effects and subject teacher fixed effects in my empirical setting. To illustrate this, I plug Equation (B.1) into a third-order polynomial of a student's 7^{th} grade test scores in Equation (2.3). Then, I have the following equation:

$$\begin{aligned} y^{T}_{ijs} &= \beta_0 + \beta_1 Rank^{7^{th}}_{ijcs} + f(score^{7th}) + \theta_{jc} + \epsilon_{ijs} = \beta_0 + \beta_1 Rank^{7^{th}}_{ijcs} + f(\mu_{js} + \delta_{js}a_{ij} + e_{ijs}) + \theta_{jc} + \epsilon_{ijs} \\ &= \beta_0 + \beta_1 Rank^{7^{th}}_{ijcs} + (\mu_{js} + \mu^2_{js} + \mu^3_{js}) + (\delta_{js}a_{ij} + \delta^2_{js}a^2_{ij} + \delta^3_{js}a^3_{ij}) + (e_{ijs} + e^2_{ijs} + e^3_{ijs}) + \theta_{jc} + \epsilon_{ijs} \end{aligned} \tag{B.2}$$

In Equation (B.2), the teacher-level effects $(\mu_{js} + \mu^2_{js} + \mu^3_{js})$ can be absorbed by classroom by subject fixed effects; thus, estimating Equation (2.1) itself may address the bias by the teacher-level effects. However, these effects can be also driven by subject teacher-level shocks.[70] Hence, as the sensitivity test, I estimate Equation (2.1) by adding subject teacher fixed effects. Panel C of Appendix Table F5 shows that including subject teacher fixed effects does not make a large difference in the estimated coefficients in Panel A of Table 2.6.

However, the teacher-transformation effects (δ_{js}) would not be accounted for by these fixed effects because students within a class are affected differently by teachers according to their ability (a_{ij}). To examine if this kind of teacher-induced spreading of test scores is driving my results, I follow the approach of (2) and perform a set of placebo tests with permutation exercises and specification checks.

First, I make an assumption that these teacher-transformation effects (δ_{js}) are invariant across classes. In South Korean middle schools, teachers are not assigned to students within a given class by considering their teaching skills. Therefore, a given student would have similar teacher–transformation regardless of what class they had within their own school. If this scenario is true, then remixing students across classes will have no impact on the estimated coefficients for the rank effects. To test for this, I randomly reassign students to classes within their own school while recalculating their class ranks. Then, I reestimate the effects of this artificial rank on later test scores using the most parsimonious specification like Panel A of Table 2.6, with 1,000 replications.

[70] South Korean subject teachers rotate classrooms or take up their own class for ability tracking. Hence, these fixed effects would additionally absorb teacher-level shocks that are not overlapped with classroom by subject-level shocks.

The results are shown in Appendix Figure E3. If artificially mixing students across classes impacts the rank effect, then my original estimates from 8^{th} to 12^{th} grade are located in the middle of the placebo coefficients distribution. However, my estimates seem to be located in the right tale of the placebo coefficients distribution. Hence, the teacher-transformation effects are not generating all of the rank effects.

Second, the teacher-transformation effects might be reflected in the variance of test scores within each class, as teachers with a high value of δ_{js} would create a higher variance of test scores within such classes. To test this, I include an interaction of the standard deviation of class test scores with each student's rank. Here, the effect of the standard deviation itself is subsumed by classroom by subject fixed effects. In Appendix Table F3, I find that allowing for interactions with the standard deviation of test scores does not significantly impact the main results.

Finally, I allow for the impact of test scores to additionally vary by classroom (or subject) by interacting test scores with classroom (or subject) fixed effects to test for whether teachers have a direct influence on the spread of test scores in 7^{th} grade. Appendix Table F4 shows that the point estimates do not significantly differ from the main estimates.

B.1.1 The Role of Teachers in the Differential Effects by Teacher–Student Gender Matches

In Section 2.5.4.1, I find that female teachers increase class rank for female students, compared to male students matched with female teachers. Regarding this finding, one may raise the question to ask whether the difference by teacher gender are due to only a small amount of teachers, or if they are more systematic. To tackle this question, I develop the mixed linear model to estimate a subject teacher random effects:[71]

$$R_{ijbs} = \alpha + X_i^{'}\delta_1 + T_j^{'}\delta_2 + \gamma_{bs} + \eta_j + \varepsilon_{ijbs} \tag{B.3}$$

where η_j is the subject teacher (j)-specific random effect. This random effect measures the difference between the average class rank of a student i with a subject teacher j and the average class

[71]This approach is similar to (103) who apply this statistical model to look at whether the differential effects of police officers with respect to race are driven by a handful of them.

rank (0.5) in the entire sample. Other variables are student (X_i') and teacher (T_j') controls. I begin by regressing class rank on school by subject fixed effects and other student and teacher controls, and keep the residuals. I then exploit those residuals as the outcome in a random effect model, and then compute the random effect for each subject teacher j.[72]

In Appendix Figure E4, I show the distribution of random effects for female and male teachers through kernel density plots. I trim about 1.5% extreme teachers in the far left tails for more clear visualization of any differences in the distributions. From the distribution, I am able to graphically distinguish the effects due to differences in the middle of the distribution from the effects due to differences in the tails. Panel A shows that there are fewer male teachers on the left side and the middle of distribution and more in the right side of the distribution, compared to female teachers. However, I do not find a noticeable shift in the male teacher distribution from the middle of the distribution toward the right tail. Theses findings suggest that the increased class rank by male teachers may be driven by a handful of teachers at some points in the distribution. Panel B shows that for female students, there are still fewer male teachers on the left side of the distribution, but not more in the middle of the distribution, compared to female teachers. However, there is no visual evidence of a rightward shift in the distribution of female teachers. Hence, the increased class rank of female students by female teachers may not be driven by a sizable fraction of teachers. The distribution of both female and male teachers for male students in Panel C is similar to that in Panel A. While I find the positive treatment effect of female teachers on female students' class rank, this evidence makes it difficult to conclude that this effect is systematic.

B.2 Peer Effects

Class peers in 7^{th} grade can have a permanent or transitory impact on a student's outcomes. The first one could be absorbed in the test scores in 7^{th} grade. The second one could generate false rank effects if low-achieving peers who would lower students' test scores provide them with a higher rank. This kind of the possibility that a measure of ability might be confounded by peer effects would lead the coefficients of ordinal rank to be biased upward. Because class peer effects

[72]To estimate a random effect model, I use the STATA command *mixed*.

can be considered as classroom-level shocks, classroom by subject fixed effects can capture such peer effects.

However, because a student cannot be his or her own peer, one may still be concerned that any peer effects that are individual-specific might not be absorbed by those fixed effects. To test for this, I include peer controls in the specification, such as leave-out means of student controls and test scores, and the standard deviation of peers' test scores.[73] Panel D of Appendix Table F5 shows that including peer controls does not change the benchmark estimates (in Panel B of Table 2.6) much, implying that the individual-specific peer effects would not be significantly driving the main results.

[73](104) note that failure to control for the distribution of peer quality would bias the estimates of rank effects in the presence of large variation in peer composition across reference groups. To rule out this concern, I include here the mean and standard deviation of peers' test scores in a set of peer controls.

APPENDIX C

MODELS FOR EFFORT ALLOCATION AND PARENTAL LEARNING

C.1 Model for Student Effort Allocation

In Subsection 2.5.1, I showed that the rank effect increases students' self-confidence. There are two channels to illustrate how increased confidence could improve student academic achievement by triggering his or her effort allocations. ((105); (106); (2)). First, students use their ordinal class rankings to learn about their own subject-specific abilities. They eventually exploit this information when deciding how to allocate effort by subject in the future. Second, a student's ordinal ranking has an impact on their academic confidence, which generates non-cognitive skills in a subject such as grit, resilience, and perseverance. In the educational literature, the first model is defined as the "Learning-About-Ability effect" model while the second one is denoted as the "Big-Fish-Little-Pond Effect" model (the non-cognitive skill model).

Following (2), I develop a simple conceptual framework that embodies these two models. For simplicity, I consider only two periods: middle school (t_1) and high school (t_2). I assume that in the middle school period, students compare their academic achievement to peers by observing their class rank and then change their confidence when entering the high school period. I also assume that there are only two subjects (s), maths (m) and English (e). The education production function for student i is:

$$TS_i = f(A_{im}, E_{im}) + f(A_{ie}, E_{ie}) = s_{im}A_{im}(E_{im}^{\kappa}) + s_{ie}A_{ie}(E_{ie}^{\kappa}) \tag{C.1}$$

where TS_i is the total test score for student i, which is defined as the sum of their test scores across subjects s. A_{is} is a subject-specific ability. E_{is}^{κ} is a subject-specific effort, where I assume decreasing returns to effort, 0<κ<1. s_{is} is the subject-specific school factor.

Next, I assume that a student i maximize his or her total test scores subject to a cost function.

This cost function is represented as:

$$T_i \geq C_{im}E_{im} + C_{ie}E_{ie} + C_{ig}(E_{im} + E_{ie}) \tag{C.2}$$

where T_i is the total (fixed) cost of effort for student i. C_{is} is the cost of effort for student i in each subject s. C_{ig} is the general cost of effort for student i. Here, the general cost has a linear form of the sum of effort across two subjects, $E_{im} + E_{ie}$.

After solving the problem of maximizing Equation (C.1) subject to Equation (C.2), I obtain the optimal effort level E^* for student i and each subject s as the first order condition:

$$E^*_{is} = \left[\frac{\lambda(C_{is} + C_{ig})}{\kappa s_s A_{is}}\right]^{(1/(\kappa-1))} \tag{C.3}$$

where λ>0 is the marginal test score per effort. Equation (C.3) means that the optimal choice of student effort (E^*_{is}) will be determined by either perceived ability (A_{is}) or the cost of effort ($C_{is} + C_{ig}$). Specifically, as 0<κ<1, at fixed λ and s_s any increased subject-specific ability A_{is} will increase the optimal effort for that subject s, and higher costs will decrease the optimal effort.

As (2) note, an important distinction between the learning-about-ability effect and the non-cognitive skill models is to whether students know their abilities. In the learning-about-ability effect model, students know their costs of effort but do not know their abilities, whereas in the non-cognitive skill model, students know their costs of effort and their abilities in each subject. The key assumption in the learning-about-ability effect model is that students form beliefs about abilities by using test scores and relative rank in class. In this case, students who, for example, had a high rank in math in the primary period will devote more effort to math in the secondary period because they believe that this will generate high marginal returns to effort in that subject than doing so in other subjects. Moreover, this model assumes that higher rankings in other subjects do not change students' general cost of effort C_{ig}. Therefore, there might not be the spillover effects of a higher rank in one subject on efforts and hence test scores in other subjects. In contrast, the non-cognitive skill model assumes that rank in any subject negatively impacts students' general cost of effort C_{ig} by developing positive non-cognitive skills in that subject. Hence, in this model, for example, a high rank in math can also increase test scores in English by leading students to provide more effort to all subjects.

I look at the estimates for the spillover effects of rank in one subject on test scores in other subjects in my data. Appendix Table F7 shows spillovers across subjects. English rank has a positive and statistically significant effect on some of math and Korean test scores. However, the estimates for Korean rank effects on math and English test scores are not statistically significant. In addition, the estimates for math rank effects on English and Korean test scores are not statistically insignificant. Hence, I find no strong evidence of spillover effects of rank in one subject on test scores in another subject.

