Essays on the Effectiveness of Educational Inputs and Organization on Student Achievement, Adult Behaviors and Longevity

By

Kai Hong

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of
DOCTOR OF PHILOSOPHY
in
Economics
August, 2016
Nashville, Tennessee

Approved:

Kathryn H. Anderson, Ph.D.

Ron Zimmer, Ph.D.

Peter Savelyev, Ph.D.

Tong Li, Ph.D.

To my son, Thomas Hong

ACKNOWLEDGMENTS

I thank my advisor, Professor Kathryn Anderson, for her infinite guidance on this project. She taught me how to do research as an economist. Her patience and support helped me overcome many difficulties and finish this dissertation. I also extend profound thanks to Professor Ron Zimmer, Professor Peter Savelyev, and Professor Tong Li for their individual contributions of time and effort. Their support make this dissertation possible.

I would like to thank my father, Ruiting Hong, and mother, Yanhua Sun, whose love and encouragement are always with me. To my loved wife, Na Liu, I extend profound thanks for her support and care, which help me stay focused on my study.

Finally, I appreciate the financial support from Vanderbilt University, NSF and the Beijing Planning Office of Philosophy and Social Science that funded parts of the research discussed in this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter I

School Bond Referendum, Capital Expenditure, and Student Achievement

I.1 Introduction

Capital expenditures on infrastructure for schools in the United States are financed through voter-approved bonds issued by each local district in many states, including Michigan, California and Texas.[1] This funding system for schools is under debate because the disparity in fiscal capacity of local districts leads to inequality in capital investment across districts (Pratt et al., 2012). People hope to decrease inequality in student achievement across districts by eliminating inequality in capital investment through the reform of the school financing system. Alternative systems of finance such as matching grant program are under discussion in some states, such as Michigan. However, the alternative systems can work as expected only if capital investment impacts student achievement. The discussion about reform benefits from the answer to the question addressed in this paper: what is the average effect of capital expenditure financed by a passed bond on student achievement?

Capital expenditures on infrastructure represent a significant educational investment in the United States. Each year since 2000 as much as $60-70 billion (constant 2012-2013 dollars) was invested in public school construction and repairs (National Center for Education Statistics, 2015). In urban districts, billions of dollars were invested to improve school facilities in cities such as Los Angeles, Cleveland and Dallas (Dejong and Glover, 2003). Despite the substantial investment, little is known about the effect of capital expenditure on educational outcomes. Many students are educated in old and programmatically inadequate school buildings, especially in urban districts (Dejong and Glover, 2003). Old school buildings may have problems such as poor indoor air quality, inadequate lighting and acoustics,

[1] The financing system for capital expenditure varies by state. There are only nine states and Washington, D.C that do not require elections for school bonds for new projects. For a detailed description see http://ballotpedia.org/Voting_on_school_bond_and_tax_measures.

which are negatively correlated with student performance (Schneider, 2002).

Much of the early research on the impact of capital expenditures, which only emerged in the past 15 years, did not fully account for endogenous expenditures (Blincoe, 2009; Jones and Zimmer, 2001; Picus et al., 2005). Several recent papers, including Cellini et al. (2010), Martorell et al. (2015) and Hong and Zimmer (2015), employ a more rigorous identification strategy — the regression discontinuity design (RDD) — to examine outcomes for districts in which capital expenditures narrowly are approved relative to districts in which capital expenditures narrowly fail.[2] The existing studies portray a mixed picture of the effect of capital expenditure on student achievement; some studies find positive effects (Hong and Zimmer, 2015; Nielson and Zimmerman, 2014; Welsh et al., 2012), and others show zero to modest effects (Cellini et al., 2010; Martorell et al., 2015).

The RDD approach solves the issue of endogenous bond passage with strong internal validity (Lee and Lemieux, 2010), but the external validity of the RDD approach is limited. Without additional assumptions, we can only identify the causal effect for a bond that is marginally passed with a vote share that is precisely at the cutoff (Hahn et al., 2001; Imbens and Lemieux, 2008; Lee and Lemieux, 2010).[3] This so-called weak external validity feature of the RDD approach prevents us from making broader inferences for districts that are far from the cutoff.

From a policy viewpoint, it is important to know the average treatment effects in addition to the local effect identified by a standard RDD (Hainmueller et al., 2014). Estimation of the average effect helps resolve the inconsistency of the estimated effects in the existing literature, which may be the result of using different cutoffs in RDD to estimate effects. Policy makers are more interested in the average effect than the local effect because they

[2]In another strand of literature employing difference-in-differences (DID) approaches, Nielson and Zimmerman (2014) and Welsh et al. (2012) examine the impact of large-scale construction projects of new schools in New Haven and Los Angeles.

[3]The papers using DID focus on large-scale projects to replace old buildings. The authors provide only limited implications for smaller-scale projects, including maintenance and additions. These DID studies examine how capital investment works in either a poor urban district (Nielson and Zimmerman, 2014) or a large metropolitan setting (Welsh et al., 2012). It is unclear if their results are applicable in other cases.

need to know the effect for all students, districts, and bonds, not just students who attend districts with vote shares close to the cutoff (Wing and Cook, 2013).

I adopt a latent factor model to estimate the average effect of bonds for all districts. Latent factor models are established in economics to identify the joint distribution of potential outcomes (Cunha et al., 2005; Heckman and Vytlacil, 2005) and underlying latent personality (Heckman et al., 2013a; Hong et al., 2014; Savelyev, 2014). Rokkanen (2015) adopts a latent factor model to extrapolate the effect of exam schools on students with test scores away from the cutoff. Other extrapolation approaches are also developed; Angrist and Rokkanen (2015) use the conditional independence assumption, and Dong and Lewbel (2015) identify the derivative of the treatment effect at the cutoff. I use the latent factor model because the other approaches either impose stronger assumptions (Angrist and Rokkanen, 2015) or only extrapolate the effect to a small neighborhood near the cutoff (Dong and Lewbel, 2015).[4]

One advantage of my approach is that it allows me to predict the effect of capital expenditure in different environments, such as a newly designed bond referendum system or a completely different financing system for capital investment. Heckman and Vytlacil (2007) show that adjusting causal effects to a new environment is beyond the traditional treatment effect approach, in which the mechanisms underlying causal effects are not explicitly modeled. Without knowing the underlying mechanisms, the only way to obtain the effect for a different environment is to re-estimate the model when the data from the new environment are available, and the new policy has been implemented. My model of the average effect helps me predict the effect on student achievement under a hypothetical financing system – a matching grant program – before it has been implemented.

I use data from the state of Michigan and find some evidence of positive average effects of bond passage on 4th grade reading proficiency. I examine the effectiveness and

[4]There are some other extrapolation approaches that are not applicable to this paper because of the data limitation or the current context of education. Examples include Wing and Cook (2013) using pretest data and Bargain and Doorley (2013) who explore the structure of the utility function.

efficiency of the current capital financing system using bond referenda, and a hypothetical alternative system – a matching grant program. I show that although both the current and the hypothetical alternative systems have the potential to improve student achievement with positive pecuniary net benefits under certain reasonable scenarios, they are unlikely to be the most cost-effective investments that to improve student achievement.

I.2 Institutional Background and Data Description

I use Michigan bond election data from 1996 to 2009, a 14-year panel data recorded by the Michigan Department of Treasury. The data include 1,265 bond elections in 549 local districts. The election data are matched with student achievement data from the Michigan Department of Education and the common core data from the National Center for Education Statistics (NCES).[5]

The School Bond Qualification and Loan Program (SBQLP) in Michigan was created in 1955 by constitutional amendment. The Michigan Constitution of 1963 established the parameters for the program. Most of the parameters are still in place with modifications from a series of public acts. The initial purpose of SBQLP was to assist local school districts with construction of facilities for growth in the student population due to the baby boom (Pratt et al., 2012). The program today still helps local districts with debt service for school construction. School districts rely on a local referendum to approve the issuance of a bond for capital expenditures. The whole process involves six steps (Michigan Department of Treasury, 2015): Strategic Planning and Defining Project Scope, Preliminary Qualification Approval, Election, Final Qualification Approval and Issuance of Bond, Construction, and Audit of Bonded Construction Funds.

For strategic planning and defining project scope, the school district identifies the need for capital expenditure, construction priorities, potential regulations and restrictions, and

[5] For the election data see https://treas-secure.state.mi.us/apps/findschoolbondelectinfo.asp. For the student achievement data see http://www.michigan.gov/mde/0,1607,7-140-22709_31168_31530---,00.html. For the common core data see http://nces.ed.gov/ccd/bat/.

forms a preliminary plan that includes a detailed budget. The plan, cost and budget are approved first by the local board of education then by the Michigan Department of Treasury. After the preliminary qualification approval, the district holds a local election for the qualified bond proposal.[6] If the majority of the local electorate approves the bond, the district applies for final qualification and issues the bond upon approval. The type of construction is restricted to a number of specific categories such as new buildings, remodeling, land purchase, athletic and physical education facilities and school buses.[7] An independent audit is required after all projects are finished.

These referenda are summarized across years in Figure I.1, which provides general trends in election results. The number of proposed measures decreased substantially from 1996 to 2009, which may be the result of the decline in Michigan′s economy since the early 2000s. We do not observe a clear time trend in percentage of passage. 433 districts (75% of the total amount of districts) held at least one bond election between 1996 and 2009. On average a district held three elections.[8] Elections were quite close in terms of vote share; about 30.5% of the elections were decided by less than 5% of the vote, and 57.9% were decided by less than 10% of the vote. Details of descriptive statistics of bond characteristics are shown in Table A.1 in the Appendix.

I have detailed information on district expenditures, demographics, and student achievement (Table A.2 in the Appendix). The measures of academic achievement are 4th and 7th grade district-level reading proficiency rates.[9] Proficiency rate is the metric by which

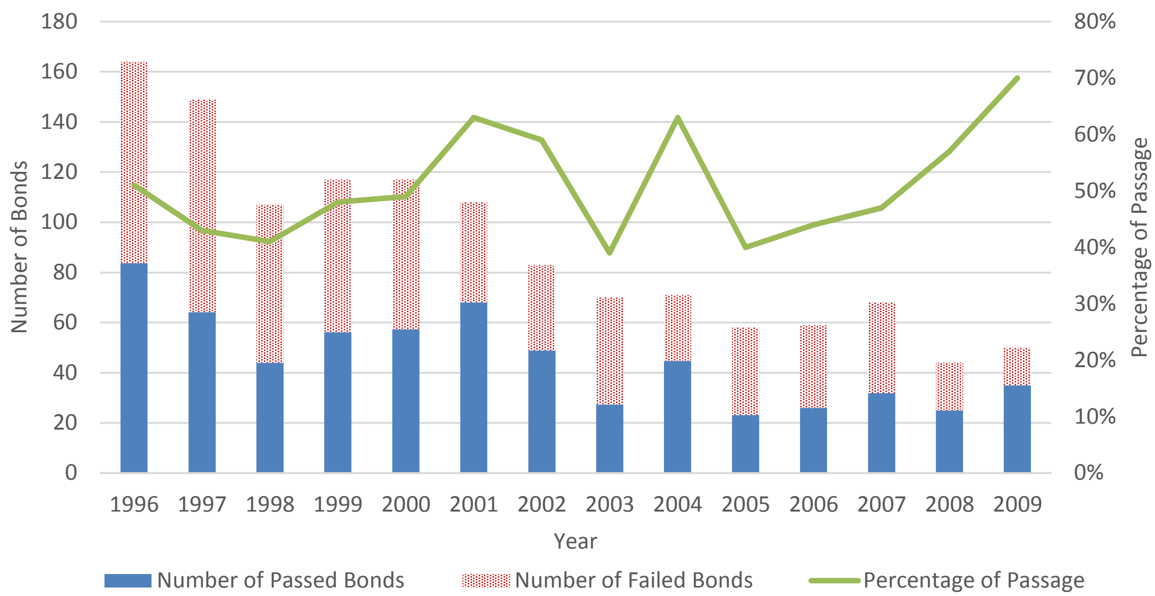
[6]The elections must be held on a certain day, which is generally one of the four regular election dates: 4th Tuesday in February, 1st Tuesday after the 1st Monday in May, 1st Tuesday after the 1st Monday in August, and 1st Tuesday after the 1st Monday in November. A limited number of "floater dates" can also be chosen if certain requirements are satisfied.

[7]Others include energy conservation improvement, asbestos abatement, development and improvement of site and playground, loose furnishings and equipment purchasing, computer hardware and direct bond program cost.

[8]There are 127 districts holding one election, 97 districts holding two elections, 84 districts holding three elections and 125 districts holding more than three elections. The largest amount of elections held by a district is 14.

[9]I do not have a consistent measure of math proficiency over the entire time horizon, and therefore, I focus exclusively on reading proficiency. Nielson and Zimmerman (2014) find positive effects of capital investment on reading scores but no effect on math scores.

Figure I.1: School Bond Referendum in Michigan, Years 1996–2009



Notes: The sample includes 1,265 bonds with non-missing values in both passage and vote share.

schools and districts are held accountable under No Child Left Behind. Michigan requires all public school students in grades 3–9 to take the Michigan Educational Assessment Program (MEAP) each year. The MEAP is a standardized test covering various subjects. The state sets score cutoffs to separate students into categories of performance: advanced, proficient, partially proficient and not proficient. The district-level proficiency rate is the percentage of students in the district who are advanced or proficient.[10]

In Table I.1, I describe observable district characteristics at baseline, which is one year prior to a bond election.[11] I first examine all districts (Column (1)) and then compare districts in which bonds passed and failed. Columns (2) and (3) provide means of observable

[10]For details of MEAP and proficiency, see http://www.michigan.gov/mde/0,4615,7-140-22709_70117_40135---,00.html. Since spring 2015, MEAP was replaced by the Michigan Student Test of Educational Progress (M-STEP). For M-STEP see http://www.michigan.gov/mde/0,4615,7-140-22709_70117---,00.html.

[11]Districts in Michigan are allowed to propose multiple elections in the same year. As Cellini et al. (2010), I keep the election with the highest vote share only in each year. This gives us a deterministic relationship between vote share and bond passage. In total there are 936 elections (74% of the total number of elections) with non-missing vote share after dropping multiple elections in each year.

Table I.1: District Descriptive Statistics in Year Prior to Elections

	All elections (1)		Passed elections (2)		Failed elections (3)		Difference (2)-(3), t test (4)	
	mean	standard deviation	mean	standard deviation	mean	standard deviation	difference in mean	*p* -value
Expenditure per pupil[(a)]								
Total	7872	1664	8092	1783	7547	1414	544.9	0.000
Current	6857	1093	6930	1167	6749	965.8	180.9	0.020
Teacher salary	2929	513.3	2974	528.4	2863	483.5	110.8	0.002
Capital[(b)]	317.8	930.4	405.5	1030	188.9	743.1	216.6	0.001
Construction	256.4	866.8	319.3	941.0	163.9	736.1	155.3	0.011
Land and structure	61.47	316.0	86.28	401.8	24.99	85.34	61.29	0.006
Instructional equipment[(c)]	53.02	63.83	50.54	66.17	56.68	60.14	-6.140	0.175
Demography								
Enrollment	3167	5044	3478	5917	2710	3330	768.3	0.032
White students[(d)]	89.18	15.49	89.27	15.71	89.05	15.18	0.217	0.844
Free lunch[(d)(e)]	29.04	16.47	26.62	16.99	32.47	15.07	-5.846	0.000
Achievement (proficiency)[(d)]								
4th grade reading	65.31	18.55	67.72	17.64	61.83	19.30	5.893	0.000
7th grade reading	56.41	18.06	58.49	17.52	53.40	18.44	5.090	0.000
Sample size	825		491		334		825	

Notes: [(a)]All expenditures per pupil are measured in constant year 2000 dollars. [(b)]Capital is defined as the sum of construction, land, and structure. [(c)]Instruction equipment includes expenditures for all instructional equipment recorded in general and operating funds under "instruction". [(d)]White students, free lunch and proficiency are measured in percentage point. [(e)]A student can receive free lunch if the households income is within the limits on the Federal Income Eligibility Guidelines or one of several other conditions is satisfied. For details see http://www.michigan.gov/mde/0,4615,7-140-66254_50144-194552--,00.html.

district characteristics one year prior to a bond passing and failing, respectively. Column (4) compares the means of the two samples with a t-test. In general, the results in Columns (2), (3) and (4) suggest that districts that passed an election are more advantaged in most baseline characteristics, which indicates that a simple estimate (e.g., OLS estimate) of the differences in outcomes across these two populations in which bonds pass and fail is likely to be biased.

I.3 Methodology

I first briefly discuss the relationship between bond passage and student achievement within the RDD framework. Then I discuss how to estimate the average effect of bond passage using a latent factor model. I show that, after controlling for the underlying preferences for educational investment, I can estimate the average effect of bonds with any vote share under several regular assumptions.