C.1.1 Test for Student Effort Allocation

In this section, I indirectly test for the learning–about-ability effect. To do so, I still rest on the approach of (2) by assuming that if local ranks are very different from global (national) ranks, and thus less informative about actual ability, then the misinformation on ability can lead to students obtaining lower average test scores.[74] Similar to them, I define the degree of misinformation as the absolute difference between students' local (class) rank and their global (in-sample) rank:

$$M^g_{ijcs} = \left|Rank_{ijcs} - Rank^g_{ijcs}\right|$$

where $Rank^g_{ijcs}$ is global rank. This measure should be zero if local rank is exactly equal to global rank. Also, a large number of M^g_{ijcs} means the large difference between them. To perform the empirical task, I first average this metric across subjects within student. This provides a mean indicator of the degree of misinformation for each student. Then, I include the misinformation variable in the specification similar to Equation (2.3) while removing the subject variation by

[74] Following their approach, I also assume that if the rank effects are caused by actual changes in the costs associated with the education production function, then the large difference between local ranks and global ranks would not result in a misallocation of effort in terms of maximizing test scores. Hence, average test scores would not decrease.

averaging test scores in subject. I estimate the following equation:

$$\overline{y}_{ics}^{T} = \beta_0 + \beta_1 \overline{Rank}_{ics}^{7^{th}} + f(score^{7^{th}}) + \delta_g \overline{M}_{ijcs}^{g} + \eta_{cs} + \epsilon_{ics} \tag{C.4}$$

where $\overline{y}_{ics}^{T}$ denotes the average test scores across subjects in period T. $\overline{Rank}_{ics}^{7^{th}}$ is the average rank in 7^{th} grade. $f(score^{7^{th}})$ is the 3^{rd} polynomial in 7^{th} grade test scores in each subject. $\overline{M}_{ijcs}^{g}$ is the average degree of misinformation. η_{cs} is classroom fixed effects. ϵ_{ics} is the error term.

Students who use misinformation to form beliefs about their ability may have lower average test scores across subjects. I therefore expect $\delta_g < 0$ if the higher degree of misinformation causes students to misallocate effort across subjects (the learning-about-ability effect). I construct the null hypothesis ($\delta_g = 0$) that local rank causes changes to the actual production function.

$$H_0\text{: } \delta_g = 0$$
$$H_1\text{: } \delta_g < 0$$

Appendix Table F8 shows the results. Panel A displays the benchmark estimates without including the additional misinformation variable. All rank parameters still indicate the positive effect of rank on test scores. Specifically, the effect of average rank in 7^{th} grade on average test scores in 11^{th} grade is estimated at 1.544 standard deviations. Among these estimates, three are statistically significant. In Panel B, the misinformation parameter in 11^{th} grade is also negatively estimated and statistically significant at the 1% level. The other coefficients of misinformation are also negative, though these coefficients are imprecisely estimated; however, their magnitude is fairly sizable. In Panel C, the interaction terms of the misinformation variable with dummies for student gender are included in the specification. The learning-about-ability effect is revealed for male students in 8^{th} grade. The corresponding estimate is statistically significant at the 5% level. In addition, this effect persists until 12^{th} grade: the corresponding coefficient is negative and statistically significant at the 5% level.

C.2 Model for Parental Learning

In this section, I present a simple model for parental learning about child skill.[75] This model includes the role of local (class) rank to generate information distortions in parental beliefs about children's cognitive skills. I assume that parents are uncertain about the cognitive skill of their child and the cognitive skill of the average child (peer) in the local school. I also assume that they know the average cognitive skill level in the economy, which is normalized to zero. The cognitive skill of their children and school can be proxied by test scores. The true skill level for the average student in school j is s_j. The true skill level for a student i in school j is c_{ij}. These two variables are defined as:

$$s_j = \epsilon_j \;\; \& \;\; c_{ij} = s_j + \epsilon_{ij}$$

where ϵ_j and ϵ_{ij} are normally distributed random variables with the mean zero and nonzero standard deviations. Parents also have normally distributed prior beliefs about child and school skill with means $\hat{c}_{ij}$ and $\hat{s}_j$, given the basis of information they have already obtained.

In addition, I assume that parents additionally receive information about child and school skills through class rank ($Rank_{ijcs}$), which is associated with these skills:

$$Rank^{sig}_{ijcs} = \gamma_c c_{ij} + \gamma_s s_j + \eta_{ijcs} \tag{C.5}$$

where η_{ijcs} is a normally distributed random variable with the mean zero and nonzero standard deviations (σ_η). γ_c is between zero and one. Parents might receive information about class rank from teachers, their children, parents of children's school peers, and so on. If γ_c=-γ_s, then a signal through class rank is the signal of child skill relative to school average skill. Note that for a signal from class rank to generate the negative impact of school average skill on parental beliefs about

[75] This model is motivated by (48). I follow the definition of variables and the approach in their model. I slightly modify their model by assuming that parents receive a mixed signal of child and school skill through information on class rank, rather than direct information on them from teachers, counselors, and so on.

child skill, γ_s must be negative.[76]

Next, I assume that parents are Baysian updaters and develop the equation of posterior parental beliefs about children's cognitive skills:

$$c_{ij}^{'} = (1 - \Omega\gamma_c)\hat{c}_{ij} - \Omega\gamma_s\hat{s}_j + \Omega\gamma_c c_{ij} + \Omega\gamma_s s_j + \Omega\eta_{ijcs} \tag{C.6}$$

where $c_{ij}^{'}$ is the posterior parental belief about a child's skill. Ω is the ratio of the covariance between c_{ij} and $Rank_{ijcs}^{sig}$ divided by the variance of $Rank_{ijcs}^{sig}$; this value will be positive and less than one.[77]

Prior beliefs of parents are likely to contribute to the impact their posterior expectation about their children's skills. To show this, parent mean prior beliefs about school and child skills can be presented as:

$$\hat{s}_j = \phi_s s_j + \eta_{ij}^s \ \ \& \ \ \hat{c}_{ij} = w_s s_j + w_c c_{ij} + \eta_{ij}^c$$

where $\phi_s > 0$ and $w_s \leq 0$.[78]

I substitute these two equations for mean prior beliefs into the posterior belief in Equation (C.6) and obtain the following equation:

$$\hat{c}_{ij}^{'} = \left[\underbrace{\Omega\gamma_s}_{Rank_{ijcs}^{sig}} - \underbrace{\Omega\gamma_s\phi_s}_{\hat{s}_j} + \underbrace{(1 - \Omega\gamma_c)w_s}_{\hat{c}_{ij}} \right] s_j + \Gamma c_{ij} + \tilde{\epsilon}_{ij} \tag{C.7}$$

where $\tilde{\epsilon}_{ij}$=$\Omega\epsilon_{ij}$+(1-$\Omega\gamma_c$)η_{ij}^c-$\Omega\gamma_s\eta_{ij}^s$ and Γ=(1-$\Omega\gamma_c$)w_c+$\Omega\gamma_c$.

[76]Regarding a signal about school average skill, I assume that parents indirectly perceive it through class rank. For example, if the child's test score is high but their class rank is low, then the parent can guess that their school peers' average test scores might be high. I also regress class rank in 7^{th} grade on a student's test scores in 7^{th} grade and school average test scores in 7^{th} grade. The corresponding OLS estimates are 0.287 (s.e.=0.003) and -0.138 (s.e.=0.006). This result suggests that a higher school skill might have the negative effect on beliefs of parents who perceive class rank as child skill.

[77]To be explicit, Ω = $(\gamma_s\Sigma^{sc} + \gamma_c\Sigma^c)/(\gamma_s^2\Sigma^s + \gamma_c^2\Sigma^c + 2\gamma_s\gamma_c\Sigma^{sc} + \sigma_\eta^2)$ where associated mean squared errors of mean prior beliefs about child ($\hat{c}_{ij}$) and school($\hat{s}_j$) skill are Σ^c, Σ^s, and Σ^{sc}.

[78]The condition $w_s \leq 0$ satisfies the idea that past signals from school skill could create local distortions in parents' prior mean beliefs.

Equation (C.7) indicates that local distortions (the combined coefficient on s_j) in parental beliefs can be driven by a number of channels, including a signal of school skill through class rank (the first term), prior mean belief about school skill (the second term), and prior mean belief about child skill (the third term). Given that $\hat{c}_{ij}^{'}$ and $\hat{c}_{ij}$ are parental beliefs about children's future and current skills, the negative coefficient on school average skill s_j in Equation (C.7) can be interpreted as the negative effects of current school skill on future parental beliefs about child skill.[79]

If the combined coefficient on s_j in Equation (C.7) is larger in absolute value than w_s, then the local distortion will increase over time. However, this is not so if prior beliefs about the school average skill are nearly the same as the true skill ($\phi_s \approx$1); thus, the local distortion might decrease over time.[80] In addition, if the negative coefficient ($\Omega\gamma_s$+$(1-\Omega\gamma_c)w_s$) on s_j is offset by the positive term (-$\Omega\gamma_s\phi_s$) on s_j, then local distortions might not arise over time. If this is not so, then local distortions might persist.

C.2.1 Conceptual Framework for Rank Effects on Parental Investment

To further understand how parental beliefs about child skill affect parental investment through class rank, I assume that parental investment is linearly associated with parental beliefs about child skill. Thus, parental investment is a linear function of parental beliefs about child skill:

$$P^I = a \times P^B(c_{ij}, s_j) \tag{C.8}$$

where P^I is parental investment. $P^B(c_{ij}, s_j)$ is the parental belief about child skill with child skill (c_{ij}) and average school skill (s_j). Following (49), if P^I is a complement with $P^B(c_{ij}, s_j)$, then a > 0; otherwise, a < 0.

[79]The existence of local distortions in parental beliefs means that parents of children attending schools with low (high) average skills tend to believe that their child is higher (lower) in the overall skill distribution than they actually are (48).

[80]If $\phi_s \approx$1, then the first two terms in the coefficient on s_j will cancel out. Because $\Omega\gamma_c$ is less than 1, the last term in the coefficient on s_j is smaller in absolute value than w_s.

By taking the derivative to both sides of Equation (C.8) with respect to s_j, I have

$$\frac{\partial P^I}{\partial s_j} = a \times \frac{\partial P^B(c_{ij}, s_j)}{\partial s_j} \tag{C.9}$$

$\frac{\partial P^I}{\partial s_j}$ can be rewritten as $\frac{\partial P^I}{\partial Rank_{ijcs}} \times \frac{\partial Rank_{ijcs}}{\partial s_j}$. Then, I obtain:

$$\begin{aligned} \frac{\partial P^I}{\partial Rank_{ijcs}} \times \frac{\partial Rank_{ijcs}}{\partial s_j} = a \times \frac{\partial P^B(c_{ij}, s_j)}{\partial s_j} \Longrightarrow \\ \frac{\partial P^I}{\partial Rank_{ijcs}} = a \times \frac{\partial P^B(c_{ij}, s_j)}{\partial s_j} \times \frac{1}{\frac{\partial Rank_{ijcs}}{\partial s_j}} \end{aligned} \tag{C.10}$$

Equation (C.10) indicates that the effect of class rank on parental investment ($\frac{\partial P^I}{\partial Rank_{ijcs}}$) can be divided into three parts: the relationship (a) between parental investment and parental beliefs about child skill, the relationship ($\frac{\partial P^B(c_{ij},s_j)}{\partial s_j}$) between school quality and parental beliefs about child skill, and the reciprocal ($\frac{1}{\frac{\partial Rank_{ijcs}}{\partial s_j}}$) of the relationship between school quality and class rank.