$Y_{jt\tau}$ denotes the proficiency rate in school district j at time $t+\tau$, which is modeled as a function of one previous bond passage τ years ago at time t:

$$Y_{jt\tau} = \beta_\tau B_{jt} + e_{jt\tau}, \quad \text{for all } \tau \text{ when there was a bond election.} \tag{I.1}$$

τ is the time gap between the election and the year when the proficiency is measured; bonds passed at time t have differential effects on $Y_{jt\tau}$ based on the time between bond passage and the outcomes. A dummy variable B_{jt} indicates whether or not the bond is passed at time t, and $e_{jt\tau}$ captures other factors affecting the outcome at time $t+\tau$. Bond passage is endogenous; B_{jt} is correlated with $e_{jt\tau}$. The electorate's choice determines if a bond is approved; the choice largely depends on underlying variables such as voters' preferences for educational investment. In previous literature an RDD approach is adopted to fully account for the endogenous bond passage; the literature compares outcomes of districts in which capital expenditures are marginally approved with districts in which capital expenditures marginally fail. In principle, districts that marginally pass a bond should be similar to those that marginally fail a bond in both observed and unobserved ways. By such a comparison authors obtain the effect of a bond that passes with a vote share at the cutoff. Specifically, we estimate the following equation:

$$Y_{jt\tau} = \sum_{\tau=-2}^{12} (\beta_\tau g^\tau B_{jt} + \omega_\tau g^\tau f(v_{jt}, \gamma_{jt})) + G_\tau + T_{t\tau} + E_{jt} + e_{jt\tau},\ 1996 \leq t \leq 2009, \tag{I.2}$$

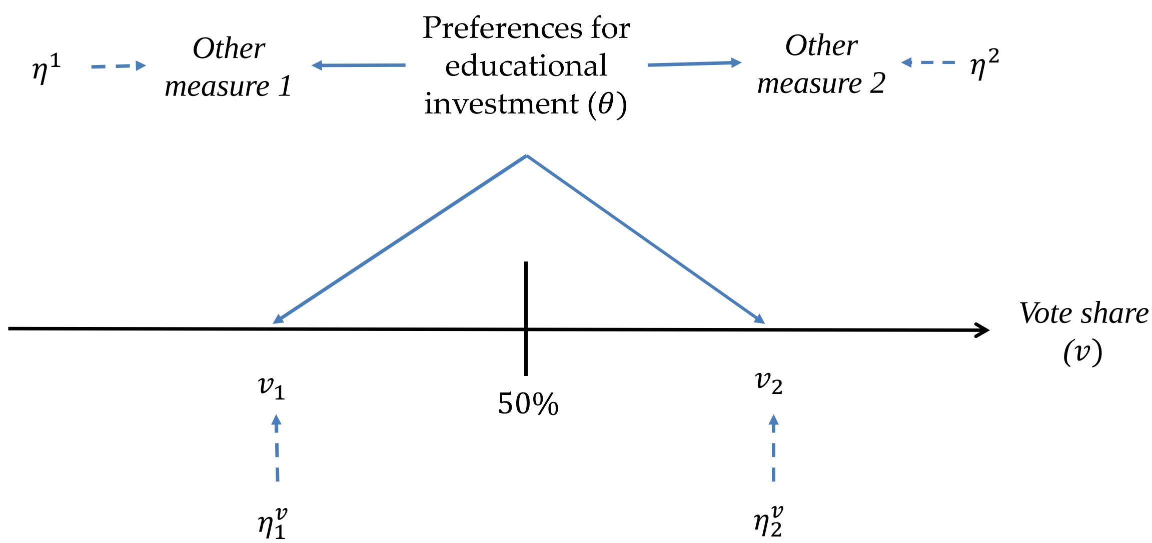
where $e_{jt\tau}$ and B_{jt} are uncorrelated. g^{τ} is a dummy variable indicating year gap τ. β_{τ} is the coefficient on the interaction between g^{τ} and B_{jt} and estimates the gap-specific intent-to-treat (ITT) effect of the bond passed τ years ago.[12] v_{jt} is vote share of the election held in district j at time t. $f(\cdot)$ is a polynomial function of v_{jt} with parameter γ_{jt}. ω_{τ} is the coefficient on the interaction between g^{τ} and $f(v_{jt}, \gamma_{jt})$. G_{τ} is the gap fixed effect, $T_{t\tau}$ is the calendar year fixed effect and E_{jt} is the election fixed effect, which also incorporates the district fixed effect. The ITT effect incorporates the effects of passing a bond on the subsequent bond passages which also affect proficiency. The ITT analysis provides policy insight into the impact of successful passage of an initial referendum, without isolating the impacts through the subsequent bonds.

The RDD approach uses vote share as the forcing variable. The underlying assumption of RDD is that districts with similar vote share have similar preferences for educational investment. As there are some unanticipated random components in vote, districts with similar preferences for educational investment have different election results. Vote share, as well as other variables such as pre-existing student achievement, represent underlying preferences for educational investment with noise.

An example in Figure I.2 illustrates the relationship between preferences for educational investment and bond passage. Vote share (v) and two other measures (measures 1 and 2) represent latent preferences for educational investment (θ) with noises η^{v}, η^{1} and η^{2}, respectively. Consider two districts with the same preferences for educational investment, but different realized noises about vote share (η_1^{v} and η_2^{v}). If there is no noise, the districts have the same election result. However, because of the noise in vote share, district 1 receives a vote share below the cutoff (50%) and district 2 receives a vote share above the cutoff. We observe both passed and failed bonds for the same underlying preferences for educational investment. The noise might be, for example, the weather on election day which affects voter turnout (Lee, 2008).

[12]As Cellini et al. (2010), we assume that the effects on subsequent outcomes depend only on the time gap, not on the time of election or election history.

Figure I.2: Preferences for Educational Investment and Bond Passage



Let the potential outcomes (proficiency rate) be Y^0 with failed bond and Y^1 with passed bond, which depend on preferences for educational investment but do not depend on the noise in vote share. Y^0 and Y^1 are independent of vote share conditional on preferences for educational investment. Formally I have the following conditional independence assumption:

$$(Y^0, Y^1) \perp\!\!\!\perp v|\theta \tag{I.3}$$

If I am able to identify the latent preferences for educational investment and the conditional distribution of θ given vote share, I can estimate the conditional average treatment effect $E((Y^1 - Y^0)|\theta)$ and the average treatment effect at any vote share $E((Y^1 - Y^0)|V = v) = E((Y^1 - Y^0)|\theta = E(\theta|V = v))$.

Latent factor models are designed to obtain a small number of latent factors linked to multiple noisy and correlated observable variables (Bollen, 1989). In my latent factor model I use pre-existing student achievement and election result to identify underlying preferences for educational investment.[13] The average effect of bond passage on student achievement can be obtained by controlling for the uncovered underlying preferences for educational investment as a potential confounder.[14]

To account for confounding variables as determinants of the bond election result and student achievement, I link the latent factor of preferences with observed measures P and estimate the following latent factor model:

$$P_{l\,jt} = \psi_l \Theta^P_{jt} + \psi_0 + \eta_{l\,jt},\ l = 1,2,3,\ 1996 \leq t \leq 2009, \tag{I.4}$$

where the subscript jt indicates that the corresponding variable is measured in school district j at time t. $P_{l\,jt}$ are measures that proxy latent school district's preferences on financing education Θ^P_{jt}: vote share, $P_{1\,jt}$; 4th grade reading proficiency, $P_{2\,jt}$; and 7th grade reading proficiency, $P_{3\,jt}$. ψ_l is the factor loading for the observed measure P_l, which represents the relationships between latent factor Θ^P and preference measure P_l. ψ_0 is an intercept, and η_l denotes measurement error.

I make standard assumptions for factor model identification and clear interpretation of results (Anderson and Rubin, 1956): $\eta_{l\,jt} \perp\!\!\!\perp \eta_{l'\,jt}$, for $l \neq l'$; $\eta_{l\,jt} \perp\!\!\!\perp \Theta^P_{jt}$ for all l; and standard normalizations ($\mathrm{Var}(\Theta^P_{jt}) = 1$).[15] Finally I assign ψ_1 to be positive without loss of generality.[16]

I identify the average effect of bond passage by the following conditional independence

[13] There may be other confounding variables related to preferences that influence educational outcomes. These variables include household income, investment from other sources, and student ability. Because of the limitations of the data set, I am not able to include all of these other confounding variables into the analysis. Instead, I identify a latent factor using multiple achievement measures. The intuition is that pre-existing achievement captures those potential confounding variables.

[14] The latent factor approach also controls for measurement error in measures of preferences.

[15] I set the mean and variance of preferences for educational investment to zero and one respectively to pin down the location and scale of the latent factor. An alternative is to set the loading of the first measure (ψ_1) to a constant such as one. The two normalization approaches produce the same model fit.

[16] An alternative is to assign ψ_1 to be negative. Then the identified latent factor is interpreted as the opposite of the current identified factor.

assumptions:

$$(Y^0_{jt\tau}, Y^1_{jt\tau}) \perp\!\!\!\perp P_{ljt}|\Theta^P_{jt},\ l = 1,2,3,\ 1996 \leq t \leq 2009, \tag{I.5}$$

where $Y^0_{jt\tau}$ is the outcome τ years later if B_{jt} is not passed and $Y^1_{jt\tau}$ is the outcome τ years later if B_{jt} is passed. Θ^P_{jt} is latent preferences for educational investment in school district j at time t. This exclusion restriction assumes that potential outcomes $(Y^0_{jt\tau}, Y^1_{jt\tau})$ are correlated with the preference measures P_{ljt} through only the latent preferences for educational investment Θ^P_{jt}.[17] I estimate the average ITT effect of bond passage through the following equation:

$$Y_{jt\tau} = \sum_{\tau=-2}^{12} (\beta_\tau g^\tau B_{jt} + \omega^P_\tau g^\tau \Theta^P_{jt}) + G_\tau + T_{t\tau} + E_{jt} + e_{jt\tau},\ 1996 \leq t \leq 2009, \tag{I.6}$$

where Θ^P_{jt} is identified in the latent factor model (I.4).[18] ω^P_τ is the coefficient of the interaction between g^τ and Θ^P_{jt}. All other parameters are the same as in Equation (I.2). For the ITT estimation I use the stacked sample as defined in Cellini et al. (2010).[19]

An initial bond passage may also affect subsequent bonds. A part of the ITT effect β_τ may affect later student achievement through the impact on subsequent bond passage. I identify the average effect of bond passage controlling for subsequent bonds, through the following one-step TOT estimation, which is an analogue of the RDD one-step estimation in Cellini et al. (2010):

$$Y_{jt} = \sum_{\tau}^{\bar{\tau}} (\alpha_\tau M_{jt-\tau} + \beta_\tau B_{jt-\tau} + \omega_\tau \Theta^P_{jt-\tau}) + D_j + T_t + \mu_{jt},\ 1996 \leq t \leq 2009, \tag{I.7}$$

where $\bar{\tau}$ is the largest possible lag, which is 0 for $t = 1996$, 1 for $t = 1997$ and so on. $M_{jt-\tau}$

[17]The conditional independence assumption (I.5) implies $(Y^0_{jt\tau}, Y^1_{jt\tau}) \perp\!\!\!\perp B_{jt}|\Theta^P_{jt}$.

[18]As Cellini et al. (2010), I assume that a bond passage does not affect pre-existing outcomes; β_{-2}, β_{-1}, β_0, ω_{-2}, ω_{-1} and ω_0 are restricted to zero.

[19]To generate the stacked sample, I first "stack" all district-year observations for the district that has an election in year t in a window from $t-2$ through $t+13$, for each election. Second, the stacked data sets for each separate election are combined into one large panel data set.

is a dummy variable with coefficient α_τ and equals to one if district j held a bond election τ years before time t. $B_{jt-\tau}$ indicates whether district j passed a bond election τ years ago. β_τ is identified as the total treatment effect on the treated (TOT). D_j is the district fixed effect and T_t is the calendar year fixed effect. μ_{jt} is the error term. β_τ provides insight on the impact of bond authorization by isolating the subsequent changes in the district's capital expenditure. For the TOT estimation I use a standard panel data set.

I estimate the models with the following three-stage procedure (Heckman et al., 2013a). First I estimate the measurement system for the latent preferences for educational investment (Equation (I.4)). Second I estimate the preferences for educational investment for each observation of district-election. Finally, I estimate the outcome equation – Equation (I.6) or (I.7) – with the estimated preferences for educational investment. I calculate standard errors using the bootstrap with 500 draws.

I.4 Empirical Results

I.4.1 Latent Factor Model

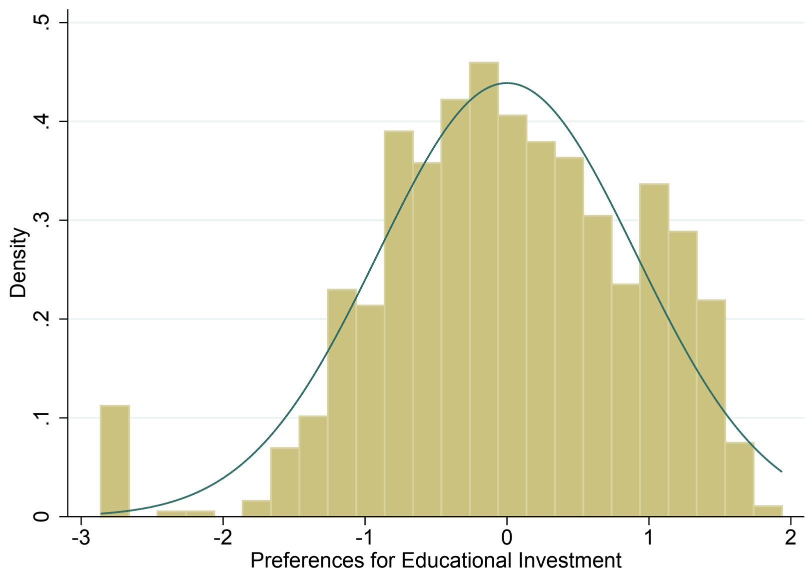
Table I.2: Factor Loadings on the Means of the Observed Measures

	Vote share	Reading proficiency	
		4th grade	7th grade
Preferences for educational investment	0.140 ***	0.841 ***	0.969 ***
	(0.009)	*(0.019)*	*(0.022)*

Notes: The table presents the estimated factor loadings on the means of the observed measures of latent preferences for educational investment. The specification is the latent factor model described in Equation (I.4). Standard errors are shown in parentheses. The sample size is 936. *** indicates significance at the 1% level.

Table I.2 shows the estimated factor loadings on the means of vote share, 4th grade reading proficiency and 7th grade reading proficiency (ψ_1, ψ_2 and ψ_3, respectively). All of the factor loadings are statistically significant. The latent preferences for educational investment positively correlate with vote share and pre-existing proficiency. The latent preferences for educational investment affects 4th and 7th grade proficiency in a similar way, with a larger factor loading on 7th grade reading proficiency. In spite of different grades (perhaps schools, too), students in the same district have many educational inputs in common, such as the district's policies and services.

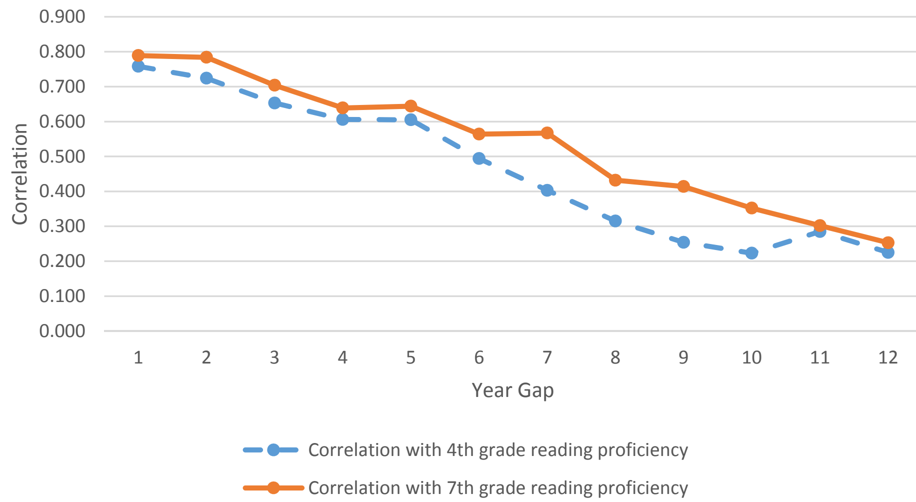
Figure I.3: Density of Preferences for Educational Investment



Notes: The histogram presents the density of the estimated preferences for educational investment. The width of each bin is 0.2. The curve represents an appropriately scaled normal density with the same mean and standard deviation as the data.

Figure I.3 shows the density of the uncovered preferences for educational investment. Overall the distribution of preferences for educational investment is close to a standard normal distribution. The mean and standard deviation are 0 and 0.91. Figure I.4 presents the correlations between the latent preferences for educational investment and subsequent proficiency. The preferences for educational investment is highly correlated with the subsequent proficiency in the short run, and this correlation declines over time.

Figure I.4: Correlation Between Preferences for Educational Investment and Subsequent Reading Proficiency



Notes: The figure presents the correlation between the estimated preferences for educational investment and subsequent reading proficiency of grade 4 and 7.

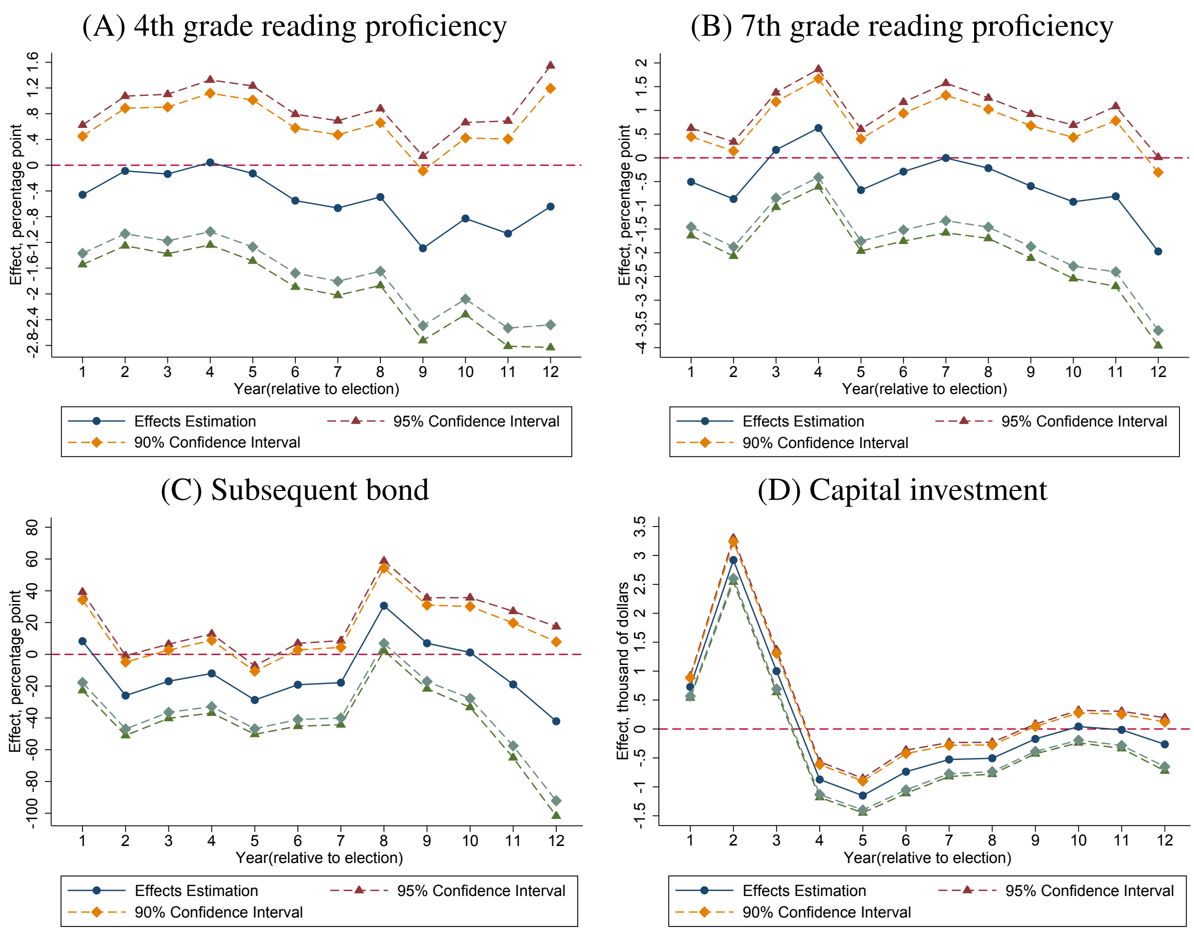
I.4.2 Average Effect of Passing a Bond on Proficiency

For the whole population of districts I estimate the average effects of bond passage on proficiency, as well as various mediating outcomes, including subsequent bond passage and capital expenditures. The results show modest evidence that a passed bond can increase reading proficiency on average, especially for the 4th grade.

Figure I.5 presents the estimated average ITT effects from the latent factor model (see Columns (1) and (2) of Table A.3 in the Appendix for details.). The ITT effects estimated by the latent factor model (Equation (I.6)) show little evidence that bond passage affects subsequent proficiency (Panels (A) and (B)). These effects may understate the true effect of bond passage because an initial bond passage decreases the possibility of passing subsequent bonds. Alternatively, we do not observe any effect because an initial bond does not increase subsequent expenditures on capital.

Passing a bond can potentially affect various subsequent outcomes. Figure I.5 also shows how passing an initial bond affects the probability of passing subsequent bonds and

Figure I.5: ITT Effect of Bond Passage on Achievement, Subsequent Bond Passage and Capital Investment



Notes: The figure shows the coefficients, 95% and 90% confidence interval for ITT effects of bond passage on subsequent 4th grade reading proficiency(Panel (A)), 7th grade reading proficiency (Panel (B)), subsequent bond passage (Panel (C)) and subsequent capital investment (Panel (D)). The specification is the ITT regression described in Equation (I.6). The sample sizes are 9,665, 9,663, 2,665 and 9,829, respectively.

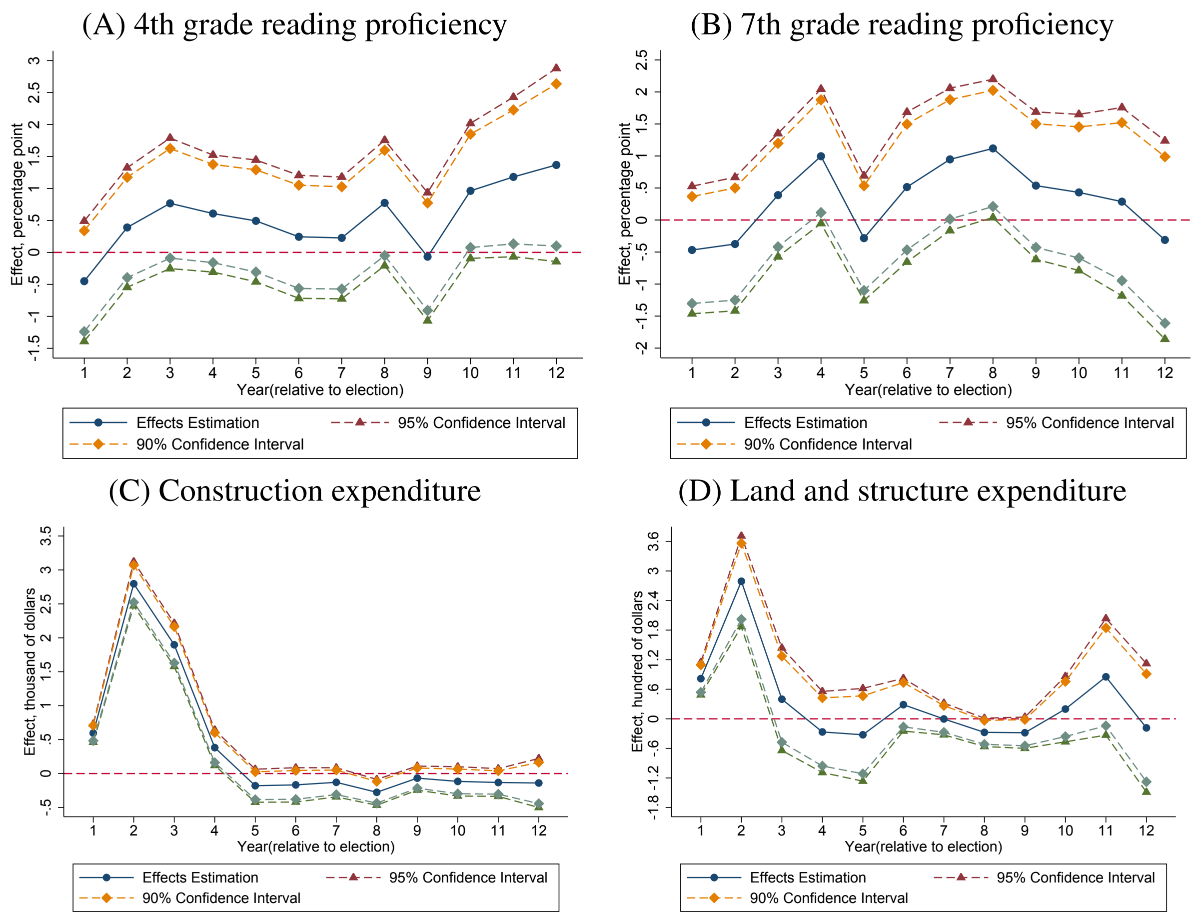
total capital expenditures on construction, land and structure (see Columns (3) and (4) of Table A.3 in the Appendix for details.). The results regarding subsequent bonds in Panel (C) are consistent with the findings of Cellini et al. (2010) in California and Martorell et al. (2015) in Texas. Passing an initial bond decreases the probability of passing another bond in the short term — two to five years — by about 20–30%, although the effects are not always significant. There is no clear long-term effect on the probability of passing subsequent bonds, except for a positive effect of 0.3 emerging eight years later. The cumulative short-run effect of an initial bond passage in five years is about -0.55, and the cumulative long-run effect of an initial bond passage in 12 years is about -0.25. These results indicate that on average failing an initial bond election reduces the expected total number of passed bonds by 0.45 in the short run and 0.75 in the long run.

A bond passage significantly increases total capital expenditure as well as expenditure on construction, land and structure in the short run (Panel (D)). Total capital expenditure starts to increase in the year of bond passage, and peaks after 2 years, when the maximum expenditure on construction per pupil is about $3,000 higher than the expenditure in the districts that fail a bond. The effects start to decline in the third year and become negative after the fourth year because of the short-run negative effects on subsequent bonds. The effects diminish after 9 years. The effects on capital investment confirm the findings in Panel (C) about subsequent bonds. In the middle term a passed initial bond decreases capital expenditure through its negative impact on subsequent bond passage in the short run.

Controlling for the impact on subsequent bond passage, the TOT effects in Figure I.6 show that a passed bond marginally increases reading proficiency in 4th grade by 1.0–2.0 percentage points after 10 years (Panel (A)). For 7th grade there is no consistent effect (Panel (B)), with some evidence of a positive effect after 4 and 8 years. As I will discuss later, the estimation for 7th grade reading proficiency is less reliable. Details of TOT effects can be found in Columns (1) and (2) of Table A.4 in the Appendix. While the ITT estimation suggests that passing a bond does not improve student performance because passing an initial bond reduces the likelihood of passing a later bond, the TOT estimation shows that the actual capital expenditure generated from an individual passed bond positively affects long-term 4th grade reading proficiency.

The positive effect of passing a bond on subsequent achievement is likely due to more total expenditure on either construction or land acquisition or both. To understand the role

Figure I.6: TOT Effect of Bond Passage on Achievement, Construction, Land and Structure Expenditure



Notes: The figure shows the coefficients, 95% and 90% confidence interval for TOT effects of bond passage on subsequent 4th grade reading proficiency(Panel (A)), 7th grade reading proficiency (Panel (B)), subsequent expenditure on construction (Panel (C)) and land and structure (Panel (D)). The specification is the TOT regression described in Equation (I.7). The sample sizes are 7,244, 7,219, 7,746 and 7,746, respectively.

of increased capital investment in explaining the observed positive effects on achievement, I estimate a two-step model. In the first step I estimate the TOT effect of bond passage on subsequent capital investment on construction or land and structure by Equation (I.7) using capital investment as the dependent variable. In the second step I estimate the TOT effect of bond passage on subsequent achievement by Equation (I.7). Specifically, I estimate the following two-step model:

$$\begin{cases} E_{jt} = \sum_{\tau}^{\bar{\tau}}(\alpha_{\tau}^{E} M_{jt-\tau} + \beta_{\tau}^{E} B_{jt-\tau} + \omega_{\tau}^{E} \Theta_{jt-\tau}^{P}) + D_{j}^{E} + T_{t}^{E} + \mu_{jt}^{E},\ 1996 \leq t \leq 2009, \\ Y_{jt} = \sum_{\tau}^{\bar{\tau}}(\alpha_{\tau}^{Y} M_{jt-\tau} + \beta_{\tau}^{Y} B_{jt-\tau} + \omega_{\tau}^{Y} \Theta_{jt-\tau}^{P}) + D_{j}^{Y} + T_{t}^{Y} + \mu_{jt}^{Y},\ 1996 \leq t \leq 2009, \end{cases} \tag{I.8}$$

where E_{jt} denotes the capital investment on construction or land and structure in district j at time t. All other parameters are defined in the same way as Equation (I.7), with superscript Y and E referring to equations for reading proficiency and capital investment respectively. Figure I.6 shows the TOT effects of bond passage on subsequent expenditure on construction (Panel (C)) and land and structure (Panel (D)). Passing a bond increases construction capital in the first four years and land capital in the first two years. Therefore, I use the sums of construction capital and land capital during those periods as the treatment variable, and estimate the average effect of the increased construction capital and land capital over four years and two years on achievement in subsequent years after the construction/land acquisition is complete. Specifically, the effect of construction or land acquisition that is finished after ρ years on reading proficiency τ years later, is expressed as:

$$\gamma_{\tau} = \frac{\beta_{\tau}^{Y}}{\sum_{\rho=1}^{\bar{\rho}} \beta_{\rho}^{E}},\ \bar{\rho} + 1 \leq \tau \leq 12, \tag{I.9}$$

where $\bar{\rho} = 4$ for construction capital and $\bar{\rho} = 2$ for land capital.[20]

Table I.3 shows that an increase of $\$1,000$ of construction expenditure per pupil increases 4th and 7th grade proficiency by 0.15–0.25 and 0.15–0.2 percentage points respectively. The results for land capital are larger. An increase of $\$1,000$ of land and structure expenditure per pupil increases 4th and 7th grade proficiency by 2.5–3.5 and 2.5–3 percentage points respectively.

[20] Effect (I.9) can be interpreted as the local average treatment effect (LATE) for the compliers, which are districts that are induced to finish construction or land acquisition by passing a bond.

Table I.3: Effect of Capital Investment on Achievement

	Construction		Land and Structure	
Relative year	4th grade	7th grade	4th grade	7th grade
3	-	-	2.129	1.078
			(1.451)	*(1.557)*
4	-	-	1.686	2.761 *
			(1.289)	*(1.651)*
5	0.087	-0.050	1.366	-0.784
	(0.085)	*(0.094)*	*(1.371)*	*(1.500)*
6	0.043	0.091	0.676	1.429
	(0.096)	*(0.100)*	*(1.494)*	*(1.619)*
7	0.040	0.167 *	0.629	2.624 *
	(0.087)	*(0.094)*	*(1.402)*	*(1.519)*
8	0.137	0.197 **	2.147	3.098 **
	(0.083)	*(0.093)*	*(1.326)*	*(1.501)*
9	-0.012	0.095	-0.186	1.488
	(0.084)	*(0.095)*	*(1.350)*	*(1.517)*
10	0.170 *	0.076	2.670 *	1.194
	(0.098)	*(0.111)*	*(1.543)*	*(1.814)*
11	0.208 *	0.051	3.274 *	0.797
	(0.123)	*(0.125)*	*(1.951)*	*(1.963)*
12	0.241 *	-0.055	3.793	-0.862
	(0.144)	*(0.126)*	*(2.329)*	*(2.033)*

Notes: The table shows the effects of finished capital investment on construction and land and structure on subsequent reading proficiency. It is assumed that unfinished construction or land and structure has no effect on student achievement. The specification is the TOT regression described in Equation (I.8). Clustered standard errors by school district are shown in parentheses. * and ** indicate the statistical significance at the 10% and 5% levels, respectively. The sample sizes are 7,746 for all estimations.

I.4.3 Validity of Latent Factor Model

In the literature a standard validity test of an RDD approach is to check if the observed pre-determined variables are all balanced around the cutoff. I first use a similar idea to test the validity of the latent factor model by checking if the bond passage has any effect on pre-determined variables after I control for the latent preferences for educational investment. Second, I estimate the effects of bond passage on reading proficiency at the cutoff using the latent factor model. Then I compare the estimated effects with the effects from a standard RDD model. If the latent factor model is valid, the two models should show similar results.

The comparison indicates that the latent factor model is a valid approach for my data, especially for 4th grading reading proficiency. Finally, I compare the estimation results from alternative specifications controlling for different potential confounders and find that controlling for latent preferences for educational investment changes the estimated effects.

To examine whether districts that pass or fail a bond are different, I conduct a balance check, which examines whether baseline observable characteristics before treatment are significantly different between the treatment and control groups. If there is no significant difference in these observable characteristics, then researchers typically conclude that the treatment and control groups are “balanced”.[21]

In Table I.4 I present coefficients (β_{-1}) from estimation of Equation (I.6) but using each background variable as the dependent variable.[22] The results suggest balance among all of the observed characteristics, especially pre-existing achievement. I also show the results of a similar check for the RDD model (β_{-1} from estimation of Equation (I.2)). Overall, the results of the balance check provide support for the latent factor model.

Besides the average effect of bond passage, I also estimate the effect of bond passage at any vote share through the following (TOT) regression:

$$Y_{jt} = \sum_{\tau}^{\bar{\tau}} (\alpha_\tau M_{jt-\tau} + \beta_\tau B_{jt-\tau} + \gamma_\tau B_{jt-\tau} \Theta^P_{jt-\tau} + \omega_\tau \Theta^P_{jt-\tau}) + D_j + T_t + \mu_{jt},\ 1996 \leq t \leq 2009, \tag{I.10}$$

where γ_τ is the coefficient on the interaction of bond passage and latent preferences of educational investment. All other parameters are defined in the same way as Equation (I.7). The TOT effect of a bond passed with vote share v' on achievement τ years later is given by

$$(\beta_\tau + \gamma_\tau \Theta^P_{jt-\tau})|_{(\Theta^P_{jt-\tau} = E(\Theta^P | v = v'))} \tag{I.11}$$

[21]The same idea of balance check is adopted in other inference approaches based on the regression discontinuity design. See Cattaneo et al. (2015) for an application in randomization inference in the regression discontinuity design.

[22]The normalization is also adjusted accordingly. I normalize the effects before the first one I want to check to zero. For example, when I check if $\beta_{-1} = 0$, I normalize β_{-2} to zero.