APPENDIX D

DATA APPENDIX

D.1 KLoSA Variable Definitions

D.1.1 Healthcare Utilization

This paper uses self-reported healthcare utilization as the primary outcome of interest. The KLoSA includes questions that ask individuals about their inpatient, outpatient, and dental services use:

Inpatient Service Use

1. Were you ever hospitalized since the last interview?

2. If yes for the question 1, how many did you do so ever?

3. How much did you pay for it?

Outpatient Service Use

1. Did you use outpatient services in hospitals or clinics since the last interview?

2. If yes for the question 1, how many did you do so ever?

3. How much did you pay for it?

Dental Service Use

1. Did you use dental services since the last interview?

2. If yes for the question 1, how many did you do so ever?

3. How much did you pay for it?

D.1.2 Health Outcomes

The KLoSA asked individuals about a variety of their diseases and mental health status:

Physical Diseases

1. Do you suffer from high blood pressure/diabetes/cancer/lung/live/heart/brain/bone disease?

Mental Health

1. Do you suffer from psychiatric illness?

D.1.3 Other Variables

The KLoSA includes questions about the national pension received and the amount of the national pension benefit:

Pension Received

1. Did you receive the national pension benefit last year?

Pension Benefit

1. How much did you receive the national pension benefit last year?

D.2 The Construction of Healthcare Utilization Indices

To develop healthcare utilization indices, I follow the method of (4):

$$h_{ij} = \frac{1}{W_{ij}} \sum_{k \in K_{ij}} w_{jk} \frac{y_{ijk} - \bar{y}_{jk}}{\sigma^y_{jk}} \tag{D.1}$$

where k indexes individual healthcare outcomes (e.g. the number of hospital visits) within specific area j. These outcomes are normalized with mean zero and standard deviation one. K_{ij} is the set of non-missing outcomes for observation i in area j. σ^y_{jk} is the control group standard deviation for outcome k in area j. w_{jk} is the outcome weight from the inverted covariance matrix $\hat{\sum}_j^{-1}$. W_{ij}= $\sum_{k \in K_{ij}} w_{jk}$.

Specifically, w_{jk}=$\sum_{l=1}^{K_j} c_{jkl}$,

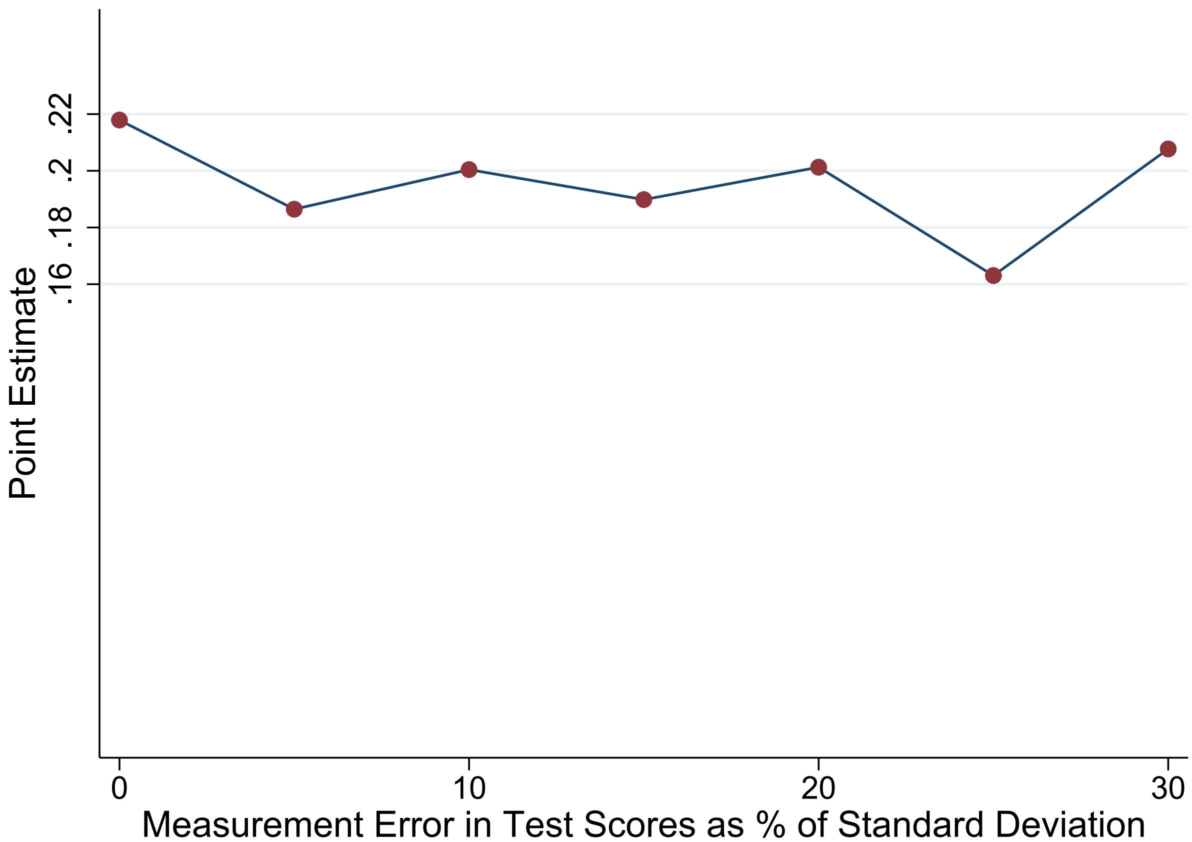
$$\hat{\sum}_j^{-1} = \begin{bmatrix} c_{j11} & c_{j12} & \cdots & c_{j1K} \\ c_{j21} & c_{j22} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots \\ c_{jK1} & \vdots & \ddots & c_{jKK} \end{bmatrix}$$

and $\hat{\sum}_j$ consists of elements: $\hat{\sum}_{jmn} = \sum_{i=1}^{N_{jmn}} \frac{y_{ijm} - \bar{y}_{jm}}{\sigma_{jm}^{y}} \frac{y_{ijn} - \bar{y}_{jn}}{\sigma_{jn}^{y}}$. where K_j is the total number of outcomes of area j. N_{jmn} is the number of observations not missing for both outcome m and outcome n in area j.

APPENDIX E

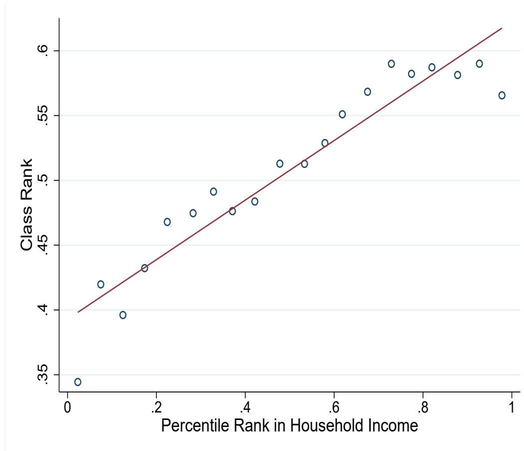
APPENDIX FIGURES

Figure E.1: Monte Carlo Simulation Result: Additional Measurement Error in Test Scores

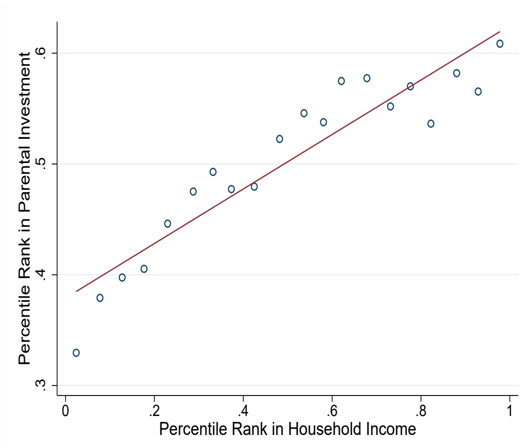


Notes: This figure presents the mean rank estimate from 1,000 simulations of specification (1) in Table 2.5 with increasing additional measurement error added to 7^{th} grade test scores. The rank is re-computed for each test score reflecting the error. The measurement error is randomly drawn from a normal distribution with mean zero and a standard deviation that is proportional to the standard deviation (one) in the test score.

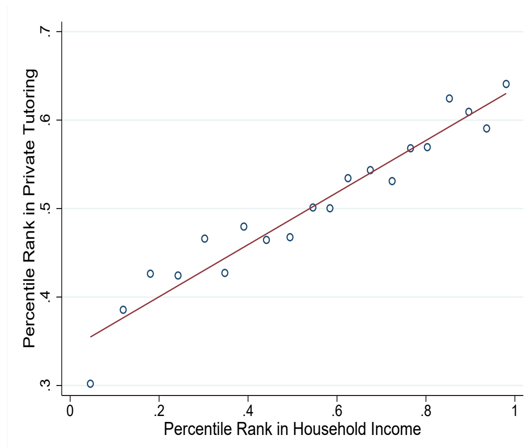
Figure E.2: Relationship between Household Income Rank and Class Rank and Parental Investment, and Private Tutoring Rank



(a) The Rank-Rank Correlation



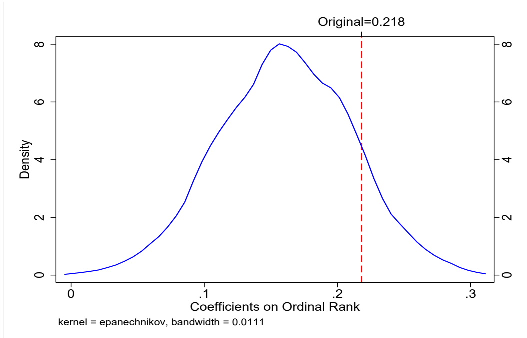
(b) The Rank-Rank Correlation



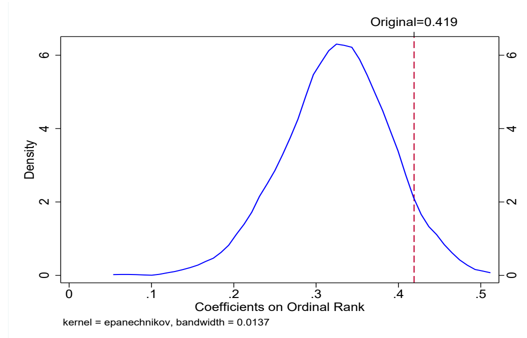
(c) The Rank-Rank Correlation

Notes: This figure presents the correlations between household income rank and class rank in 7^{th} grade (Panel A) and parental investment rank (Panel B), and private tutoring rank (Panel C). Household income rank, parental investment rank, and private tutoring rank are computed by Equation (1) and on the x-axis in each panel. These ranks are local percentile ranks within a classroom. Parental investment represent the parental investment index. Class rank in 7^{th} grade is on the x-axis in each panel.