Table I.4: Balance Check for Background Variables

	Latent Factor Model			RDD		
	estimate	standard error[a]	*p*-value	estimate	standard error[a]	*p*-value
Expenditure per pupil[b]						
Total	-39.67	155.8	0.799	-28.03	148.5	0.850
Current	-20.85	81.76	0.799	-31.43	71.63	0.661
Teacher salary	-18.77	34.23	0.584	-16.19	32.12	0.614
Capital[c]	26.51	102.1	0.795	-6.065	110.6	0.956
Construction	14.69	97.46	0.880	4.857	107.0	0.964
Land and structure	11.82	46.58	0.800	-10.92	42.86	0.799
Instructional equipment[d]	-0.305	11.34	0.979	-1.180	7.989	0.883
Demography						
Enrollment	13.63	48.13	0.777	-49.48	44.29	0.265
White students[e]	-0.121	0.246	0.622	0.812	0.831	0.329
Free lunch[e][f]	-0.002	0.006	0.754	0.006	0.009	0.474
Achievement (proficiency)[e]						
4th grade reading	-0.116	1.405	0.934	0.286	1.062	0.788
7th grade reading	-0.213	1.639	0.897	0.667	1.201	0.579
Sample size		9833			9833	

Notes: [a]Standard errors are clustered by district. [b]All expenditures per pupil are measured in constant year 2000 dollars. [c]Capital is defined as the sum of construction, land, and structure. [d]Instruction equipment includes expenditures for all instructional equipment recorded in general and operating funds under "instruction". [e]White students, free lunch and proficiency are measured in percentage points. [f]A student can receive free lunch if the household's income is within the limits on the Federal Income Eligibility Guidelines or one of several other conditions is satisfied. For details see http://www.michigan.gov/mde/0,4615,7-140-66254_50144-194552--,00.html

In Table I.5 I estimate the TOT effects at the cutoff as Expression (I.11) with $v' = 50\%$. For 4th grade a marginally passed bond increases reading proficiency by about 1.5 percentage points after 10 years. For 7th grade there is no significant effect on reading proficiency. The significant effects on 4th grade reading proficiency are similar to those from the RDD approach.[23] However, the results about 7th grade reading proficiency are quite different. The disparity indicates that the latent factor model may not provide reliable average effects of bond passage on 7th grade reading proficiency, probably because the uncovered pref-

[23]The effects of bond passage on 4th grade reading proficiency after 10 years are insignificant in the RDD approach, probably because the RDD approach has relatively low statistical power (Schochet, 2009). Nevertheless, the magnitudes of these effects are similar with the ones from the latent factor model.

Table I.5: TOT Effects of Bond Passage on Achievement at the Cutoff

	Latent Factor Model		RDD	
Relative year	4th grade	7th grade	4th grade	7th grade
1	-0.361	-0.695	-0.434	-0.004
	(0.470)	*(0.489)*	*(1.079)*	*(1.069)*
2	0.531	-0.527	0.077	0.068
	(0.478)	*(0.515)*	*(1.239)*	*(1.331)*
3	0.734	0.118	0.345	0.613
	(0.512)	*(0.494)*	*(1.139)*	*(1.247)*
4	0.569	0.563	0.682	1.033
	(0.479)	*(0.529)*	*(1.106)*	*(1.223)*
5	0.386	-0.744	1.781 *	0.777
	(0.503)	*(0.505)*	*(1.082)*	*(1.321)*
6	-0.017	-0.113	1.370	1.480
	(0.523)	*(0.607)*	*(1.263)*	*(1.388)*
7	-0.197	0.694	1.470	2.816 *
	(0.512)	*(0.577)*	*(1.195)*	*(1.459)*
8	0.383	0.834	2.092 *	3.493 **
	(0.538)	*(0.573)*	*(1.099)*	*(1.445)*
9	-0.290	0.668	1.565	3.497 **
	(0.550)	*(0.624)*	*(1.275)*	*(1.594)*
10	1.390 **	0.508	1.774	3.113 *
	(0.613)	*(0.684)*	*(1.333)*	*(1.815)*
11	1.712 **	0.197	2.126	4.129 **
	(0.715)	*(0.831)*	*(1.450)*	*(1.692)*
12	1.544 *	-0.049	1.658	3.589 **
	(0.862)	*(0.898)*	*(1.665)*	*(1.704)*
Sample size	7244	7219	7244	7219

Notes: The table shows the TOT effects of bond passage on reading proficiency at the cutoff of 50%. The specification is the TOT regression at any vote share described in Equation (I.11). Clustered standard errors by school district are shown in parentheses. * and ** indicate the statistical significance at 10% and 5% levels, respectively.

erences for educational investment are insufficient to eliminate the relationship between vote share and proficiency for grade 7.[24] Nevertheless, the average effects of bond passage on 4th grade reading proficiency seem to be trustworthy. In view of the results of the robustness checks, the discussion of policy implications that follows focuses on 4th grade only.

I estimate alternative specifications that control for other potential confounders but not latent preferences for educational investment (See Table A.5 in the Appendix for details.).

[24] Using a extrapolation based conditional independence assumption, Angrist and Rokkanen (2015) find that the extrapolation results of one of the two examined outcomes are more reliable than the other.

The year fixed effects capture a large deal of variation in 4th grade reading proficiency that is related to the time trend. The bond fixed effect in the ITT estimation and the district fixed effect in the TOT estimation also significantly shrink the effects estimated. Compared with the results in Table A.5 in the Appendix, controlling for latent preferences for education investment substantially changes the estimates.

I.5 Policy Implications

I examine the effectiveness and efficiency of the current capital financing system and a hypothetical alternative system — matching grant program. I show that (1) both systems have the potential to improve student achievement with positive net benefits; and (2) investing in school capital through the system of bond referendum is unlikely to be the optimal educational investment to improve student achievement in terms of the internal rate of return.

I.5.1 Cost-Benefit Analysis

Investing in school capital can potentially increase student achievement, but it is not clear if capital investment is the best use of tax dollars targeted for education. Krueger (2003) conducts a cost-benefit analysis for class size reductions in kindergarten using the Tennessee STAR class size experiment. Following his assumptions and methodology, I conduct a cost-benefit analysis for capital investment through a bond passed in year 2000 from the viewpoint of an average household. Given the high level of average investment required, I find that although the benefits of capital expenditures may still cover the costs under certain reasonable scenarios, overall the investment in school capital under the current financing system through referendum is unlikely to be the optimal choice in terms of rate of return.

For a household the cost of school capital investment is calculated as the property tax levy. A household repays the passed bond through the property tax; the repayment time and the millage rate are set in the bond proposal. The present value of the cost for household i is given by

$$PVC_i = \sum_{t=0}^{T} mW_t(1+r)^{-t} \tag{I.12}$$

m is the average millage rate of the passed bonds, which was 2.44 per \$1,000 in year 2000. W_t is house value measured in thousands of dollars at time t. I start my cost analysis in year 2000, when the median house value in Michigan was $W_0 = \$115,600$.[25] Without bond passage, I assume a 3% annual increase in house value.[26] In addition, according to Cellini et al. (2010) and Nielson and Zimmerman (2014), a passed bond can increase the house price by about 4% in the year following the bond, 6% in years 2–5 and 8% in year 6. A passed bond does not have a significant effect on housing prices after 6 years. Therefore, for a district that passes a bond I have $W_1 = W_0(1+4\%)$, $W_t = W_{t-1}(1+6\%)$ for $t = 2--5$, $W_6 = W_5(1+8\%)$, and $W_t = W_{t-1}(1+3\%)$ for $t \geq 7$. For a school district that does not pass a bond, $W_t = W_{t-1}(1+3\%)$ for all t. r is the discount rate, which is assumed to be 3%, 4%, 5% or 6%.[27] T is the average repayment time of the passed bonds, which was 23 years in year 2000. In conclusion, the present value of cost in year 2000 over a 23-year period for an average household i is $C_i = \sum_{t=0}^{23}(2.44 \times W_t)(1+r)^{-t}$, where W_t is described above and $r = 3\%$, 4%, 5% or 6%.

Following Krueger (2003), I use the following formula to calculate the present value of benefits for an average household with one child who is in grade 4 (aged 9) in year 2011, 11 years after a bond passed in year 2000;[28] begins work at age 18 (year 2020); and retires

[25]Source: http://www.zillow.com/mi/home-values/.

[26]The annual increase rate from August 2014 to August 2015 is 3.2%. For details see http://www.zillow.com/mi/home-values/.

[27]In 2000, the daily treasury real long-term rates is about 3-4%. For details see http://www.treasury.gov/resource-center/data-chart-center/interest-rates/Pages/TextView.aspx?data=reallongtermrate. I also assume higher rates because society may desire a higher discount rate to reflect the uncertainty in the future benefit.

[28]Note that the significant TOT effect comes up after 10 to 12 years.

at age 65 (year 2067):

$$PVB_i = \sum_{t=(2020-2000)}^{(2067-2000)} E_t(1+r_w)^{(t-4)} \times s(\beta_r + \beta_m) \times (1+r_b)^{-t}, \tag{I.13}$$

where E_t denotes earnings in year t. I update the age-earning profile using the report from the Michigan Department of Treasury and the Office of Revenue and Tax Analysis (Michigan Department of Treasury AND Office of Revenue and Tax Analysis, 2009). r_w is real wage growth rate. I assume zero, one or two percentage of real wage growth after year 2000.[29] s is the effect of a one standard deviation increase in achievement on earnings. I assume $s = 8\%$ as in Krueger (2003). In the current study one percentage point of proficiency approximately translates into 0.05 standard deviations (SD). β_r is the average effect of bond passage on reading score, which is about 0.063 SD. β_m is the average effect of bond passage on math score. I do not have consistent measures of math proficiency. Nielson and Zimmerman (2014) do not find positive effects of capital expenditure on math scores, although they find positive effects comparable to mine on reading scores. I assume two scenarios with the same effects on math and reading proficiency or zero effect on math proficiency, so β_m is either 0 or 0.063 SD.

Panel (A) of Table I.6 summarizes the cost, net present value (NPV) and internal rate of return (IRR) for an average household. Whether NPV is positive depends on parameters such as the discount rate and whether there are effects on math. In the most optimistic scenario (same effect on math, 2% annual increase in wage and 3% discount rate), the IRR is 5.6%. This is smaller than the IRR (6.2%) of reduction in class size from 22 to 17 students (Krueger, 2003).[30]

[29]Real earnings have grown by 1% or 2% per year over the 20th century (Krueger, 2003).

[30]I may omit several other potential benefits from capital expenditure, such as the positive effect on noncognitive skills. The increased achievement in elementary and middle school may also increase the attainment of college education. Moreover, both noncognitive skills and college education increase future income, influence adult behaviors such as smoking and marriage, and increase longevity (Hong et al., 2014). All of these effects have private and social benefits with economic consequences. Given that I may underestimate the benefit by omitting these considerations, investing in capital gives reasonable returns. However, since the cost-benefit analyses of other investments in education, such as Krueger (2003), also omit some their

Table I.6: Cost Benefit Analysis

	Discount rate	Cost ($)	Net present value ($)[a]					
			no effect on math			same effect on math		
			anuual real wage growth rate			anuual real wage growth rate		
			0	1%	2%	0	1%	2%
(A) Average effect								
	0.03	6,871	-3,221	-1,234	1,906	428.9	4,403	10,683
	0.04	6,169	-3,838	-2,614	-699.7	-1,507	941.9	4,770
	0.05	5,569	-4,056	-3,290	-2,105	-2,543	-1,010	1,359
	0.06	5,054	-4,056	-3,569	-2,825	-3,058	-2,084	-595.4
	IRR[b]		0.012	0.025	0.037	0.032	0.044	0.056
(B) Matching grant								
	0.03	10,117	-5,308	-2,690	1,446	-499.4	4,737	13,010
	0.04	9,000	-5,929	-4,316	-1,794	-2,858	368.6	5,412
	0.05	8,057	-6,063	-5,054	-3,493	-4,070	-2,051	1,070
	0.06	7,256	-5,942	-5,300	-4,320	-4,627	-3,344	-1,383
	IRR[b]		0.009	0.021	0.034	0.028	0.041	0.054

Notes: Cost and net present value are in dollars as present value in year 2000. [a]Net present value equals to net present value of benefit minus net present value of cost. [b]IRR is the discount rate that gives zero net present value. Cost is calculated by Formula (I.12). Net present value is calculated by Formulas (I.12) and (I.13).

I.5.2 Counterfactual Simulation

The current funding system for capital investment in Michigan is a local funding system, which makes capital investment depend on local decisions and economic conditions rather than need. Since school districts vary greatly in willingness and ability to pay for educational investments, the current system likely increases the disparity in capital outlays across districts, leading to high intrastate inequality in per-pupil capital expenditure (Wang and Duncombe, 2009). To address this inequality issue, Michigan is considering a shift from the current system to a (partially) centralized funding system (Pratt et al., 2012).

One potential alternative under consideration is a matching grant program, by which the state would provide eligible districts with a specified level of state funding to match each millage of a local levy. This matching system by design should decrease disparities in

potential benefits, it is still unlikely that capital investment is the best educational investment choice.

capital outlays, but it remains unclear if the subsidized districts which ultimately pass the bond benefit and if such a system is beneficial from the viewpoint of cost-benefit analysis.

Suppose the state provides funding, with 100% matching rate, to districts that fail a bond to enhance overall capital investment, and to reduce the inequality. A $d\%$ matching rate is approximately equivalent to increasing the bond amount by the same percentage, without changing other bond characteristics. To estimate the matching rate that induces the average failed bonds to pass, I regress vote share on a set of bond characteristics including bond amount per pupil, repayment time, millage rate, type of construction, election date and year, and total number of voters. The results show that a 100% increase in bond amount per pupil, conditional on other covariates, increases vote share by 5.6 percentage points. As a result, the failed bonds with vote share between 44.4% and 50% will get passed under the 100% matching grant program.

The households can benefit from the matching grant program; bond passage has, on average, a positive effect on 4th grade reading proficiency. The average effect of bond passage on 4th grade reading proficiency after 11 years is 0.083 SD, which is obtained from the following expression:

$$(\beta_\tau + \gamma_\tau \Theta^p_{jt-\tau})|_{(\Theta^p_{jt-\tau} = E(\Theta^p|44.4\% \leq v \leq 50\%))}. \tag{I.14}$$

However, given a matching rate as high as 100%, it is not clear if the program provides a net benefit. In 2000, the average bond amount per pupil of failed bonds with vote share between 44.4% to 50% was \$8,574. The average repayment time and millage rate for the bonds affected by the program are 25 years and 3.32, respectively. All of the other parameters are assumed to be the same as in Section I.5.1.

I use the same method as in Section I.5.1 to calculate the pecuniary benefit, NPV and IRR. The results are summarized in Panel (B) of Table I.6. Whether or not the matching grant program is beneficial depends on discount rate, annual real wage growth rate and the effect in other dimensions. The IRR of 0.009 in the most conservative case is likely to be the lower bound. Again, the matching grant program is cost-effective under certain scenarios, but with modest internal rate of returns it is unlikely to be the optimal investment choice.

I.6 Conclusions

I use administrative data from Michigan to estimate the average treatment effect of bond passage on student achievement. Using a latent factor model, I uncover the underlying confounders of preferences for educational investment, and extrapolate the effects of bond passage to the districts with vote share away from the cutoff. On average, a passing bond positively affects achievement, especially for grade 4.

The paper contributes to the literature by offering deeper understanding of the average effect of financing capital expenditure through bond referendum on student achievement. School districts can improve student achievement by investing in infrastructure through the bond financing system. However, there are ways to improve the effectiveness of the capital investment. An alternative system that promotes capital investment in relatively disadvantaged districts (e.g., matching grant program) can potentially have larger effects on student achievement in the state.[31] Nevertheless, compared with other existing investments that also aim to improve student achievement, such as the class size reduction program, investing in school capital is less cost-effective.

Increased capital investment explains part of the observed effects. There are other potential channels through which a passed bond affects student achievement, such as demographic changes including household reallocation across districts. Future study may examine the mechanisms linking bond passage and student achievement by exploring the relative roles of capital expenditure and other channels in producing the total effects.

[31] Recall that districts that failed an election were more disadvantaged (Table I.1).

Chapter II

Do Selective High Schools Improve Student Achievement? Effects of Exam Schools in Urban China

More than 80,000 ninth grade students took the high school entrance exam on June 26, 2014 in Beijing; the exam was given in about 3000 classrooms and lasted for three days. Performance on the exam would decide which middle school students would attend high school in the fall semester. High schools vary in quality, and the competition to get into elite schools is severe; the minimum requirement for admission to some elite schools is to score as high as 95% on the entrance exam. College admission is also based on an exam taken after high school, and graduation from a higher-ranked college usually indicates that the student will obtain a good job with a higher salary and better working conditions. Parents and students think that graduation from an elite high school with high quality teachers and high-ability peers increases the probability that the student is admitted to a good college; graduation from an elite high school can be the ticket to a students successful future career.

The exam school system is one of the most important parts of Chinese education; many public resources are involved, but its effectiveness in improving student achievement remains unclear. The evidence on the effectiveness of elite schools which admit students by exam score is under debate in the United States. A recent article in Slate magazine advocates that super-elite public schools arent necessary any more (Salam, 2014). The author of the article looks at Stuyvesant High School in New York City and describes several potential problems at elite high schools like Stuyvesant; most of these problems are associated with the fierce competition at the schools which encourages languishing among lower-ranked students and cheating on exams. Salams solution is to close such one-size-fits-all elite schools and spread gifted students across a wide range of high schools.

The elite school model is found in countries other than China and the United States.

Romania, Trinidad and Tobago, and the United Kingdom have elite schools that admit students based on an admission test score, especially at the high school level, and the majority of public high schools in Singapore are exam schools. Evaluations of student performance in exam schools produce mixed results. The positive effect of elite schools on the final year test scores is found in studies by Pop-Eleches and Urquiola (2013) in Romania, Jackson (2010) in Trinidad and Tobago and Park et al. (2015) in rural China. Other studies show little to no effect of exam schools on student outcomes, including Abdulkadiroglu et al. (2014) in Boston and New York City, Dobbie and Fryer Jr. (2014) in New York City, Clark (2010) in the United Kingdom, Lucas and Mbiti (2014) in Kenya and Aiayi (2014) in Ghana. The studies of elite schools which admit students by lottery also show no consensus. The positive effect on test scores is found by Hastings and Weinstein (2008) in Charlotte-Mecklenburg, North Carolina, but little evidence of a positive effect on student achievement is found by Zhang (2014) in urban China and Cullen et al. (2006) in Chicago.[1]

One possible reason for the mixed results is that the elite school is not a homogenous educational experience; many characteristics related to achievement can affect academic outcomes. In most cases elite schools admit higher ability students. Students are influenced by their peers, and students may learn more if they interact with smarter peers (Hoxby and Gretchen, 2006). However, studies of exam schools find positive peer effects in some cases (Jackson, 2013, in Trinidad and Tobago) and no peer effects in others (Abdulkadiroglu et al., 2014, in Boston and New York City). Higher school quality is another expected characteristic of elite schools, although it is difficult to precisely measure quality. Some literature shows that school quality and teacher qualification have positive effects on student performance (Lai et al., 2011, in China). Since school quality is multi-dimensional, elite schools do not necessarily outperform other schools in all school quality measures which have significant impacts on student exam performance.

[1] Zhang (2014) studies the effect of elite middle schools that admit students by lottery in a provincial capital city in China. He finds that attending elite middle schools has no significant impact on high school entrance exam scores.

Another possible explanation for the mixed results concerns differences in samples across school settings (Park et al., 2015). An RDD study compares students near the cutoff, and the result of this comparison may depend on the school setting. If the studied system of schools is in a rich metropolitan area and the elite schools are extremely selective, a significant positive effect of the elite school is less likely. The students who fail to enter an elite school in this setting may still get admitted to a high-quality school. Their families may also have sufficient resources that are used to complement school quality. On the flip side, if the studied system of schools is in a poor rural area, the potential gap in quality between elite schools and other schools can be very large. Families also lack resources to fill the quality gap between schools, and we may observe significant effects of elite schools on academic performance. In an extreme case where the educational system is inefficient, even elite schools cannot effectively improve student performance.

One difficulty in measuring the effect of exam schools is the admission rule itself. Students admitted by higher-ranked schools perform better than other students on the admission test. Performance on the test reflects how well students did in school before the test and is positively correlated with student ability and parental resources. Students in higher-ranked schools are more likely to obtain higher scores on exams even if they are enrolled in a lower-ranked school because these students, on average, are more able and have more parental resources. The crucial empirical problem in the evaluation of the effectiveness of elite exam schools is modelling this selection by ability. Several recent studies implement regression discontinuity design (RDD) to solve the selection problem (Hahn et al., 2001; Imbens and Lemieux, 2008; Lee and Lemieux, 2010).

Our paper evaluates the effect of exam schools in Beijing on academic performance using the RDD strategy. Each school has a minimum test score for admission. None of the individual school cutoffs are deterministic; students can attend an exam school with scores below the relative cutoff, and students with scores above the cutoff can attend another lower-ranked school. The structure for admission into high schools fits the fuzzy RDD setting.[2] The forcing variable by which students are assigned to schools is their score on the entrance exam to high school (SEEH); the main outcome of high school selection is the score on the entrance exam to college (SEEC). Our study contributes to the existing literatures in several ways. First, we examine the effect of a couple of elite schools and all of the other non-elite exam schools in a large metropolitan area of China. All of these schools follow the same student admission rules but with different admission cutoff points. We have multiple settings for these schools and are able to examine whether the sample differences across settings explain the mixed findings in literature. Second, the distribution of student achievement is broad across the exam schools in our study. With RDD we can only identify the effect at the cutoff, but our study provides stronger external validity by identifying the effects of elite schools at multiple cutoffs. Finally, our paper examines whether there are heterogeneous effects by gender and parental education and occupation, and we provide insight into whether selective exam schools in Beijing decrease the gender gap in achievement and variation in achievement by parental background.

We find no effect of attending elite schools on SEEC. However, we find that on average the exam school system indeed increases SEEC, which indicates that students can potentially benefit from attending a more selective non-elite exam school. On average attending a more selective school with a higher admission cutoff increases SEEC by about 0.36 standard deviations. We attribute these findings to the observed educational inputs in exam schools including peer effects and teacher quality.

II.1 Data and Institutional Background

II.1.1 Descriptive Statistics

Park et al. (2015) study magnet high schools in rural counties in Western China; we ex-

[2] Students can attend a higher-ranked school because they are able to obtain extra scores if certain requirements are satisfied, such as being of minority race or the child of a martyr. Students choose a lower-ranked school rather than higher-ranked schools for which they are eligible because, when students report their school preferences, they do not rank those higher-ranked schools above the school they actually attend.

amine the effect of exam schools in Beijing City under an urban setting.[3] Our data include administrative information on student demographic characteristics and outcomes from an anonymous district of Beijing. The district used to be a suburb surrounding the Beijing metropolitan area and became an urban district in 2001 with the expansion of Beijing City. Information on the school level is also reported. The data include all 11 high schools (named AS to KS) and 3868 grade 12 students in the 2008 cohort. All of the 11 schools are exam schools that admit students solely by SEEH, and two of them are elite schools (AS and BS). The students in our sample took the entrance exam for high schools (SEEH) in 2005 and the entrance exam for colleges (SEEC) in 2008. Students self-chose whether to take the SEEC, and approximately 7.9% of the students (305) did not take the SEEC. We include the students without SEEC only in the analysis about the self-choice of exam participation. We also exclude one student who enrolled in an unknown high school.

We collect scores on both the high school and college exams, student characteristics, parental educational background, and school characteristics. The descriptive statistics on these characteristics by school are shown in Table II.1. We normalize SEEH and SEEC to be the number of standard deviations from their means. Parental educational background variables are whether the father and mother have a bachelor's degree.

Elite schools in Beijing originated from the key schools policy in the 1950s. This policy allocated more resources to certain schools in hopes of getting better educational outcomes with limited educational resources. Compared with other non-elite schools, elite schools receive more financial support from the government. Because of their good reputation, it is easier for elite schools to obtain support from other funding sources. Elite schools are more attractive for middle school graduates. Elite schools generally are higher quality and have more high-achieving students.[4]

Students in the elite schools do better on average on both the SEEC and SEEH. They

[3] Park et al. (2015) studies Gansu Province, which is one of the poorest provinces in China. The GDP per capita of Gansu Providence is about $750in2004, whiletheGDPpercapitaofBeijingCityin2004isabout5000$ (National Bureau of Statistics, 2005).

[4] Details about elite and non-elite schools can be found in Appendix B.

Table II.1: Descriptive Statistics

	Total	Elite			Non-elite
		Total	AS	BS	
Students					
Exam Score					
SEEC	0	0.742	1.003	0.397	-0.314
	(1)	*(0.778)*	*(0.722)*	*(0.714)*	*(0.914)*
SEEH	0	0.958	1.219	0.615	-0.368
	(1)	*(0.478)*	*(0.438)*	*(0.265)*	*(0.899)*
Student Characteristics					
Male	0.545	0.545	0.573	0.509	0.545
Age	18.72	18.6	18.46	18.72	18.77
	(0.712)	*(0.663)*	*(0.609)*	*(0.702)*	*(0.722)*
Parental Background					
College Father	0.184	0.209	0.365	0.004	0.175
College Mother	0.166	0.208	0.361	0.009	0.149
Sample Size	3867	1073	609	464	2794
School					
Number of Enrolled Students	1232	1894	2286	1502	1085
Number of Teachers	106	155	153	156	95
Students/Teachers	12.01	12.29	14.94	9.63	11.95
Percentage of Teachers with Advanced Certificate	0.267	0.421	0.412	0.429	0.233
Percentage of Teachers Younger than 35	0.604	0.431	0.451	0.41	0.643
Minimum Score of SEEH for Admission in 2005	435	482	490	474	424

Notes: Sample means for the characteristics of students in the 2008 cohort are reported. Sample size is the number of observations with non-missing values on the SEEH. AS and BS are the elite exam schools. Data source: Administrative data from an anonymous district in Beijing.).

are also slightly younger and come from families with more advantaged social status; their parents are more likely to hold a bachelor's degree. There is no significant difference in student gender.

The elite schools have larger enrollment and more teachers, but their student/teacher ratios are also larger. The quality of teachers is higher in the elite schools; the teachers are more experienced, and a higher percentage of teachers has an advanced certificate. As a consequence, the elite schools are popular among high-achieving students. This leads to a much higher minimum SEEH for enrollment as well as higher-quality peers in the elite schools.

We find significant differences between student characteristics in our urban sample and in the sample of rural students in Park et al. (2015). For example, in our sample 45.5% of the students are female while in their sample about one third (35.9%) are female.[5] The elite schools in our sample are more selective; they admit 27.7% of the students while in the Park et al. (2015) sample 53.5% of the students enroll in elite schools. There is also a gap in educational outcome. In the Park et al. sample 49.7% of the students are eligible for college after three-years of study in high school, while in our sample 79.1% of the students are eligible for college. All of these differences indicate that the students in our sample come from a quite different setting; our results complement previous studies of elite schools in China but from a different perspective.

II.1.2 Exam School Applications and Admissions

The Education Bureau in Beijing administers a city-wide uniform high school entrance exam. Middle school graduates must take the exam before they can be admitted to public high school. To be eligible for a high school, students need to get exam scores above the cutoff set by the corresponding school. When the entrance exam scores become available, the Education Bureau assigns students to the elite schools starting with the student with the highest score until the school's capacity is filled. The cutoff for admission for each school is the exam score of the last student who filled the school's capacity. The remaining students can be admitted by non-elite high schools if their scores are higher than the cutoff lines set by the non-elite high schools. A student who is not admitted by any public high school can attend a private or vocational high school or can drop out of school.[6]

Manipulation about taking the exam, admission and enrollment is unlikely. It is difficult for students to precisely predict whether expected scores are higher or lower than a specific cutoff; the cutoffs are set after the test scores are available and largely depend on the performance of other students. The Education Bureau in Beijing strictly follows these

[5] In China high school education is not compulsory. In poor rural areas families are more likely to let girls dropout after middle school for housework.

[6] The high school admission procedures in Beijing are described in detail in Appendix C.

procedures, making manipulation at the administrative level also unlikely.

II.2 Baseline Analytical Framework

II.2.1 Regression Discontinuity

Equation (1) models the SEEC outcome:

$$SEEC_i = \alpha + \beta H_j + e_i, \tag{II.1}$$

where $SEEC_i$ is the $SEEC$ of student i. H_j is an indicator of enrollment in school j. α and β are coefficients. e_i is the error term that includes all of the unobserved factors which are correlated with $SEEC$. If the high school assignment is uncorrelated with the error term, an OLS regression gives consistent and unbiased estimation of the effect of attending school H_j on $SEEC$.

However, high school assignment is not exogenous. To account for this endogeneity, we implement RDD with $SEEH$ as the forcing variable, which assumes that the unobserved factors in the error term in (II.1) are uncorrelated with high school assignment at the corresponding cutoffs after controlling for a polynomial of $SEEH$. We estimate the following equation:

$$SEEC_i = \alpha + \beta H_j + \gamma f(SEEH_i - c_j) + e_i. \tag{II.2}$$

c_j is the cutoff of school j. Attendance at high school H_j is positively correlated with the indicator of whether the $SEEH$ is higher or equal to the cutoff for that school, but this relationship is not deterministic; students may propose and attend a school with a lower $SEEH$ cutoff, and it is possible for students with $SEEH$ lower than the cutoff to attend the corresponding high school, as long as certain extra requirements are satisfied. Taking such fuzziness into account, we estimate two types of treatment effect. First, we replace the school attendance variable in equation (II.2) with an indicator of whether the $SEEH$ is higher or equal to the cutoff and get equation (II.3) in a reduced form:

$$SEEC_i = \alpha + \beta 1\{SEEH_i \geq c_j\} + \gamma f(SEEH_i - c_j) + e_i. \tag{II.3}$$

β in equation (II.3) identifies the intent-to-treat (ITT) effect of high school; it is the effect of high school eligibility rather than the effect of attending high school.

We stack the data for each school and estimate the average effect over the samples of schools.[7] The average ITT effect is equal to β in equation (II.4):

$$SEEC_{ij} = \alpha + \beta 1\{SEEH_i \geq c_j\} + \gamma f(SEEH_i - c_j) + \delta_j + e_{ij}, \tag{II.4}$$

Subscript $j = 1, 2, , 10$ indicates the j-th individual school; $j = 1, 2$ for the elite schools. δ_j is the fixed effect of school j.[8] The stacked school analysis helps us evaluate the group of exam schools as a whole and provides stronger statistical power.

In the main analysis we use local linear regression to estimate the models, with the optimal bandwidth derived by Imbens and Kalyanaraman (2012), a uniform kernel, and a piecewise linear function of SEEH, which is $f(SEEH_i - c_j) = \gamma_1(SEEH_i - c_j) + \gamma_2(SEEH_i - c_j)1SEEH_i \geq c_j$. We check the robustness of our results with respect to different estimation settings.[9]

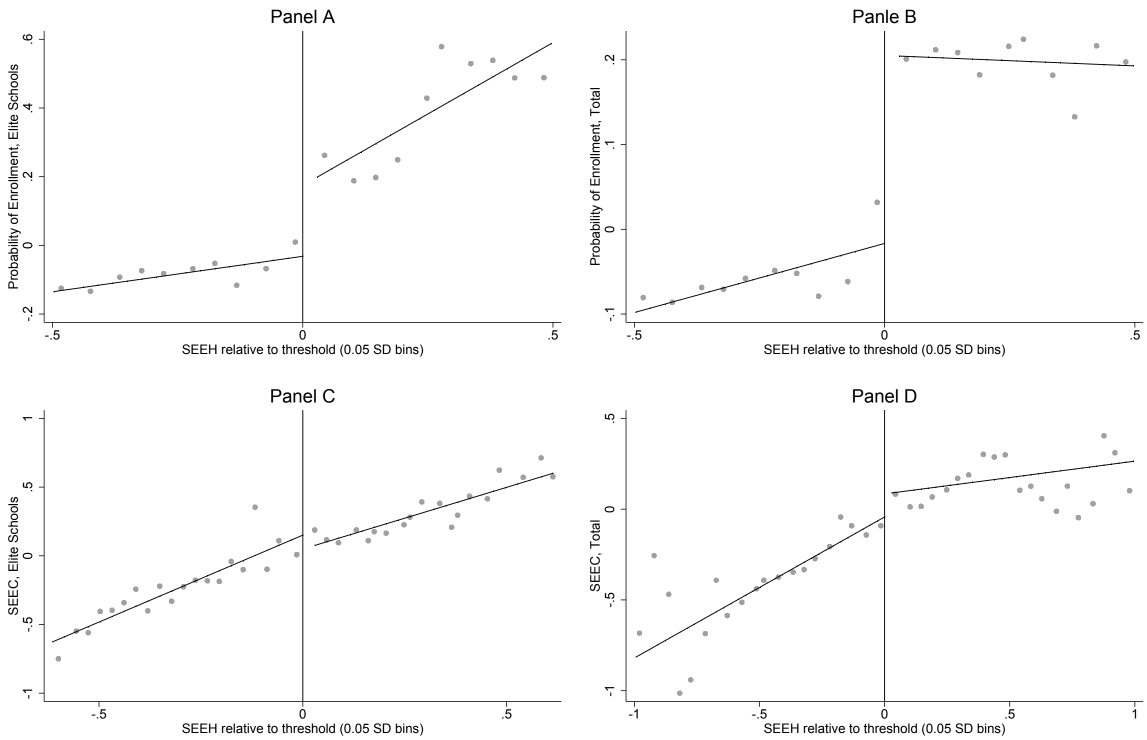
In Figure II.1 we present the probability of attending more selective schools by SEEH around the cutoff (Panel (A) for the elite schools and Panel (B) for the whole sample).

[7] We generate the stacked data set in a way similar to Pop-Eleches and Urquiola (2013). First, we estimate the optimal bandwidth for each school separately. Second, we stack all observations with SEEH in the optimal bandwidth for each school. Finally, the stacked data sets for each school are combined into one large data set. In the stacked data set one student may be included more than once if his/her SEEH is in the bandwidth of more than one school. For example, school A has a zero cutoff and a bandwidth $[-1, 1]$, and school B has a cutoff of one and a bandwidth $[0, 2]$. All students with SEEH in window $[0, 1]$ are included in the analyses for both schools. In the stacked data set each observation is identified by school j and forcing variable $SEEH_i - c_j$. A similar approach with stacked data is used in the literature on other research topics (Cellini et al., 2010, 2010, for example).

[8] There are 11 schools in the analysis. However, we can only estimate the effects for 10 of them, as there is no clear control group for the school whose cutoff is the lowest.

[9] Various bandwidths for each school can be found in Appendix Table A.6.

Figure II.1: Student Enrollment and SEEC around Cutoffs



Notes: Dots represent the means of residuals from a regression of the dependent variable on school fixed effects, in each bin with the width of 0.05 standard deviations. The lines are linear fits to approximate the underlying sample mean of the residuals conditional on SEEH. Data source: Administrative data from an anonymous district in Beijing.