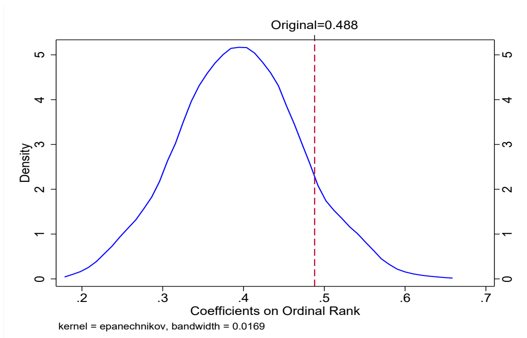
Figure E.3: Placebo Tests



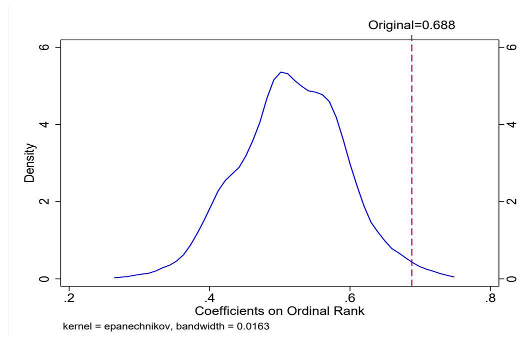
(a) Panel A: Effect on Test Score in 8^{th} grade



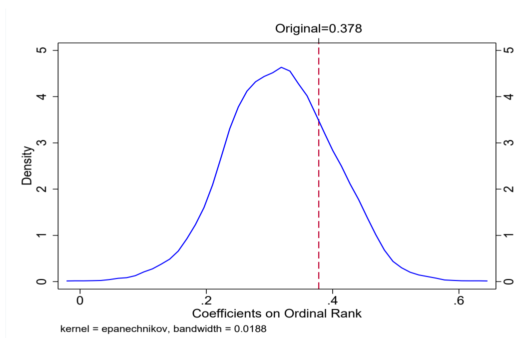
(b) Panel B: Effect on Test Score in 9^{th} grade



(c) Panel C: Effect on Test Score in 10^{th} grade



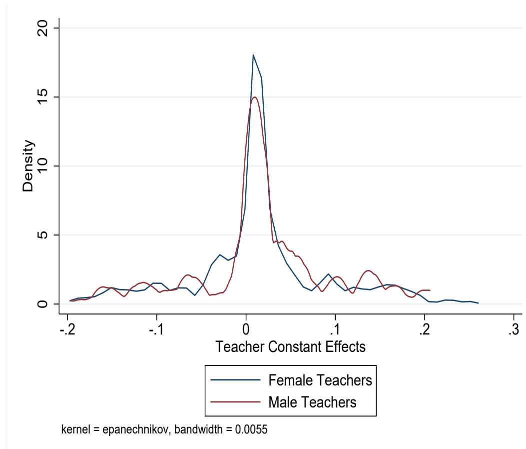
(d) Panel D: Effect on Test Score in 11^{th} grade



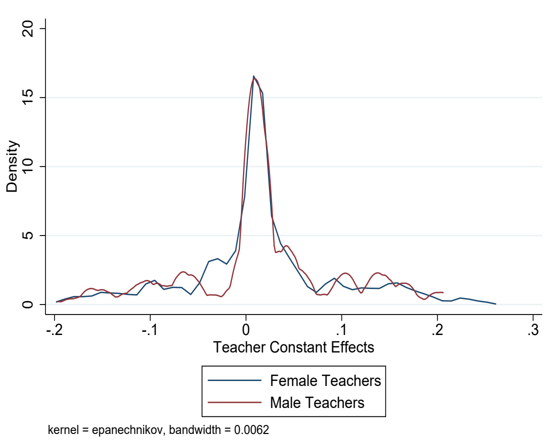
(e) Panel E: Effect on Test Score in 12^{th} grade

Notes: These graphs depict the results from a permutation exercise where students are randomly reassigned into classrooms of the same size as the existing classrooms for 1,000 replications. In each panel, the dashed line denotes the original estimate from Panel A of Table 6.

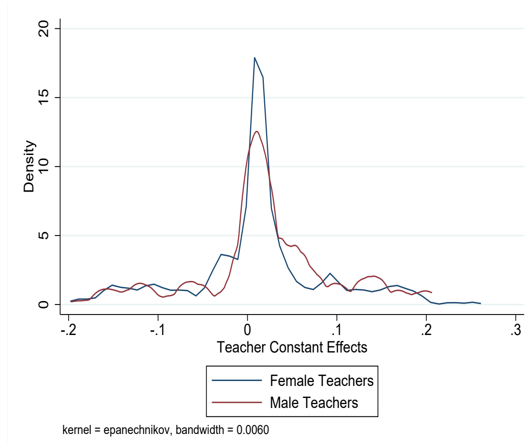
Figure E.4: Individual Teacher Effects on Class Rank



(a) Panel A: Teacher Effects



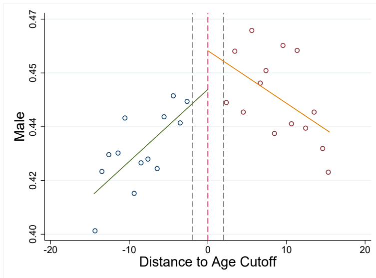
(b) Panel B: Teacher Effects for Female Students



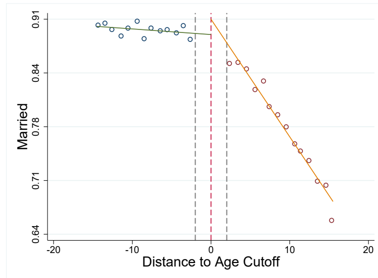
(c) Panel C: Teacher Effects for Male Students

Notes: These graphs display the distribution of random effects of individual (subject) teacher effects by teacher gender (a) and student gender (b) and (c). These random effects are calculated by first eliminating school by subject fixed effects and other student and teacher controls. Student controls are indicators for a single-parent household, number of siblings, and indicators for having at least one parent with a BA degree or higher and for free lunch. Teacher controls are indicators for teacher age over 40, for teacher's graduation from teachers college, for teacher having a Master's degree or higher, and for administrative teacher. By using residualized class rank, I then obtain a constant random effect for each teacher in panel (a), a constant random effect for each teacher and female students in panel (b), or a constant random effect for each teacher and male students in panel (c).

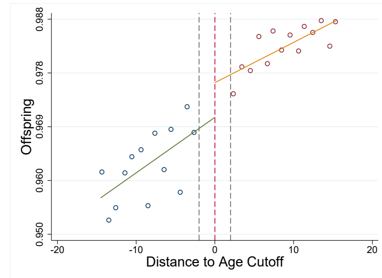
Figure E.5: Balance on Baseline Covariates or Predicted Indices for Healthcare Utilization



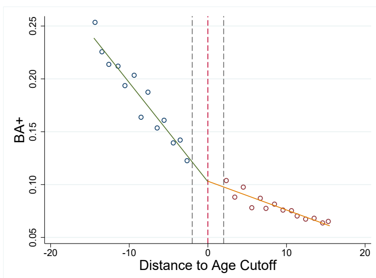
(a) Panel A: Male



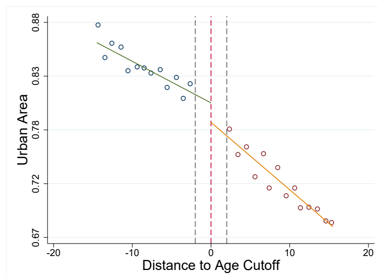
(b) Panel B: Married



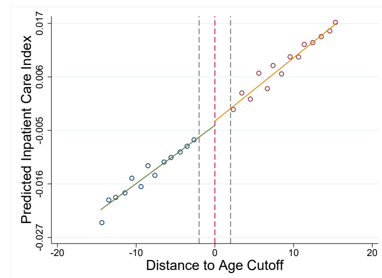
(c) Panel C: Offspring



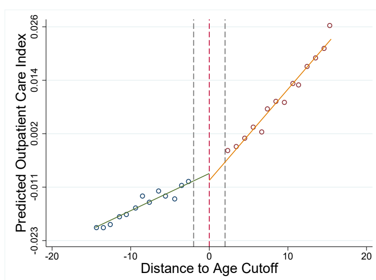
(d) Panel D: Low Education



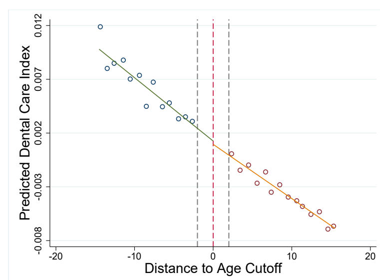
(e) Panel E: Urban Area



(f) Panel F: Inpatient Care Index



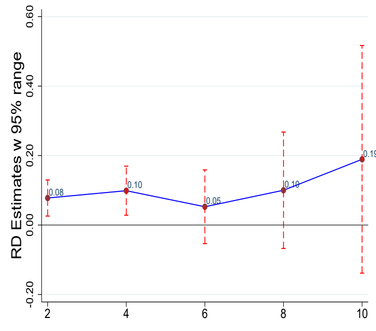
(g) Panel G: Outpatient Care Index



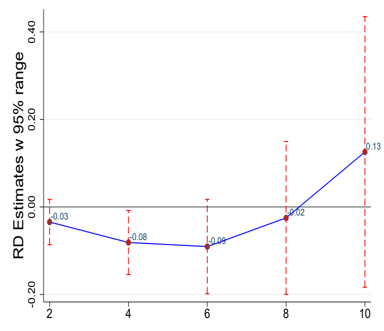
(h) Panel H: Dental Care Index

Notes: This figure depicts the discontinuity in baseline covariates (Panels A–E) or predicted healthcare utilization indices (Panels F–H). These predicted indices are obtained by regressions of each healthcare utilization index on covariates. Covariates include the indicators for whether an individual is male, for whether an individual has a spouse, for whether an individual has at least one offspring, for whether an individual has a BA degree or higher, and for whether an individual lives in an urban area. In each panel, the value on the y-axis reports the mean outcome by the distance from the cutoff (dotted line) with one width of bins. Two gray dotted lines represent the donut hole size, excluding 4 bins around the age cutoff. The best-fit solid lines are linear fitted on each side of the cutoff.