Graphically the student whose SEEH is higher than the cutoff is more likely to attend the corresponding more selective schools. The magnitude of the jump is about 30% for both samples.

In Panels (C) and (D) we present the graphic evidence of whether or not SEEC increases discontinuously around the cutoff. For the elite schools Panel (C) suggests at best no effect of being eligible for the elite schools on SEEC. Nevertheless, Panel (D) suggests that potentially students can benefit from being eligible for a more selective school on average.

II.2.2 Validity Test

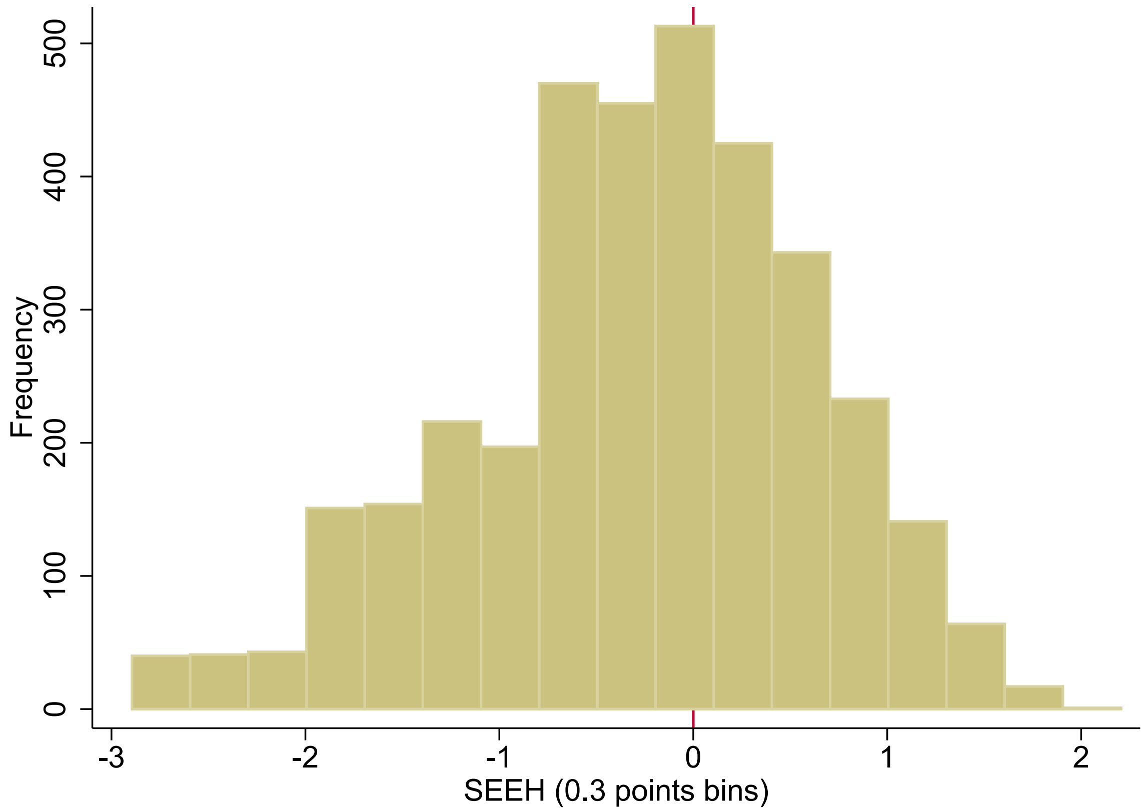
RDD produces unbiased estimates of the effects of high school on subsequent student achievement if there is no perfect manipulation of the exam and admission. Perfect manipulation is unlikely at the cutoff; the cutoff is unknown to anyone,[10] and nobody can perfectly control it. Graders of the SEEH exam do not know whose test they are grading because the name is sealed. There are many graders, and a few cheating graders, if any, should not lead to systematic manipulation. Students can attend cram schools, but re-taking the exam at least one year later is costly for teenagers and does not guarantee admission. We formally examine whether there is any evidence of manipulation by displaying the distribution of students observed in our data. According to McCrary (2008), discontinuity in the density around the threshold indicates the risk of endogenous sorting which violates the RDD assumptions. Figure II.2 is a histogram that shows the distribution of students by SEEH. There is no evidence of manipulation around the cutoff of elite schools. We also perform a formal density test for elite schools and the whole sample. We do not find evidence of manipulation around the cutoff in either case.[11]

The internal validity of the RDD requires that no relevant variables other than the treatment jump at the cutoff. Table II.2 shows the details of the validity tests of the background variables. In Column (a) we compare the background variables by type of school. A simple t-test shows that students who attend elite schools are significantly different from students who attend non-elite schools in their age and parental college degree. This comparison indicates that an OLS regression model is likely to give us biased results. In Column (b) we perform the formal balance check by estimating Equation (II.4) with each background variable as the dependent variable. In most cases we do not have unbalancedness, except that students in the elite schools are slightly younger. Nevertheless, a difference of 0.09

[10] Students can obtain information about the cutoffs in previous years and predict the current cutoff. Prediction error always exists, and on average it is unlikely for a substantial number of students to perfectly predict the cutoff.

[11] In both cases the changes in density around the cutoff are -0.001, and neither is significant.

Figure II.2: Student Distribution by SEEH



Notes: The width of each bin is 0.3 standard deviations. SEEH is censored at 3 standard deviations away from the cutoff. Data source: Administrative data from an anonymous district in Beijing.

years, which is approximately 5 weeks, should not be a big concern. As a result, we do not control for covariates in the model, but we show that the addition of the background variables as covariates does not change our results.

II.3 Effect of Exam Schools on Student Achievement

II.3.1 ITT Effect of Exam Schools on Enrollment and SEEC

The ITT analysis provides insight into the effect of high school eligibility. We first estimate Equations (II.4) using the actual enrollment as the dependent variable. Column (a) of Table II.3 shows the results of eligibility on the enrollment in the corresponding

Table II.2: Balance Check of Density, Background Variables and Self-Choice Variables

	Summary by School Type (a)				Test (b)		
	Elite	Non-Elite	Diff (t test)		Elite		Total
Male	0.544	0.558	-0.013		-0.014		0.000
			(0.018)		(0.029)		(0.023)
Age in 2008	18.57	18.76	-0.193	***	-0.091	**	-0.030
	(0.662)	(0.714)	(0.026)		(0.035)		(0.027)
College Father	0.208	0.185	0.023		-0.032		-0.029
			(0.014)		(0.029)		(0.038)
College Mother	0.209	0.157	0.052	***	-0.02		-0.015
			(0.014)		(0.022)		(0.030)
Sample Size	1060	2520	3562		3197		12832

Notes: The standard error which is robust and clustered on school is shown in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% significance levels, respectively. Column (a) uses the original data set, and Column (b) uses the stacked data set. Data source: Administrative data from an anonymous district in Beijing.

schools.[12] The results indicate that eligibility for a seat in an exam school increases the probability that the student enrolls in the corresponding school.

Column (b) shows the effect of eligibility on SEEC, from the reduced form estimation of Equation (II.4) with SEEC as the dependent variable. For elite schools, eligibility for an exam school with a higher cutoff does not lead to improvement in student achievement, after controlling for pre-existing achievement. For the whole sample we find evidence that students benefit from the exam school system if they are eligible for more selective schools with higher cutoffs. Since we do not find any effect of elite schools, the positive effect for the whole sample indicates that attending a more selective non-elite school probably leads

[12]We conduct robustness checks with respect to other optimal bandwidths such as the data-driven optimal bandwidth (Calonico et al., 2014) and the cross-validation optimal bandwidth (Ludwig and Miller, 2007). Our results are robust to the choice of bandwidth. We also conduct piecewise parametric estimations with all observations within a window of 1.5 standard deviations and get similar results; the parametric results are available upon request.

Table II.3: Effect of Eligibility on SEEC and Enrollment

	Enrollment (a)		SEEC (b)		SEEC (OLS) (c)	
	Elite	Total	Elite	Total	Elite	Total
Effect (β)	0.309 **	0.297 ***	-0.101	0.107 *	0.502 ***	0.458 ***
	(0.132)	(0.060)	(0.091)	(0.048)	(0.085)	(0.044)
F statistic	384.0	11.88	75.82	109.2	34.78	110.4
Controlling for SEEH	Yes	Yes	Yes	Yes	No	No
Statistics of dependent variable	0.247	0.191	0.313	-0.237	0.313	-0.237
			[0.763]	[0.900]	[0.763]	[0.900]
Sample Size	3053	11616	3053	11616	3053	11616

Notes: The standard errors which are robust and clustered on school are shown in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% significance levels, respectively. Statistics on the dependent variable are the mean and standard deviation (in brackets for the continuous variables only) of enrollment and SEEC. Data source: Administrative data from an anonymous district in Beijing.

to higher student achievement. Finally we estimate an OLS regression model which does not control for SEEH (Column (c) of Table II.3). In both cases the results indicate that, without controlling for SEEH, eligibility has a strong positive effect on SEEC; the effect of SEEH from OLS estimation is very likely to be upward biased.

II.3.2 LATE of Exam Schools on SEEC

The local average treatment effect (LATE) provides insight on the effects of attending a high school rather than just eligibility. We estimate a fuzzy RDD model in (II.5) using the indicator $1SEEH_i \geq c_j$ as an instrument for school enrollment.

$$\begin{cases} H_i = a + b1\{SEEH_i \geq c_j\} + cf(SEEH_i - c_j) + d_j + u_i \\ SEEC_{ij} = \alpha + \beta H_i + \gamma f(SEEH_i - c_j) + \delta_j + e_{ij}. \end{cases} \tag{II.5}$$

Equation (II.5) identifies the LATE of attending a high school as (β/b); it estimates the average effect for compliers at the cutoff who would attend the high school when the SEEH exceeded the cutoff and would not attend the high school when the SEEH was less than the

Table II.4: LATE of Enrollment on SEEC

	Two-Step Estimation (a)			OLS Estimation (b)				
	Elite	Total		Elite		Total		
Effect (β/b)	-0.326	0.361	**	1.133	***	1.891	***
	(0.370)	(0.176)		(0.432)		(0.460)	
F statistic	506.9	71.54		4.50		4.83	
Statistics of dependent variable	0.313	-0.237		0.313		-0.237	
	[0.763]	[0.900]		[0.763]		[0.900]	
Sample Size	3053	11616		3053		11616	

Notes: The standard errors which are robust and clustered on school are shown in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% significance levels, respectively. Statistics on the dependent variable are the sample mean and standard deviation (in brackets) of SEEC. Data source: Administrative data from an anonymous district in Beijing.

cutoff.

The results of LATE are shown in Table II.4 (Column (a)). We find little evidence of a positive effect of elite school on exam performance. However, we find that on average attending a school with a higher cutoff increases SEEC by 0.36 standard deviation. The OLS estimation in Column (b) shows that the results without controlling for SEEH have large upward bias.

Our results are consistent with the findings in the previous literature. The elite schools in our sample are comparable to elite schools in large metropolitan areas in developed countries in selectivity[13], and we do not find any effect of elite schools on test scores. The other non-elite schools are close to the elite schools in poor rural areas in school quality. Our result of 0.361 is similar with the effect of elite schools in rural areas from Park et al. (2015), which is 0.387. Non-elite schools in Beijing generally perform well, especially on

[13] The elite schools in Beijing are probably comparable to elite schools in developed countries in school quality and educational outcomes. Household income is lower, and schools receive fewer educational resources in China; however, the Chinese people put great importance on education. Chinese students in schools in metropolitan areas generally perform well. One piece of evidence for this good performance is Shanghais number one ranking on the global PISA test.

exams. For example, in our sample about 70% of the graduates from non-elite schools are eligible for college; attendance at a more selective non-elite school should lead to better student performance on the college admission exam. Our evaluation of the whole system of exam schools in a metropolitan environment confirms that the effect of selective schools depends on the research setting. We are less likely to find a positive effect of the most selective schools, especially in a developed metropolitan area. Students marginally rejected by the elite schools are also excellent, and the top non-elite schools also have good quality.

II.3.3 Effect of Exam Schools on College Admission

We examine the effect of exam schools on other outcomes of college admission. College admission outcomes are correlated with the SEEC, and they also reflect the relative rank of students in a larger area, as the college admission procedure is centralized at the Beijing City level. Table II.5 shows the results of non-parametric estimation of the effects of eligibility on college admission, including qualifications to elite universities, 4-year universities and 3-year colleges. We find some significant positive effects from the whole sample; students attending more selective exam schools on average are more likely to be eligible for colleges at all levels.

II.3.4 Effect of Exam Schools on the Choice of Track and SEEC Participation

Students can strategically sort in terms of self-choice based on other variables. We examine the probability of choosing the science track and the probability of taking the entrance exam to college. In middle school there is no track difference among students, and students have the same curriculum in all schools. The subjects and materials covered on the entrance exam to high school are also the same for all middle school students. High school students have the same curriculum until the end of the first year when students indicate their track preferences. Before the second year and based on their preferences, students are

Table II.5: Effects of Exam Schools on College Eligibility

	Elite			Total		
	Statistics of dependent variable	ITT	LATE	Statistics of dependent variable	ITT	LATE
Elite 4-year	0.144	-0.019	-0.062	0.063	0.040 *	0.133 **
		(0.036)	(0.122)		(0.018)	(0.056)
4-year	0.379	-0.041	-0.133	0.191	0.073 **	0.246 ***
		(0.042)	(0.168)		(0.029)	(0.094)
3-year	0.550	-0.060	-0.195	0.306	0.096 **	0.322 ***
		(0.058)	(0.247)		(0.033)	(0.117)
Sample Size	3053			11616		

Notes: The standard errors which are robust and clustered on school are shown in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% significance levels, respectively. Statistics on the dependent variable are the probabilities of college eligibility. Data source: Administrative data from an anonymous district in Beijing.

re-allocated to the science or art track, with different curricula and exam materials.[14]

Students self-choose the track and decide whether to take the college entrance exam according to their expectation of performance. Table II.6 presents the LATE effect of being enrolled in more selective schools on the choice of track and SEEC participation. We do not find any significant effect of elite school enrollment on choice of track or SEEC participation.[15]

The student's self-choice may be strategic. For example, students who are marginally admitted may be more likely to choose the track with their highest expected score on the entrance exam to college than students who are marginally rejected. Students who are marginally admitted may also be less likely to take the SEEC exam. Their strategic sorting around the cutoff may lead to biased estimates from the RDD model. The results in Table II.6 provide validity checks for potential strategic sorting. We conclude that self-choice is

[14] Appendix C describes the choice of tracks in detail.

[15] We also conduct our main analysis for each track separately. We do not find any heterogeneous effect by track, except for a larger effect from the two-step estimation on SEEC for the art track with the total sample. The results are available upon request.

Table II.6: LATE of Enrollment on the Choice of Track and SEEC Participation

	Being at Science Track (a)		Probability of SEEC Participation (b)	
	Elite	Total	Elite	Total
Effect (β/b)	-0.076	0.096	-0.003	-0.070
	(0.177)	(0.068)	(0.039)	(0.068)
F statistic	3.00	17.97	1.35	5.88
Statistics of dependent variable	0.319	0.387	0.045	0.095
Sample Size	3053	11616	3197	12832

Notes: The standard errors which are robust and clustered on school are shown in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% significance levels, respectively. Statistics on the dependent variable are the probability of being in the science track and the probability of SEEC participation. Data source: Administrative data from an anonymous district in Beijing.

not a threat to the RDD.[16]

II.4 Linking Exam Schools with Student Achievement

We analyze the relation between exam schools and educational inputs to determine the possible reasons why elite schools do not have strong positive effects on achievement while the whole selective school system improves student performance on the college entrance exam on average. Educational outcomes are produced with inputs through a production function. A typical educational production function based on Hanushek (1979) is:

$$SEEC_{ij} = f(B_i^{(t)}, P_{ij}^{(t)}, S_{ij}^{(t)}, I_i, e_{ij}^{(t)}),$$

[16] A remaining issue is that students around the cutoffs may have missing SEEC for different reasons. For example, a student who is marginally admitted chooses not to participate in the exam because he expects a low SEEC, while a student who is marginally rejected is more likely to choose not to participate on the exam for other reasons. Our results may still be upward biased even with a balanced probability of taking the SEEC exam around the cutoff. In our estimation within each school the relationship between SEEH and the probability of taking the exam is insignificant; the magnitude is -0.036 with a p-value of 0.582. This result indicates that the issue that students miss the SEEC for different reasons is not a big concern for our estimation.

where $B_i^{(t)}$, $P_{ij}^{(t)}$ and $S_{ij}^{(t)}$ are accumulated family background characteristics, peer group characteristics and school inputs for student i in school j at the time of the SEEC; I_i is the student's innate ability, which does not change over time; and $e_{ij}^{(t)}$ denotes unobserved variables.

t and t' denote the times when the SEEC and SEEH are measured respectively. At time t' we have a similar educational production function, and we use SEEH to capture the effect of all inputs accumulated until t'. We rewrite the production function in the following value-added form:

$$SEEC_{ij} = f(SEEH_i, B_i^{(t-t')}, P_{ij}^{(t-t')}, S_{ij}^{(t-t')}, e_{ij}^{(t-t')}).$$

We explore two input channels: peer quality and schooling quality. If the changes in these inputs are due to attendance in different high schools, H_{ij}, the production function for SEEC becomes:

$$SEEC_i = f(SEEH_i, B_i^{(t-t')}, H_{ij}, e_{ij}^{(t-t')}).$$

Peer quality is likely to jump when students transit from middle school to high school under the admission rule, especially in the most selective schools. Peer quality can affect achievement from both sides. On one hand, according to the Big-Fish-Little-Pond effect (BFLPE) by Marsh et al. (1995), such a jump in peer achievement is perceived to have a negative effect on the student's own achievement. The marginally admitted students usually are not in the bottom percentage of their middle school class, and those students who get into the most selective schools are the higher-achieving students in middle school. Since they are accepted at the margin compared with their classmates in high school, they are the weaker students in high school and are the least likely to rise to the top. Such a dramatic drop in relative rankings at both the classroom and school levels can harm the academic self-perception of students in these programs and affect their performance on exams; the lower ranked students may under-invest in education (Elsner and Isphording, 2015). On the other hand, higher-achieving peers mean better study groups. Teachers may also cover harder material. Any student can benefit from those features of higher ranked schools.