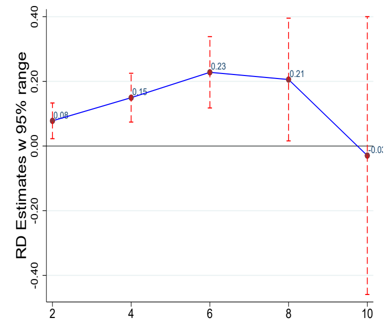
Figure E.6: RD Estimates by Donut Size



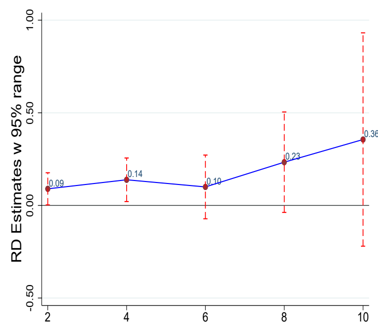
(a) Panel A: Inpatient Care Index (All)



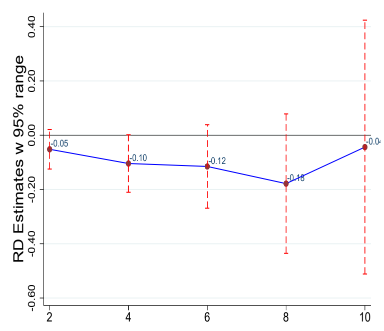
(b) Panel B: Outpatient Care Index (All)



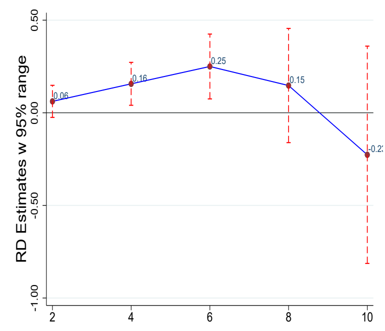
(c) Panel C: Dental Care Index (All)



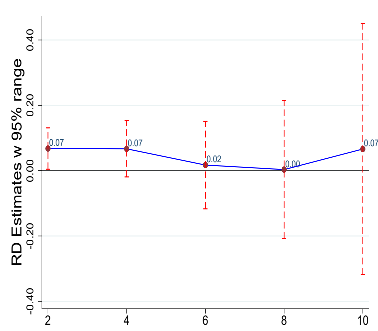
(d) Panel D: Inpatient Care Index (Men)



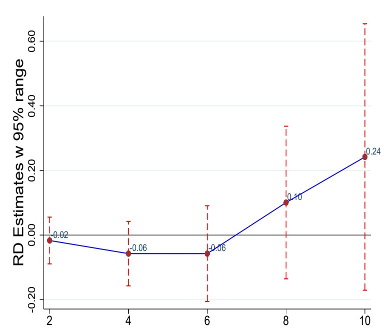
(e) Panel E: Outpatient Care Index (Men)



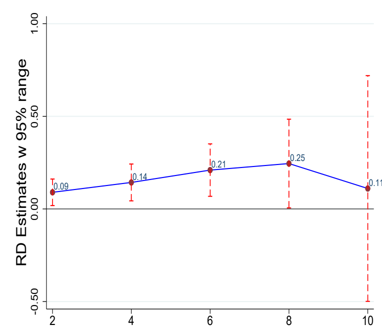
(f) Panel F: Dental Care Index (Men)



(g) Panel G: Inpatient Care Index (Women)



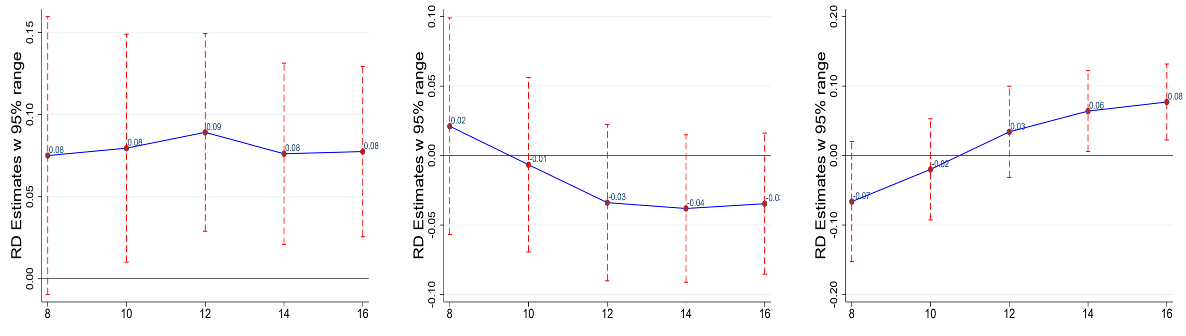
(h) Panel H: Outpatient Care Index (Women)



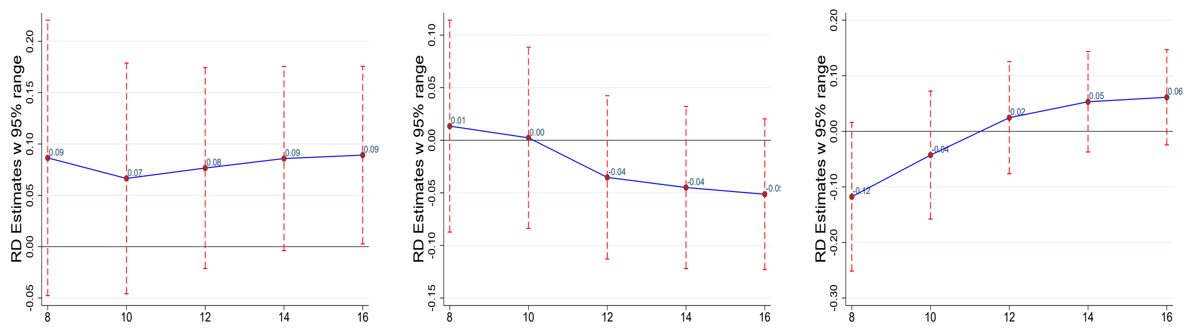
(i) Panel I: Dental Care Index (Women)

Notes: This figure depicts sensitivity to the donut size. Each point estimate for a different size of the donut hole is obtained by a local linear regression for the donut RD design. I use a range of donut holes, from 2 to 10. The dashed lines represent the 95% confidence intervals.

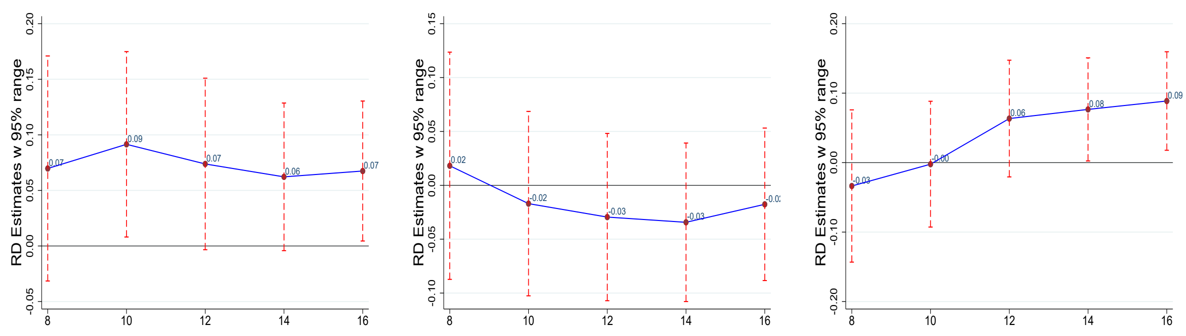
Figure E.7: Checks for Sensitivity to Bandwidth (1)



(a) Panel A: Inpatient Care Index (All) (b) Panel B: Outpatient Care Index (All) (c) Panel C: Dental Care Index (All)



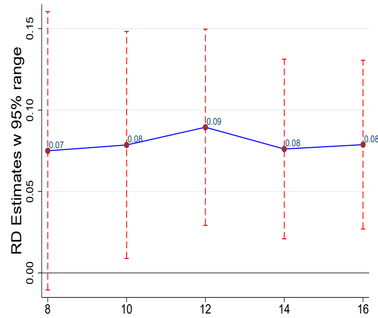
(d) Panel D: Inpatient Care Index (Men) (e) Panel E: Outpatient Care Index (Men) (f) Panel F: Dental Care Index (Men)



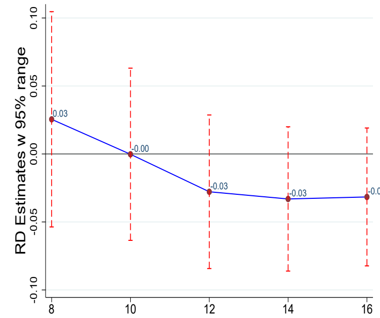
(g) Panel G: Inpatient Care Index (Women) (h) Panel H: Outpatient Care Index (Women) (i) Panel I: Dental Care Index (Women)

Notes: This figure depicts sensitivity to bandwidths. Each point estimate for a different choice of bandwidth is obtained by a local linear regression for the donut RD design. I use a range of bandwidths, from 8 to 16. The dashed lines represent the 95% confidence intervals.

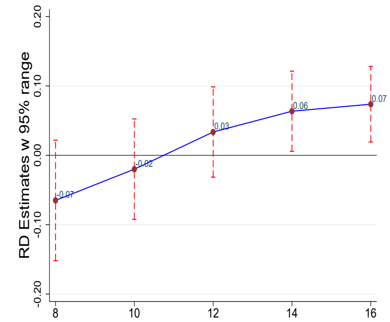
Figure E.8: Checks for Sensitivity to Bandwidth (2)



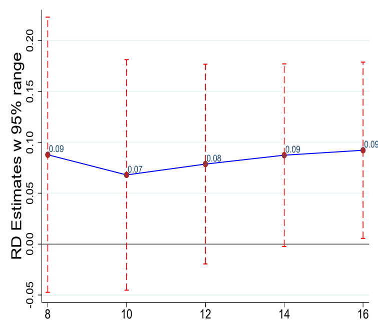
(a) Panel A: Inpatient Care Index (All)



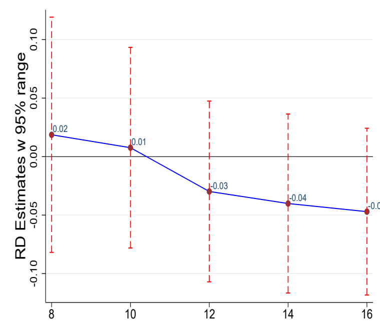
(b) Panel B: Outpatient Care Index (All)



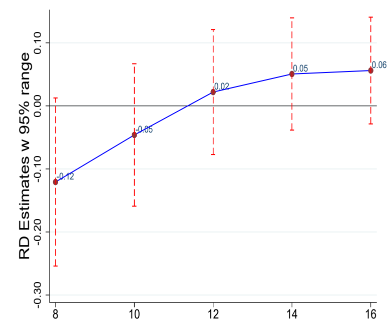
(c) Panel C: Dental Care Index (All)



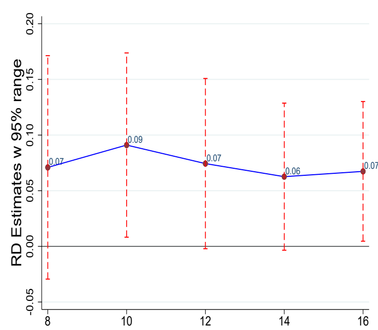
(d) Panel D: Inpatient Care Index (Men)



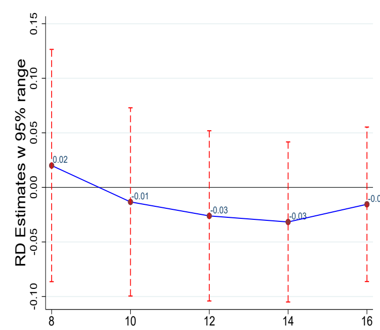
(e) Panel E: Outpatient Care Index (Men)



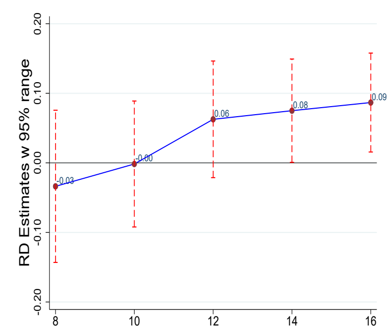
(f) Panel F: Dental Care Index (Men)



(g) Panel G: Inpatient Care Index (Women)