We use the following definition of peer gap to denote the relative ranking of a student:

$$GAP_{ij} = S\bar{EEH}_j - SEEH_{ij},$$

where $S\bar{EEH}_j$ is the average SEEH of the same track at school j. Column (a) of Table II.7 shows the ITT effect of eligibility on peer quality. Unlike the probability of enrollment, the peer gap has positive and insignificant discontinuity at the cutoff of elite schools. We observe a significantly increased peer gap for the whole sample, but we find only weak jump in peer quality for the most selective schools.

The elite exam schools are perceived to have better facilities, more experienced teachers and more advanced curricula. In Column (a) of Table II.7 we examine three indicators of school quality that are measured at the school level: student/teacher ratio, percentage of teachers with an advanced certificate and the percentage of teachers older than 35. We estimate with Equation (II.4) the ITT effect of eligibility on those educational inputs. We find significant improvement in all three inputs at the cutoffs of elite schools, but we only find weak evidence of change in schooling qualities for the whole sample.

Column (b) shows the LATE effect of educational inputs on SEEC. We obtain those effects by estimating Equation (II.5), but we replace the first stage outcome with the corresponding input. We find strong positive effects of peer gap in SEEH on SEEC. One standard deviation of peer gap can increase the student's own SEEC by about 0.5 standard deviation. We do not find any effect of student/teacher ratio.[17] For the other teacher inputs we find positive effects on SEEC.

The elite schools around the cutoff do have obvious advantages in terms of the educational inputs we examined. According to the effects of educational inputs shown in Table

[17] We note that for the whole sample we do not have a strong effect in the first stage, as indicated by column (a). However, for elite schools which show strong first-stage effects, we find no effect of student/teacher ratio on SEEC.

Table II.7: Roles of Observed Educational Inputs

	Statistics of dependent variable		ITT Effect on Educational Input (a)		LATE Effect of Educational Input (Total) (b)
	Elite	Total	Elite	Total	
Peer Gap in SEEH	-0.060	0.011	0.061	0.217 *	0.495 ***
	[0.327]	[0.456]	(0.080)	(0.103)	(0.128)
Student/Teacher Ratio	11.23	11.50	-0.443 *	0.194	0.554
	[1.852]	[1.805]	(0.241)	(0.320)	(0.913)
Percent of Advanced Certificate	0.359	0.270	0.025 **	0.035	0.031 ***
	[0.077]	[0.127]	(0.011)	(0.020)	(0.009)
Percent of Teachers Older than 35	0.484	0.393	0.034 *	0.029	0.037 **
	[0.098]	[0.135]	(0.017)	(0.021)	(0.015)
Sample Size	3053	11616	3053	11616	11616

Notes: The standard errors which are robust and clustered on school are shown in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% significance levels, respectively. Statistics on the dependent variable are the sample mean and standard deviation (in brackets) of educational inputs. Data source: Administrative data from an anonymous district in Beijing.

II.7, Column (b), the observed better school quality should increase student test scores. One possible explanation is that there are other omitted educational inputs which work in different directions; the net effect of changes in all educational inputs on achievement is zero. For example, parents may invest more in students who are marginally rejected by the elite schools (Pop-Eleches and Urquiola, 2013). Another possible explanation reflects heterogeneous effects of educational inputs by pre-existing achievement. Only students with low pre-existing achievement benefit from increased educational inputs. Our findings on peer effects and observed school quality partially explain the disparity in the effects of elite schools and non-elite schools on student performance.

II.5 Sensitivity Analysis and Heterogeneous Effects

We check whether our main LATE results are robust to different optimal bandwidths calculated with different criteria, different weighting variables and controlling for background variables. We re-estimate the model with different bandwidths – the data-driven

Table II.8: LATE Effect of Enrollment on SEEC, Robustness Checks

	Effect on SEEC										
	Bandwidth								Controlling for Background Variable		
	CCD		CV			Triangle Kernel					
	Elite	Total	Elite	Total		Elite	Total		Elite	Total	
Jump (β/b)	-0.156	0.330	-0.185	0.224	*	-0.178	0.366		-0.364	0.351	*
	(0.694)	(0.231)	(0.635)	(0.120)		(0.513)	(0.226)		(0.385)	(0.199)	
F statistic	8.660	129.8	272.4	192.4		100.6	318.9		2557	373.3	
Statistics of dependent variable	0.383	-0.228	0.356	-0.081		0.313	-0.237		0.313	-0.237	
	[0.705]	[0.887]	[0.725]	[0.957]		[0.763]	[0.900]		[0.763]	[0.900]	
Sample Size	1679	9240	2854	21503		3053	11616		3053	11616	

Notes: The standard errors which are robust and clustered on school are shown in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% significance levels, respectively. Statistics on the dependent variable are the sample mean and standard deviation (in brackets) of SEEC. Data source: Administrative data from an anonymous district in Beijing.

optimal bandwidth by Calonico et al. (2014) and the CV optimal bandwidth used in Ludwig and Miller (2007). We also re-estimate the model using a triangle kernel, which assigns higher weight to students closer to the corresponding cutoff. Finally we add the background variables listed in Table II.2 into the estimation as covariates. The results of LATE estimates are shown in Table II.8. These results are all consistent with the main findings.[18]

We explore the possibility that our effects are a function of the gender of the student and parental education. It is commonly believed in China that boys are relatively good at science and girls are relatively good at art. In our sample girls account for a larger portion of students in the art track than the science track. It is possible that the effects of exam schools are different for boys and girls, and the gender difference in effects largely depends on the track. Effects of exam schools may also vary among students from different family backgrounds, and we focus on two types of families: parents with a college degree and parents who are farmers.

We re-estimate the LATE effects over the whole sample for each subpopulation char-

[18]In some specifications we lose significance of the effect of non-elite schools. Since the sizes of the effect are similar with our main findings, we view the insignificance as a power issue and still conclude that our results are robust.

Table II.9: Effect of Enrollment on SEEC, Heterogeneous Effects

	Effect on SEEC (Total)					
	Gender		Father Bachelor Degree		Mother Bachelor Degree	
	Male	Female	Yes	No	Yes	No
Effect (β/b)	0.246	0.510 *	0.371	0.376 ***	0.007	0.381 **
	(0.164)	(0.263)	(0.927)	(0.145)	(0.503)	(0.183)
F statistic	226.0	34.81	340.2	93.96	1221	85.99
Statistics of dependent variable	-0.174	-0.314	-0.06	-0.278	-0.058	-0.205
	[0.882]	[0.917]	[0.923]	[0.890]	[0.906]	[0.872]
Sample Size	6384	5232	2195	9421	1711	9262

Notes: The standard errors which are robust and clustered on school are shown in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% significance levels, respectively. Statistics on the dependent variable are the sample mean and standard deviation (in brackets) of SEEC. Data source: Administrative data from an anonymous district in Beijing.

acterized by gender or parental educational background, and the results are shown in Table II.9. We find heterogeneous patterns in several cases; girls benefit more by attending better schools with a higher cutoff. Students from families in which the parents do not have a college degree are also more likely to benefit from attending more selective exam schools, especially students whose mothers have no bachelor degree.

II.6 Conclusion

Selective exam schools that admit students solely by exam scores are established in many countries across the world. The goal of these schools is to improve student outcomes including college entrance exam scores and college enrollment. There is no consensus among education policy researchers about the effect of selective exam schools on student achievement. In Romania (Pop-Eleches and Urquiola, 2013), Trinidad and Tobago (Jackson, 2010) and rural China (Park et al., 2015) exam schools improve student achievement while in the United States (Abdulkadiroglu et al., 2014; Dobbie and Fryer Jr., 2014), the United Kingdom (Clark, 2010), Kenya (Lucas and Mbiti, 2014) and Ghana (Aiayi, 2014) there is no evidence of positive effects from attendance in exam schools.

We use regression discontinuity design and examine the effect of exam schools in urban China; these schools include two elite schools which are of higher quality and more selective than other schools. We find no effect of these schools on test score performance. Elite exam schools have better school quality in several observed measures, but we find little evidence of positive effects of those measures in the elite schools. However, we do find some evidence that on average attending a more selective non-elite school increases student scores on the college entrance exam. Peer effects are important to these findings. The RDD identification across settings contributes to the mixed findings of the effect of elite schools. The evidence of heterogeneous effects by gender and parental education suggests that the current exam school on average can eliminate the gender gap in achievement and help students from disadvantaged backgrounds catch up to the academic performance of wealthier students.

Our evaluation of exam schools casts doubt on the policy of labeling schools by selectivity. The quality of schools is more important for improvement in student test scores than the elite school label which in the extreme may only serve as a signal for positive ability matching between students and schools.[19] Our study does not prove that students who attend elite exam schools do not experience any benefits from their high school education. First, the primary outcome we examine in this paper is the college exam score, but elite schools can help students in other ways that are not captured by scores. Students in elite schools may make influential friends and benefit from their network affiliations. Attending an elite school can also have long-term effects on education completion, income and fertility (Clark and Emilia, 2016). Second, the RDD can only provide the estimated effect of elite schools on a subsample of students – compliers at the cutoff. These are the students whose SEEH is at the cutoff and will surely attend elite schools if they are eligible; they will never attend if their SEEH is below the cutoff. Using the RDD framework, we can say nothing precise about the effect on other students who may benefit from exam schools.

[19]Even the signal may work only in the short-run. In the long-run parents and students can recognize the true quality by other indicators, such as the performance of students on exams in previous years.

Duflo et al. (2011) find that teachers in elite schools may pay more attention to median students than to students who are marginally admitted. Finally, because of the data limitations we are not able to identify the effect of other unmeasured school quality indicators, such as capital expenditure, which may have significant effects on student achievement. We are not able to examine student psychological and behavioral attributes during their transition to elite high schools. Those unmeasured personal educational inputs may work through different channels than academic performance. These other effects are important for the overall evaluation of elite school policy.

Although we show that disadvantaged students can potentially benefit from attending more selective non-elite schools, our results are disappointing for many parents. There is no conclusive evidence that elite exam schools improve a students score on the entrance exam to college. A good college exam score is probably the most important dimension of high school achievement as the score directly determines the ability of a student to get into a prestigious college and her future career. The government realized that the elite schools may not work as expected and is considering several admission and school selection reforms. These proposed reforms include the addition of more seats in elite schools, investment in school quality in suburban (and rural) areas, and reservation of a percentage of seats (30%-50%) in elite schools for students from non-elite middle schools (Beijing Municipal Commission of Education, 2013). These reforms would increase the diversity of students in elite schools and the advantage that students from elite schools have in college admission.

Chapter III

How Does Grade Configuration Impact Student Achievement? Evaluating the Effectiveness of K-8 Schools

In recent years, a number of districts have moved towards the use of K-8 schools instead of stand-alone middle and elementary schools. In part, this movement has been in response to research suggesting middle school students often become disengaged resulting in stagnant student learning and discipline problems and could ultimately lead to an increased likelihood of students dropping out during high school (Juvonen et al., 2004). Some believe that middle school students are vulnerable to risks of poor long-term outcomes because of physical, emotional, and intellectual changes they experience during these years (Juvonen et al., 2004). Many educators believe that a K-8 environment can address some of these issues for middle school students by alleviating a transition between elementary and middle schools.

However, this argument does not consider the possible consequences for elementary students in a K-8 environment. One could argue that a policy move to K-8 schools could adversely affect elementary students as they will be exposed to much older students, which could create an intimidating environment. In addition, a move to K-8 could adversely affect students of all grade ranges if the change to K-8 school leads to a larger school with less intimate relationship between students/families and teachers, which could be especially detrimental to elementary students (Feldlaufer et al., 1988; Midgley et al., 1989). Furthermore, combining students who are at different developmental stages in a K-8 setting may not allow schools to specialize in appropriate developmental environments for either middle or elementary grade students. Finally, separate middle schools can develop instructional and pedagogical strategies best suited for middle school aged students (Hough, 2005).

Advocates counter that eliminating the transition from elementary to middle schools can reduce stress for students, who are already feeling stress from social and biological changes from the onset of puberty (Eccles et al., 1984; Eccles and Midgley, 1989; Juvonen et al., 2004). This argument has some support as Elias et al. (1985) found that students report a high level of stress from the complex, new social world in middle school. Furthermore, research by Rudolph et al. (2001) shows that students with maladaptive self-regulatory beliefs, such as decreased perception of academic control and importance, report more pressure during the transition to middle schools. Therefore, by reducing student stress, learning may improve for students. Advocates also argue that teacher and student relations change during the transition from elementary school to middle school. Research suggests that teachers in middle schools are on average less caring, friendly and supportive than their counterparts in elementary schools (Feldlaufer et al., 1988). Deterioration in teacher/student relations in middle school could adversely affect students (Midgley et al., 1989) because positive relations with adults other than parents are important to the social and emotional development of young adolescents (Miller, 1970). These theoretical claims by advocates are supported by empirical research on student mobility, which suggests that student moves between schools can have adverse impacts on students (Hanushek et al., 2004; Schwartz et al., 2011; South et al., 2007; Xu et al., 2009).[1][2] Because K-8 schools could eliminate a move from an elementary to middle school, there could be positive effects for students in K-8 schools. However, all of these arguments for a K-8 school focuses on middle school aged children without consideration of the effects on elementary aged children.

[1]However, it should be noted that there are two types of mobilitynon-structural and structural mobility. Non-structural moves are the result of student choice. Students move to another school because of observed or unobserved preferences. Structural moves are related to grade configuration. Students move to another school because they finished the terminal grade at their current school and have to start the next grade in another school in a higher grade. It is important to distinguish between the two types of mobility as their policy implications are different. In this study we focus on the effect of structural moves.

[2]Despite this general evidence on student mobility, the existing literature paints an ambiguous picture on the transitional effect from elementary to middle schools. Some studies find adverse effects of mobility from elementary school to middle school (Bedard and Do, 2005; Cook et al., 2008; Schwartz et al., 2011) while other studies find no effect or a positive effect on achievement (Gunter and Bakken, 2010; Lippold et al., 2013; Weiss and Kipnes, 2006).

Therefore, it is not surprising that the current research has exclusively focused on the effect for middle school students. In addition, most of this research has not dealt with the non-randomness of students choosing to enroll in K-8 schools versus separate elementary and middle schools. A simple pairwise comparison in achievement between the students from the stand-alone middle schools and K-8 schools or a comparison of stand-alone elementary schools and K-8 schools is insufficient for us to draw any conclusion about the causal effects of attending a stand-alone middle and elementary schools, because the grade arrangement of the school could affect the choice of school by families. Such self-selection can lead to biased results as their choice to attend a K-8 school may be driven by non-random factors, including the hopes of gaining a better academic experience for the student.

Three recent studies (focusing only on middle school students) use an instrumental variable approach to address the possible endogeneity (Dhuey, 2013; Rockoff and Lockwood, 2010; Schwerdt and West, 2013). More specifically, these sets of authors instrument for the grade configuration during middle school years with the grade configuration of the school that the child attended in grade 3 or 4. The essential assumption for the instrument is that any unobserved shocks to achievement during the transition are not anticipated nor do they affect the choice when in grade 3 (or 4) of a school with a specific grade configuration. However, if this second assumption does not hold, their instrument only accounts for a switch to a K-8 school after grade 3 (or 4) and does not account for the possibility that parents choose to send their child to a K-8 or separate elementary/middle school sequence prior to that time.

This current paper contributes to the existing literature by not only examining whether the previous findings for the effect on middle school students hold up when employing an alternative identification strategy, but also examine the effects for elementary students. More explicitly, we employ a geographic quasi-experimental design much like a regression discontinuity approach taking advantage of school closures in an anonymous midsize district. Using this alternative approach, much like the Rockoff and Lockwood (2010), Schwerdt and West (2013) and Dhuey (2013) papers, we find some evidence for adverse transitional effects of attending a middle school. However, unlike these papers, we do not find long-term effects from attending a middle school. In addition, in our analysis of students in elementary grades, we find evidence that elementary grade students perform better in a K-5 school than a K-8 school, which was not only not examined in previous studies, but not considered. Therefore, we argue that our study not only provides a strong identification strategy for analyzing the effects of K-8 schools on middle school students, but it broadens the scope of knowledge by examining the effects of K-8 for students in elementary grades.

III.1 Research Approach

For our analysis, we use a geographic quasi-experimental approach that mimics a geographical regression discontinuity design, which is similar to an approach used by Black (1999) to estimate the capitalized value of education in homes. As noted above, a few previous papers have used an instrument for the grade configuration during middle school years with the grade configuration of the school that the child attended in grade 3 (or 4), which may control for students switching between separate elementary/middle schools to a K-8 school (or vice versa) (Dhuey, 2013; Rockoff and Lockwood, 2010; Schwerdt and West, 2013), but cannot address the original choice to attend a separate elementary/middle school or K-8 school. Our approach addresses the endogeneity due to both students who chose to switch between schools as sixth grade approaches and students who chose one of these school types when they enter the school system.