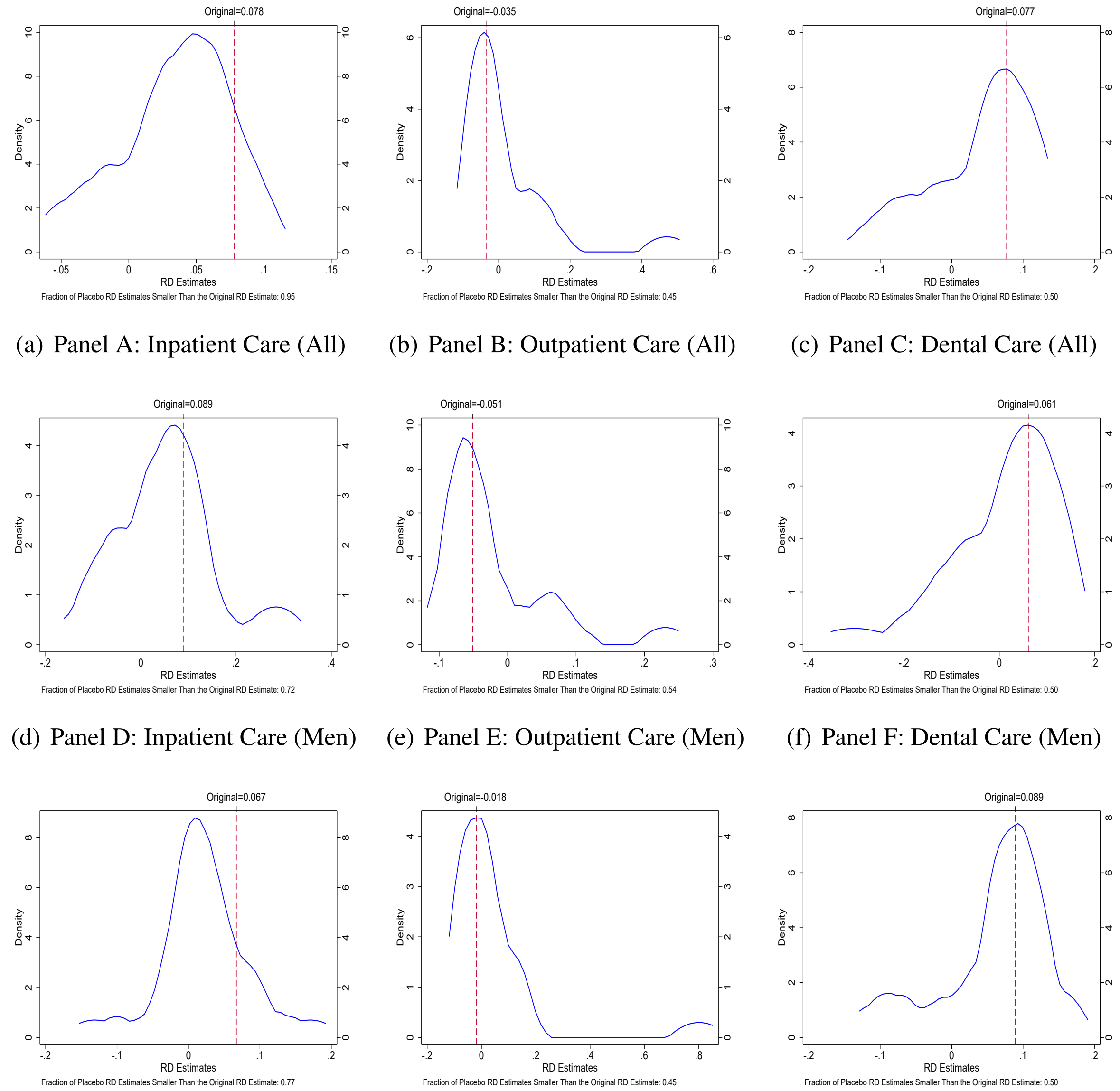
(h) Panel H: Outpatient Care Index (Women)



(i) Panel I: Dental Care Index (Women)

Notes: This figure depicts sensitivity to bandwidths. Each point estimate for a different choice of bandwidth is obtained by a global quadratic regression for the donut RD design. I use a range of bandwidths, from 8 to 16. The dashed lines represent the 95% confidence intervals.

Figure E.9: Densities of Placebo Effects on Healthcare Utilization



(a) Panel A: Inpatient Care (All) (b) Panel B: Outpatient Care (All) (c) Panel C: Dental Care (All)

(d) Panel D: Inpatient Care (Men) (e) Panel E: Outpatient Care (Men) (f) Panel F: Dental Care (Men)

(g) Panel G: Inpatient Care (Women) (h) Panel H: Outpatient Care (Women) (i) Panel I: Dental Care (Women)

Notes: This figure displays the distribution of the estimates from a local linear regression at the placebo cutoffs from [-11, -1] and [1,11] of the running variable in increments of one. The separate regression follows the donut specification dropping observations within a bandwidth of four bins from the age cutoff. The dependent variables are inpatient care index, outpatient care index, and dental care index. The red dotted line represents my original ITT estimate as reported in Table 4.5. At the bottom of each graph, I report the fraction of times placebo RD estimates are smaller than my original RD estimate.

APPENDIX F

APPENDIX TABLES

Table F.1: Rank Effects on Test Scores in Context

Study	Population	Age	Short-Run Estimate for Test Score	Long-Run Estimate for Test Score
This Paper	South Korea	Age 13 or 14	0.208 (One Year Later)	0.378 (Five Years Later)
(2)	UK	Age 11	0.290 (Thee Years Later)	0.241 (Five Years Later)
(21)	US	Age 9	-	0.147 (Five Years Later)
(26)	China	Age 13	0.650 (One Year Later)	-

Notes: Age represents the initial age that students' ranks are defined in each study. All rank measures are class (classroom) percentile ranks. Test score outcomes are normalized with a mean zero and a standard deviation one.

Table F.2: Specification Check (1): Functional Form of 7^{th} Grade Test Scores

	(1) Linear	(2) Quadratic	(3) Cubic	(4) Quartic	(5) Quintic
7^{th} Grade Rank	0.467***	0.303***	0.218**	0.218**	0.203*
	(0.105)	(0.108)	(0.108)	(0.109)	(0.109)
Observations	12,913	12,913	12,913	12,913	12,913
R^2	0.524	0.527	0.528	0.529	0.529

Notes: Each column represents results from a separate regression including cubic in 7^{th} grade test score and classroom by subject fixed effects. Standard errors, shown in parentheses, are clustered at the classroom level.
*p<0.10, ** p<0.05, *** p<0.01.

Table F.3: Specification Check (2): Interacting Rank with the Standard Deviation of Class Test Scores

	(1) Test Score 8th Grade	(2) Test Score 9th Grade	(3) Test Score 10th Grade	(4) Test Score 11th Grade	(5) Test Score 12th Grade
7^{th} Grade Rank	0.228*	0.429***	0.494***	0.667***	0.617***
$\times$ Std. of Test Scores	(0.133)	(0.147)	(0.184)	(0.191)	(0.186)
Observations	12,913	12,312	10,247	10,085	9,733
R^2	0.528	0.260	0.194	0.246	0.138

Notes: Each column represents results from a separate regression, including the standard deviation of 7^{th} grade test scores within classroom by subject, cubic in 7^{th} grade test score, and classroom by subject fixed effects. Dependent variables in Columns (1)–(5) are standardized test scores within a subject and a year from 8^{th} through 12^{th} grade, respectively. Standard errors, shown in parentheses, are clustered at the classroom level.
*p<0.10, ** p<0.05, *** p<0.01.

Table F.4: Specification Check (3): Interacting 7^{th} grade test score with Classroom and Subject Effects

	(1) Test Score 8th Grade	(2) Test Score 9th Grade	(3) Test Score 10th Grade	(4) Test Score 11th Grade	(5) Test Score 12th Grade
A. Test Scores Interacted with Classroom Fixed Effects					
7^{th} Grade Rank	0.192	0.600***	0.449**	0.757***	0.443**
	(0.116)	(0.160)	(0.215)	(0.198)	(0.187)
Observations	12,913	12,312	10,247	10,085	9,733
R^2	0.556	0.323	0.243	0.298	0.183
B. Test Scores Interacted with Subject Fixed Effects					
7^{th} Grade Rank	0.235**	0.399***	0.447***	0.668***	0.353**
	(0.109)	(0.142)	(0.153)	(0.166)	(0.146)
Observations	12,913	12,312	10,247	10,085	9,733
R^2	0.532	0.283	0.231	0.254	0.160

Notes: Each cell represents results from a separate regression, including cubic in 7^{th} grade test score interacting with classroom fixed effects (Panel A) or subject fixed effects (Panel B), and classroom by subject fixed effects. Dependent variables in Columns (1)–(5) are standardized test scores within a subject and a year from 8^{th} through 12^{th} grade, respectively. Standard errors, shown in parentheses, are clustered at the classroom level.
*p<0.10, ** p<0.05, *** p<0.01.

Table F.5: Robustness Check

	(1) Test Score 8th Grade	(2) Test Score 9th Grade	(3) Test Score 10th Grade	(4) Test Score 11th Grade	(5) Test Score 12th Grade
A. Controlling for School by Subject by Abg. FEs					
7^{th} Grade Rank	0.263**	0.361***	0.424***	0.615***	0.479***
	(0.117)	(0.111)	(0.117)	(0.129)	(0.121)
Observations	12,783	12,189	10,148	9,993	9,646
R^2	0.536	0.304	0.219	0.275	0.159
B. Balanced Sample Only					
7^{th} Grade Rank	0.267**	0.397***	0.492***	0.675***	0.403**
	(0.118)	(0.152)	(0.163)	(0.174)	(0.164)
Observations	9,209	9,209	9,209	9,209	9,209
R^2	0.541	0.255	0.200	0.248	0.141
C. Controlling for Subject Teacher FEs					
7^{th} Grade Rank	0.206	0.486***	0.424**	0.582**	0.230
	(0.126)	(0.161)	(0.193)	(0.227)	(0.181)
Observations	9,733	9,269	7,622	7,492	7,234
R^2	0.549	0.344	0.250	0.303	0.192
D. Controlling for Student Ctrls, Subject Teacher Ctrls & Peer Ctrls					
7^{th} Grade Rank	0.214	0.498***	0.451**	0.538**	0.334*
	(0.133)	(0.168)	(0.221)	(0.233)	(0.172)
Observations	8,577	8,174	6,751	6,633	6,393
R^2	0.533	0.307	0.220	0.285	0.162

Notes: Each cell represents results from a separate regression, including cubic in 7^{th} grade test score and classroom by subject fixed effects (in Panels B–E). Dependent variables in Columns (1)–(5) are standardized test scores within a subject and a year from 8^{th} through 12^{th} grade, respectively. Student controls are indicators for student gender, for a single-parent household, number of siblings, and indicators for having at least one parent with a BA degree or higher and for free lunch. Teacher controls are indicators for teacher gender, for teacher age over 40, for teacher's graduation from teachers college, for teacher having a Master's degree or higher, and for administrative teacher. Peer controls are leave-out means of student controls and test scores, and the standard deviation of peers' test scores. Standard errors, shown in parentheses, are clustered at the school (in Panel A) or classroom (in Panels B–D) level. *p<0.10, ** p<0.05, *** p<0.01.

Table F.6: Effects of Teacher–Student Gender Match on Class Rank in 7^{th} Grade

	(1)	(2)	(3)	(4)	(5)
Female Student	0.032**	0.031*	0.022*	0.023	0.022
	(0.011)	(0.017)	(0.013)	(0.024)	(0.024)
Female Teacher	-0.038	-0.031***	-0.042	-0.030**	-0.042
	(0.009)	(0.033)	(0.012)	(0.013)	(0.034)
Female Student $\times$ Female Teacher	0.051***	0.052***	0.050***	0.055***	0.055***
	(0.014)	(0.016)	(0.016)	(0.018)	(0.018)
Observations	10,056	9,197	8,854	8,854	8,854
R^2	0.045	0.304	0.303	0.303	0.303
Dep. Var. Mean	0.507	0.516	0.517	0.517	0.517
School by Subject Fixed Effects	Y	Y	Y	Y	Y
Student Controls	N	Y	Y	Y	Y
Teacher Controls	N	N	Y	Y	Y
Female Student $\times$ Teacher Controls	N	N	N	Y	Y
Female Teacher $\times$ Teacher Controls	N	N	N	N	Y

Notes: Each column represents results from a separate regression, including school by subject fixed effects. The dependent variable is class rank in 7^{th} Grade. Student controls are indicators for a single-parent household, number of siblings, and indicators for having at least one parent with a BA degree or higher and for free lunch. Teacher controls are indicators for teacher age over 40, for teacher's graduation from teachers college, for teacher having a Master's degree or higher, and for administrative teacher. Standard errors, shown in parentheses, are clustered at the school level. *p<0.10, ** p<0.05, *** p<0.01.

Table F.7: Subject-Specific Rank Effects

	(1) Test Score 8th Grade	(2) Test Score 9th Grade	(3) Test Score 10th Grade	(4) Test Score 11th Grade	(5) Test Score 12th Grade
A. Math Score					
7^{th} Grade Math Rank	-0.006	-0.207	-0.492	0.326	0.045
	(0.206)	(0.284)	(0.300)	(0.339)	(0.306)
7^{th} Grade English Rank	0.134	0.264	0.728***	0.642**	0.020
	(0.148)	(0.219)	(0.272)	(0.291)	(0.274)
7^{th} Grade Korean Rank	-0.065	-0.092	0.239	0.281	-0.157
	(0.146)	(0.233)	(0.270)	(0.267)	(0.271)
Observations	4,263	4,057	3,379	3,335	3,233
R^2	0.570	0.192	0.150	0.222	0.100
B. English Score					
7^{th} Grade Math Rank	-0.210	-0.301	-0.389	-0.034	0.081
	(0.149)	(0.253)	(0.310)	(0.288)	(0.346)
7^{th} Grade English Rank	0.369**	0.588**	0.762***	0.584	0.333
	(0.163)	(0.228)	(0.277)	(0.295)	(0.277)
7^{th} Grade Korean Rank	0.085	0.316	0.235	0.430	-0.060
	(0.146)	(0.250)	(0.286)	(0.263)	(0.269)
Observations	4,263	4,067	3,394	3,335	3,234
R^2	0.654	0.247	0.188	0.255	0.139
C. Korean Score					
7^{th} Grade Math Rank	-0.079	-0.080	0.151	0.336	0.273
	(0.154)	(0.168)	(0.204)	(0.266)	(0.234)
7^{th} Grade English Rank	0.317**	0.322	0.565**	0.358	-0.293
	(0.148)	(0.179)	(0.230)	(0.236)	(0.240)
7^{th} Grade Korean Rank	0.053	0.414**	0.509***	0.503**	0.603**
	(0.177)	(0.181)	(0.186)	(0.210)	(0.235)
Observations	4,264	4,073	3,391	3,336	3,188
R^2	0.467	0.474	0.432	0.385	0.303

Notes: Each cell represents results from a separate regression including cubic in 7^{th} grade test score and classroom fixed effects. Dependent variables in Columns (1)–(5) are the test scores in each subject from 8^{th} through 12^{th} grade, respectively. Each panel shows the impact of rank in all subjects on test scores in each subject separately, including the impacts of rank on both the same-subject and other subject outcomes. Standard errors, shown in parentheses, are clustered at the classroom level.
*p<0.10, ** p<0.05, *** p<0.01.

Table F.8: Rank Effects on Test Scores and Misinformation

	(1) Test Score 8th Grade	(2) Test Score 9th Grade	(3) Test Score 10th Grade	(4) Test Score 11th Grade	(5) Test Score 12th Grade
A. Not Controlling for the Degree of Misinformation					
Avg. 7^{th} Grade Rank	0.277	0.603*	1.295***	1.544***	0.417
	(0.215)	(0.329)	(0.420)	(0.343)	(0.384)
Observations	4,262	4,038	3,374	3,334	3,187
R^2	0.669	0.429	0.386	0.421	0.246
B. Controlling for the Degree of Misinformation					
Avg. 7^{th} Grade Rank	0.174	0.516	1.181***	1.172***	0.344
	(0.226)	(0.325)	(0.397)	(0.341)	(0.382)
Misinformation	-0.413	-0.363	-0.549	-1.744***	-0.342
	(0.244)	(0.389)	(0.436)	(0.414)	(0.416)
Observations	4,262	4,038	3,374	3,334	3,187
R^2	0.669	0.429	0.387	0.425	0.246
C. Controlling for the Degree of Misinformation Interacted with Student Gender					
Avg. 7^{th} Grade Rank	0.169	0.512	1.180***	1.178***	0.353
	(0.226)	(0.323)	(0.398)	(0.342)	(0.380)
Misinformation	-0.696**	-0.778*	-0.719	-2.204***	-0.894**
× Male Student	(0.282)	(0.397)	(0.450)	(0.423)	(0.447)
Misinformation	-0.109	0.091	-0.355	-1.231***	0.269
× Female Student	(0.269)	(0.415)	(0.462)	(0.451)	(0.435)
Observations	4,262	4,038	3,374	3,334	3,187
R^2	0.670	0.431	0.387	0.428	0.250

Notes: Each cell represents results from a separate regression including cubic in 7^{th} grade test score and classroom fixed effects. Dependent variables in Columns (1)–(5) are the average test scores across subjects from 8^{th} through 12^{th} grade, respectively. The rank is the average rank across subjects in 7^{th} grade. In Panel B, the misinformation variable represents the absolute differences between local (classroom) rank and global (in-sample) rank in 7^{th} grade. In Panel C, the interaction terms of the misinformation variable with dummies for student gender are added. Standard errors, shown in parentheses, are clustered at the classroom level. At the bottom of Panel C, the p-values of the F-test for the difference between each pair of the coefficients on the interaction term of the misinformation variable with the dummy for male or female students are presented. *p<0.10, ** p<0.05, *** p<0.01.

Table F.9: Distribution of Class Rank by School Quality

Decile	(1) Low-Quality School	(2) Middle-Quality School	(3) High-Quality School	(4) (3)-(1)
1	0.091	0.064	0.042	-0.049
2	0.232	0.157	0.102	-0.130
3	0.368	0.256	0.153	-0.215
4	0.491	0.355	0.229	-0.262
5	0.590	0.473	0.315	-0.275
6	0.674	0.580	0.425	-0.249
7	0.774	0.674	0.520	-0.254
8	0.859	0.785	0.643	-0.216
9	0.926	0.886	0.755	-0.171
10	0.977	0.959	0.890	-0.087
Mean Value-Added	-0.109	-0.027	0.145	0.254

Notes: This table categorizes middle schools into "low-quality" (Column (1)), "middle-quality" (Column (2)), and "high-quality" (Column (3)) according to school average test scores in 7^{th} grade. Each row shows the average class rank of students in three qualities of school for that decile. Column (4) presents the average rank change of students in that decile from moving from a "low-quality" to a "high-quality" school. The mean value-added is the average school value-added from 8^{th} to 12^{th} grade. The school value-added in each grade is recovered from a middle school fixed effect, obtained from the regression of test scores in each grade on a cubic of test scores in 7^{th} grade and school fixed effects.

Table F.10: Net Gain of Moving from a "Low-quality" to a "High-quality" School Net of Rank Effects

	(1) 8th Grade	(2) 9th Grade	(3) 10th Grade	(4) 11th Grade	(5) 12th Grade
A. Value-Added Gains from a "low-quality" to a "high-quality" School					
All	0.254	0.254	0.254	0.254	0.254
B. Correlations Between Class Rank and Test Scores					
All	1.943	1.053	0.854	1.290	0.727
C. Expected Class Rank Changes					
Low-Achieving Student	-0.049	-0.049	-0.049	-0.049	-0.049
Middle-Achieving Student	-0.275	-0.275	-0.275	-0.275	-0.275
High-Achieving Student	-0.087	-0.087	-0.087	-0.087	-0.087
D. Net Gains Net of Rank Effects (A-B×C)					
Low-Achieving Student	0.159	0.202	0.212	0.191	0.218
Middle-Achieving Student	-0.280	-0.036	0.019	-0.101	0.054
High-Achieving Student	0.085	0.162	0.180	0.142	0.191

Notes: This table presents the net gain of moving from a "low-quality" to a "high-quality" school (net of rank effects from 8^{th} to 12^{th} grade) by performing simulations. Panel A shows mean value-added gains from a "low-quality" to a "high-quality" school, which is the same as the value at the bottom of Column (4) of Appendix Table E9. Panel B shows the OLS estimates from the regression of test scores from 8^{th} to 12^{th} grade on class rank in 7^{th} grade. Panel C shows changes in expected class rank for a student in each decile, associated with the value in Column (4) of Appendix Table F9. Panel D shows value-added gains (Panel A) net of rank effects (Panel B $\times$ Panel C). Specifically, 0.159 (Row (1), Column (1) of Panel D) is equal to 0.254+(1.943×(-0.049)). A low- , middle-, and high-achieving student is a student in decile 1, decile 5, and decile 10, respectively, as defined in Appendix Table E9.

Table F.11: Mean Differences in Outcomes across Treatment Status

	(1) Treatment	(2) Control	(3) Difference
Pension Received	0.339	0.021	0.318***
Pension Benefits (10,000 KRW)	111.01	7.10	103.92***
Retirement	0.362	0.136	0.226***
Inpatient Care Index	0.053	-0.080	0.134***
Outpatient Care Index	0.126	-0.194	0.324***
Dental Care Index	0.027	-0.034	0.066***

Notes: This table reports means and the difference in means between individuals below the threshold age (Column (2)) and those above the threshold age (Column (1)), along with the statistical significance of the difference.
*p<0.10, ** p<0.05, *** p<0.01.

Table F.12: RD Estimates on Healthcare Utilization (ITT)

	(1) Inpatient Care Index	(2) Outpatient Care Index	(3) Dental Care Index
A. With Covariates, All			
$1[Age_{it} \geq C_{bt}]$	0.077***	-0.032	0.078***
	(0.026)	(0.026)	(0.028)
Observations	36,072	35,964	36,149
B. With Covariates, Men			
$1[Age_{it} \geq C_{bt}]$	0.090**	-0.050	0.064
	(0.044)	(0.036)	(0.044)
Observations	15,823	15,790	15,867
C. With Covariates, Women			
$1[Age_{it} \geq C_{bt}]$	0.068**	-0.017	0.089**
	(0.032)	(0.036)	(0.036)
Observations	20,249	20,174	20,282

Notes: Each cell represents results from a local linear regression. The dependent variables in Columns (1)–(3) are the indices for inpatient, outpatient, or dental cares. The process to construct these indexes is outlined by (4). Covariates include the indicators for whether an individual is male, for whether an individual has a spouse, for whether an individual has at least one offspring, for whether an individual has a BA degree or higher, and for whether an individual lives in an urban area. Standard errors, shown in parentheses, are clustered at the individual level.
*p<0.10, ** p<0.05, *** p<0.01.

Table F.13: RD Estimates on Individual Outcomes of Healthcare Utilization (ITT)

	Inpatient Care			**Outpatient Care**			**Dental Care**		
	(1) Incidence	(2) Hospital/Clinic Stays	(3) Cost	(4) Incidence	(5) Hospital/Clinic Visits	(6) Cost	(7) Incidence	(8) Hospital/Clinic Visits	(9) Cost
A. Without Covariates, All									
$1[Age_{it} \geq C_{bt}]$	0.016*	0.035***	2.657*	-0.046***	0.292	-0.243	0.028**	0.374***	1.657
	(0.008)	(0.010)	(1.575)	(0.013)	(0.283)	(2.735)	(0.012)	(0.114)	(1.994)
Observations	36,273	36,273	36,093	36,248	36,248	36,009	36,277	36,277	36,170
B. Without Covariates, Men									
$1[Age_{it} \geq C_{bt}]$	0.021*	0.037**	2.732	-0.046**	0.124	-3.765	0.031	0.178	1.566
	(0.012)	(0.015)	(2.917)	(0.021)	(0.340)	(3.567)	(0.018)	(0.170)	(3.438)
Observations	15,919	15,919	15,834	15,919	15,919	15,805	15,922	15,922	15,878
C. Without Covariates, Women									
$1[Age_{it} \geq C_{bt}]$	0.012	0.033**	2.451	-0.046***	0.486	2.598	0.024	0.523***	1.688
	(0.011)	(0.014)	(1.626)	(0.017)	(0.429)	(4.022)	(0.016)	(0.154)	(2.338)
Observations	20,354	20,354	20,259	20,329	20,329	20,204	20,355	20,355	20,292

Notes: Each cell represents results from a local linear regression. The dependent variables in Columns (1)–(9) are the indicator for whether an individual used inpatient care service, the number of hospital/clinic stays, and the amount of inpatient care cost; the indicator for whether an individual used outpatient care service, the number of hospital/clinic visits, and the amount of outpatient care cost; the indicator for whether an individual used dental care service, the number of dental hospital/clinic visits, and the amount of dental care cost. Standard errors, shown in parentheses, are clustered at the individual level. *p<0.10, ** p<0.05, *** p<0.01.

Table F.14: RD Estimates on Health Outcomes

	(1) High Blood Pressure	(2) Diabetes	(3) Cancer	(4) Lung Disease	(5) Live Disease	(6) Heart Disease	(7) Brain Disease	(8) Bone Disease	(9) Mental Illness
A. All									
$1[Age_{it} \geq C_{bt}]$	0.009	0.018*	0.004	-0.000	-0.002	0.017***	0.006	0.024**	0.008
	(0.014)	(0.010)	(0.006)	(0.004)	(0.005)	(0.007)	(0.005)	(0.011)	(0.005)
Observations	36,285	36,283	36,283	36,272	36,285	36,285	36,152	36,285	36,284
B. Men									
$1[Age_{it} \geq C_{bt}]$	0.011	0.018	0.009	-0.001	-0.010	0.018*	0.000	0.002	0.001
	(0.021)	(0.016)	(0.008)	(0.006)	(0.008)	(0.010)	(0.009)	(0.010)	(0.007)
Observations	15,927	15,926	15,926	15,919	15,927	15,927	15,866	15,927	15,927
C. Women									
$1[Age_{it} \geq C_{bt}]$	0.009	0.017	0.001	-0.000	0.004	0.016*	0.010	0.051***	0.003
	(0.018)	(0.013)	(0.009)	(0.005)	(0.005)	(0.009)	(0.006)	(0.016)	(0.008)
Observations	20,358	20,357	20,357	20,353	20,358	20,358	20,286	20,358	20,357

Notes: Each column represents results from a local linear regression. The dependent variables in Columns (1)–(9) are indicators for high blood pressure, diabetes, cancer, lung disease, live disease, heart disease, brain disease, bone disease, and mental illness. Standard errors, shown in parentheses, are clustered at the individual level.
*p<0.10, ** p<0.05, *** p<0.01.

Table F.15: RD Estimates on Healthcare Utilization (ITT)

	(1) Inpatient Care Index	(2) Outpatient Care Index	(3) Dental Care Index
A. With Health Status Controls, All			
$1[Age_{it} \geq C_{bt}]$	0.059**	-0.050**	0.075***
	(0.026)	(0.025)	(0.028)
Observations	35,922	35,815	35,999
B. With Health Status Controls, Men			
$1[Age_{it} \geq C_{bt}]$	0.070*	-0.065*	0.060
	(0.042)	(0.035)	(0.044)
Observations	15,753	15,719	15,796
C. With Health Status Controls, Women			
$1[Age_{it} \geq C_{bt}]$	0.048	-0.039	0.087**
	(0.031)	(0.035)	(0.036)
Observations	20,169	20,096	20,203

Notes: Each cell represents results from a local linear regression, including health status controls used in Appendix Table F13. The dependent variables in Columns (1)–(3) are the indices for inpatient, outpatient, or dental cares. The process to construct these indexes is outlined by (4). Standard errors, shown in parentheses, are clustered at the individual level.
*p<0.10, ** p<0.05, *** p<0.01.

Table F.16: Bounding Analysis for RD Estimates both on Retirement and Healthcare Utilization (ITT)

	(1) Retirement	(2) Inpatient Care Index	(3) Outpatient Care Index	(4) Dental Care Index
A. Original Estimates				
$1[Age_{it} \geq C_{bt}]$	0.051***	0.077***	-0.032	0.078***
	(0.012)	(0.026)	(0.026)	(0.028)
Observations	36,284	36,072	35,964	36,149
B. Lower Bound Estimates				
$1[Age_{it} \geq C_{bt}]$	0.051***	0.077***	-0.035	0.077**
	(0.010)	(0.028)	(0.027)	(0.031)
Observations	36,269	36,078	35,970	36,155
C. Upper Bound Estimates				
$1[Age_{it} \geq C_{bt}]$	0.062***	0.104***	-0.037	0.172***
	(0.015)	(0.034)	(0.027)	(0.029)
Observations	21,766	21,678	21,647	21,726

Notes: Each cell represents results from a local linear regression. The dependent variables in Columns (1)–(4) are the probability of being in retirement and the indices for inpatient, outpatient, or dental care. The process to construct these indices is outlined by (4). Original estimates (Panel A) corresponds to Panel A, Column (1) of Table 4.4 and Panel A, Columns (1)–(3) of Table 4.5. Lower bound estimates are earned by dropping individuals who have the highest amount of pension benefits in each bin above the age cutoff. Upper bound estimates are earned by dropping individuals who have the lowest amount of pension benefits in each bin above the age cutoff. Standard errors, shown in parentheses, are clustered at the individual level. The standard errors of original estimates are clustered at the individual level (Panel A). Bootstrapped standard errors are reported for the upper (Panel B) and lower bound (Panel C) estimates.
*p<0.10, ** p<0.05, *** p<0.01.

Table F.17: RD Estimates on Individual Outcomes of Healthcare Utilization (LATE)

	Inpatient Care			**Outpatient Care**			**Dental Care**		
	(1) Incidence	(2) Hospital/Clinic Stays	(3) Cost	(4) Incidence	(5) Hospital/Clinic Visits	(6) Cost	(7) Incidence	(8) Hospital/Clinic Visits	(9) Cost
A. Without Covariates, All									
Retirement	0.304*	0.675***	50.78	-0.893***	5.639	-4.788	0.537**	7.294***	32.14
	(0.169)	(0.241)	(31.64)	(0.333)	(5.525)	(53.26)	(0.260)	(2.733)	(39.29)
Observations	36,272	36,272	36,092	36,247	36,247	36,008	36,276	36,276	36,169
1st Stage Kleibergen-Paap F-Stat	19.33	19.33	19.61	19.49	19.49	19.27	19.16	19.16	19.35
B. Without Covariates, Men									
Retirement	0.212*	0.375**	27.97	-0.469**	1.264	-37.50	0.321	1.818	15.91
	(0.123)	(0.159)	(29.81)	(0.225)	(3.458)	(36.18)	(0.191)	(1.747)	(35.02)
Observations	15,919	15,919	15,898	15,919	15,919	15,805	15,922	15,922	15,878
1st Stage Kleibergen-Paap F-Stat	35.85	35.85	35.87	35.85	35.85	37.58	35.43	35.43	35.97
C. Without Covariates, Women									
Retirement	1.202	3.383	203.9	-4.499	47.68	312.9	2.544	54.97	172.1
	(2.254)	(5.731)	(306.0)	(7.491)	(83.90)	(757.5)	(4.577)	(94.68)	(365.1)
Observations	20,353	20,353	20,258	20,328	20,328	20,203	20,354	20,354	20,291
1st Stage Kleibergen-Paap F-Stat	0.365	0.365	0.423	0.391	0.391	0.259	0.343	0.343	0.364

Notes: Each cell represents results from a local linear regression. The dependent variables in Columns (1)–(9) are the indicator for whether an individual used inpatient care services, the number of hospital/clinic stays, and the amount of inpatient care cost; the indicator for whether an individual used outpatient care services, the number of hospital/clinic visits, and the amount of outpatient care cost; the indicator for whether an individual used dental care services, the number of dental hospital/clinic visits, and the amount of dental care cost. The value of Kleibergen-Paap F-statistic for weak instrument at the bottom of each panel is reported. Standard errors, shown in parentheses, are clustered at the individual level.
*p<0.10, ** p<0.05, *** p<0.01.

Table F.18: LATE Parameters in Context

(1) Study	(2) Country	(3) Strategy	(4) Instrument for Retirement	(5) Inpatient Incidence	(6) Hospital/Clinic Stays	(7) Outpatient Incidence	(8) Hospital/Clinic Visits
A. All Individuals							
Current Paper	South Korea	RDD	Pension Eligibility Age	0.304*	0.675***	-0.893***	5.639
				[0.318]	[0.389]	[0.456]	[9.890]
(76)	Germany	RDD	Pension Eligibility Age	-0.045	-	-	-0.131**
				[0.351]	-	-	[4.841]
(85)	US	IV	Pension Eligibility Age	-0.106***	-	-	-0.343
				[0.400]	-	-	[14.18]
(82)	China	RDD	Statutory Retirement Age	0.163	0.409*	0.104	0.816*
				[0.334]	[0.622]	[0.395]	[1.610]
(79)	Denmark	RDD	Pension Eligibility Age	-	-	-	-0.337***
				-	-	-	[5.150]
B. Men							
Current Paper	South Korea	RDD	Pension Eligibility Age	0.212*	0.375**	-0.469**	1.264
				[0.308]	[0.384]	[0.473]	[7.726]
(5)	Austria	IV	Pension Eligibility Age	-	-	-	-0.704**
				-	-	-	[4.240]
(83)	China	RDD	Statutory Retirement Age	0.014	0.045	-	6.857**
				[0.207]	[0.103]	-	[18.35]
(80)	Vietnam	IV	Statutory Retirement Age	0.022	-0.181	0.361***	0.510
				[0.281]	[0.672]	[0.483]	[3.423]

Notes: For each study, the LATE parameters for the effect of retirement on healthcare utilization are presented in Columns (5)–(8). In Column (3), an RDD represents a regression discontinuity design, while an IV means an instrumental variable approach. Raw parameters are used to show the estimates above in Appendix Table C8. Outcomes in Columns (5)–(8) include the indicator for whether individuals received inpatient care services, the number of times individuals received inpatient care services, the indicator for whether individuals received outpatient care services, and the number of times individuals received outpatient care services. The standard deviation of each outcome is shown in parentheses.
*p<0.10, ** p<0.05, *** p<0.01.