Specifically, to account for the original decision of whether to attend a separate elementary and middle school versus a K-8 school for both our analysis of stand-along middle and elementary versus K-8 schools, we leverage the fact that the anonymous district closed 20 schools (or about one-fourth of all schools) at the conclusion of the 2005-06 school year because of accumulated surplus capacity. Many closed middle schools were replaced by expanding existing elementary schools from K-5 to K-8 with the hope that the shift in grade configurations would reduce the disruption often associated with switching schools. As a result, there were 13 new K-8 schools at the beginning of the 2006-07 school year with new geographic boundaries. In addition, because of the new geographic boundaries, a number of students were reassigned from possibly going to a separate middle school to one of 14 existing K-8 schools and vice versa. These new school boundaries provide an opportunity for a strong identification strategy as we can compare students on one side of the boundaries attending a stand-alone middle or elementary school to students on the other side of the boundaries attending a K-8 school. In essence, this mimics a "spatial" regression discontinuity (RD) approach where equivalent comparison groups are created. We can observe students in neighborhoods that were originally assigned to the same school before the closure and assigned different schools after the closure. From these patterns, we can observe pairs of students who live very close to each other that were assigned to the same school before the closure but were assigned to different types (K-5 vs. K-8) of schools after the closure. They should have similar observed and unobserved characteristics, including similar preferences for the various types of school configurations. By comparing their achievements we are able to obtain the causal effect of the K-5/middle configuration relative to the K-8 configuration.

We should note that using long-term existing school boundaries most likely would not create equivalent groups as many families may choose their residence on one side of the boundary based upon the characteristics of the school offerings, including whether it is a K-8 school. However, in this case, the new boundaries were not anticipated by parents as the redrawing of school boundaries occurred in the spring prior to the 2006-07 academic year. In addition, unlike most districts in which the closing of schools can be influenced by political pressures, the board members accepted the superintendent's request to stay removed from the closing and reassignment process. Instead, the closures and reassignments were developed by administrators whose goal was to exactly fill the seats in the schools that remained open. In other words, the board members gave an up and down vote for the process and not the closing of individual schools or the reassignments of individual students. Therefore, the board members did not know which schools would be closed or which families would be affected when they were voting and less susceptible to political pressures from families and voters. Thus, the rezoning can be viewed as an exogenous shock creating an opportunity to create strong comparison groups.

III.2 Data

For the study, we have access to the anonymous district's longitudinal student-level data warehouse. The district has maintained student-level data on enrollment, demographics, residential location, and student outcomes since the late 1990s. The data set links student outcomes with the student's grade and school of attendance, neighborhood characteristics (based on student's residential location), time varying student characteristics, including free- and reduced-price lunch (FRL), limited English proficiency (LEP), and special education status, as well as time-invariant demographics including gender and race/ethnicity. One of the features of the data system that is fairly unique is the rich geographic information, including the addresses of students and schools, and school feeder patterns. In addition, information from Census and American Community Survey tract summary data has been linked to each student using their address for each year. Finally, the data set includes scores for the standards-based accountability tests and the district administered nationally-normed Terra Nova test. Up to the 2003-04 school year, the standards-based accountability test was administered in fifth, eighth, and eleventh grades. Third grade was added in year 2004-05 and expanded the next year to include all grades between third and eighth grade as well as eleventh grade. For the Terra Nova, we have first and second grade math test scores since 2002-03 and reading since 2003-04, which allows us to have a measure of student performance prior to grades tested for the state accountability test. While the Terra Nova is not psychometrically aligned with the state standards based accountability test, the two tests are highly correlated. To have the two tests on the same scale, we normalize these test scores by subject, grade and year with a mean of zero and standard deviation of one.

For the analyses, we include students that meet the following conditions: (1) the student was in grade 2 to 4 in the 2005-2006 school year, (2) the student attended either a K-5 school or a K-8 school in 2005 and 2006,[3] (3) the student had test scores in at least one grade and had no grade skipping or retention through 2005 and 2006, and (4) the student lived in the same place in 2005 and 2006.[4] In total there are 4,031 students that met these conditions. For these students, the average age in 2005 is 9.64 and 52 percent are male. In terms of race, 44 percent of the students are white, 47 percent are black and the remaining 9 percent are of other races. 58 percent of the students were in a K-5 school in 2005 with the remainder attended a K-8 school.

Table III.1 provides the number and percentage of first through fourth grade students in 2005-06 school year that were reassigned to and from K-8 schools as well as to and from K-5 schools as a result of the new school boundaries. These students make up our baseline set of students for the geographical analysis. We divide the students into those previously assigned to a continuing school (i.e., a school that continues to operate) and those assigned to a closing school. We present the number and percentage of students reassigned from a continuing K-5 or K-8 schools to another K-5 or K-8 school, how many are keeping their assignment at their continuing K-5 and K-8 schools, and how many are reassigned from closed K-5 or K-8 schools. In total, 270 students were reassigned from a K-5 school to a K-8 school while just 80 students were reassigned from a K-8 school to a K-5 school.

[3]There are elementary and middle schools within the district that start in grade other than Kindergarten or grade 6. We exclude them from the analysis, as we want to focus on the comparison between K-5, K-8 and stand-alone middle school (6-8).

[4]In total there 4,843 students enrolled in grade 2 to 4 in either a K-5 school or a K-8 school in 2005. Among those students who satisfy the condition (1) and (2), there are 2 students whose test scores are completely missing, 67 students who had grade skipping or retention through 2005 and 2006, and 36 students who lived in different places in 2005 and 2006. We do not require the data to be a balanced panel data, which is used in the previous studies, as we want to maximize the sample.

Table III.1: Amount of Students by the Type of Assignment

Assignment before school closure	Assignment after school closure	Number of Students			Percentage of the total students		
Continuing Schools		3,120			77.4%		
Continuing K-5			1,624			40.3%	
	Another K-5			65			1.6%
	K-8			14			0.3%
	The Same K-5			1,545			38.3%
Continuing K-8			1,496			37.1%	
	K-5			21			0.5%
	Another K-8			42			1.0%
	The Same K-8			1,433			33.5%
Closing Schools		911			22.6%		
Closing K-5			733			18.2%	
	Another K-5			475			11.8%
	K-8			258			6.4%
Closing K-8			178			4.4%	
	K-5			59			1.5%
	Another K-8			119			3.0%
Total Students		4031			100%		

Notes: Only students who meet the conditions (1)-(4) on Page 7 are included.

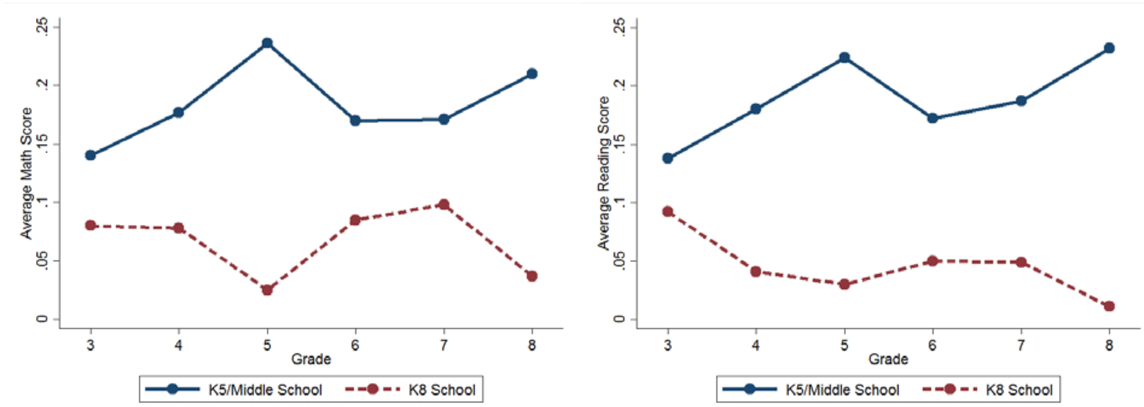
III.3 Implementation of Geographic Quasi-Experiment

Figure III.1 displays a simple and nave comparison in achievement after school closure in 2006 between the students from the K-5 school/stand-alone middle schools and K-8 schools. The students in K-5 schools outperforms in the elementary grades in both of math and reading. The gap is increasing until grade 5, when the students in K-5 schools have to move to another school to continue grade 6. For both of math and reading there is a significant drop in test score during the transition to middle school for the students in K-5 schools. At the same time the performance of the students in K-8 schools improves. However, the transitional effects are temporary. The gap in test scores of both math and reading between the students in stand-alone middle schools and K-8 schools starting increasing again at grade 7.

Figure III.1 shows us trends of achievements of the students in K-5/middle schools and

Figure III.1: Student Achievement in K-5/middle schools and K-8 schools

(A) Math (B) Reading



Notes: Only students who meet the conditions (1)-(4) on Page 7 are included.

K-8 schools at grade 3 to 8. The trends for math score and reading score are similar: the students in K-5 schools outperform the students in K8 schools at grade 3, and the gap in achievement becomes larger until grade 5, after which the students in K-5 schools transfer to middle school. The gap shrinks substantially during the transition, but still remains and increases again in middle school grades. The figure shows some descriptive evidence that K-5 school may benefit elementary students and does not have long-term negative effects beyond a negative shock during the transition to middle school. Nevertheless, we are not able to interpret those trends as causal effects of K-5 schools, because self-selection of school types can lead to biased results; for example, parents who anticipate low performance at middle school will send their kids to K-8 schools at the beginning when they enter the educational system.

As noted previously, we take the advantage of school closure program to solve the endogeneity problem of self-selection of school types. Specifically, using the newly formed school boundaries, we create comparison groups for our analysis. Two students who lived close to each other and thus previously were assigned to the same schools before the closures, but were assigned to different schools of different types after the closures, should be very similar in terms of both observed and unobserved characteristics. However, like in a regression discontinuity analysis, the broader the bandwidth (in this case, measured by geographic distance), the more likely the groups will differ in observed and unobserved ways. This is illustrated in Table III.2 in which we conduct a balance check analogous to the balance check used in experimental analysis. In the table, we report coefficient of pairwise estimates of observable characteristics for students within the bandwidths of 0.5 and 0.3 miles.[5] [6] Within the geographic bandwidths, we identify students who lived within these bandwidths of at least one other student who was previously assigned to the same school but was reassigned to a different type of school (separate K-5 and middle schools versus K-8 schools). In both cases, we have good overlap with only one statistically significant difference. We use both samples for our analyses to examine the robustness of our results.

Intent-to-treat analysis

Like many treatments, some students do not comply with their assignment. For instance, the district has magnet schools, which allows students to opt out of their assigned schools. Therefore, we first do an intent-to-treat (ITT) analysis, which examines whether merely being assigned to a K-5/6-8 or a K-8 school has an impact on student achievement. We estimate the ITT effect of being assigned to a K-5/6-8 sequence of schools, by fitting the following equation:

$$y_{igtp} = a_i + d_p + S^{gt} + S^{gtA5} + e_{igtp}, \tag{III.1}$$

where y_{igtp} is the achievement of student i in grade g, in period t, and is a member in pair p. We define period as either prior to school closures ($t = 0$) or after school closures ($t = 1$). The parameters a_i and d_p are the individual and pair fixed effects respectively. The parameter S^{gt} is a fixed effect for each combination of grade and period. The parameter

[5] We did not include a smaller bandwidth of 0.1 because the sample size becomes too small. We also conduct analysis using bandwidth of 0.4 and 0.2 miles and obtain similar results, which are available upon request.

[6] In total there are 1436 such students reassigned to 28 schools for 0.5-mile distance and 969 students reassigned to 27 schools for 0.3-mile distance.

Table III.2: Balance Check of Observable Characteristics

	0.5 Mile		0.3 Mile	
	Balance Check		Balance Check	
Male	0.000		0.001	
White	0.057		0.040	
Black	-0.060		-0.058	
Percentage of free or reduced price lunch	-0.097	***	-0.163	***
Percentage of limited English proficiency	0.003		0.016	
Percentage of gifted student	0.014		0.030	
Percentage of student with special need	0.032		0.034	
Moving from another school	0.034		-0.004	
Math Score, 2005	-0.022		0.021	
Reading Score, 2005	-0.016		0.023	
Sample size	12385		6049	

Notes: The coefficient estimates are the results of balance check by our main regression controlling for pair dummy. *** indicates significance at 1% confidence level.

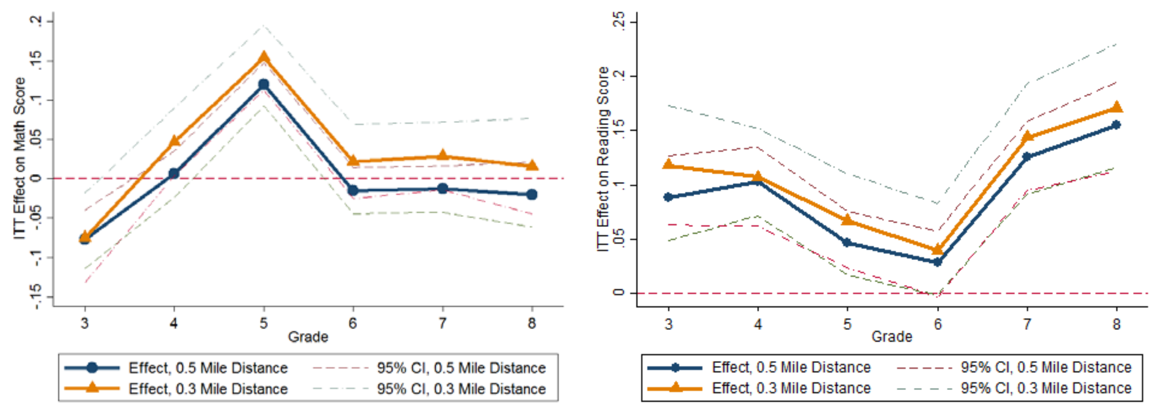
S^{gtA5} is a fixed effect for students assigned to K-5/6-8 schools in 2006, for each combination of grade and period. This parameter is the ITT effect, as it estimates the difference in relative achievement for the students assigned to K-8 following the closures.

In Figure 2 (and Appendix Table A.7), we present the ITT estimates of being reassigned a K-5 school in grade 3 to grade 8. Focusing on the effect for elementary school student first, while we do not find positive or negative math effects in grade 3, the math results generally suggest that there are statistically positive effects of large magnitude for students in fourth grade and these effects generally linger into fifth grade and becomes larger. The reading results consistently show positive effects in all three grades.

For the effect for middle school students, just as the case for Rockoff and Lockwood (2010), Schwerdt and West (2013) and Dhuey (2013), being assigned to a stand-alone middle school is the treatment. Therefore, a negative effect would suggest that there are adverse effects from being assigned to a middle school relative to a K-8 school and would

Figure III.2: Impact on Achievement of Being Assigned to K-5/6-8 School

(A) Math (B) Reading



Notes: The figure shows the coefficient and 95% confidence interval for ITT effect of being reassigned to a K-5 School on test scores of math (panel (A)) and reading (panel (B)). The specification is the ITT regression described in equation (III.1).

provide support for a K-8 policy. The opposite would be true for positive effects. Overall, the ITT results provide some evidence of a negative transition effect to a stand-alone middle school in grade 6, which supports the results found in the previous studies using the IV approach. However, unlike these previous studies, we do not observe persistent long-term effects in grades 7 and 8. In fact, we observe some statistically positive effects. These results raise questions about the robustness of the results for middle school students and the policy implications from the previous papers.

Treatment-on-treated analysis

The ITT are of interest because they capture the average impact of the policy overall of the middle school students potentially affected by it. However, the ITT results do not measure the effect of the policy on students actually complying with their assignment.To estimate the effect for students who actually enroll in a K-5/6-8 sequence of schools, we have to account for non-compliance using a treatment-on-treated (TOT) analysis. We take an approach similar to a "fuzzy" RD design and use school assignment based on the ge-

Table III.3: Percentage of Compliers by Distance

	Bandwidth					
	0.5 mile			0.3 mile		
	Total	K-5 Assignment	K-8 Assignment	Total	K-5 Assignment	K-8 Assignment
Overall	81.5%	86.7%	76.1%	81.5%	86.8%	75.2%
Grade in 2005 (immediately prior to school closures):						
Grade 2	81.3%	87.1%	74.3%	81.1%	87.3%	72.1%
Grade 3	81.8%	87.3%	77.0%	82.6%	87.3%	77.8%
Grade 4	81.4%	85.7%	76.9%	80.9%	85.7%	75.4%

Notes: The table reports the percentage of students following the school reassignment in terms of school type.

ographic boundaries as an instrumental variable to estimate the causal effect of enrolling in a middle school. More formally, to estimate the TOT effect, we use the indicator of assignment as an instrumental variable for enrollment and fit the following equations:

$$\begin{cases} E5_{ip} = f_i + l_p + S^{gt} + S^{gtA5} + u_{ip} \\ y_{igtp} = a_i + d_p + S^{gt} + S^{gtE5} + e_{igtp}, \end{cases} \tag{III.2}$$

where $E5_{ip}$ are the interactions of the indicator of whether student i enrolls in a K-5 school in 2006 after the school closure with indicators for grade, for time or both.

Table III.3 summarizes the compliers who follow the school reassignment after school closures by the neighborhood defined by distance. Overall more than 80% of the students follow the reassignment in terms of school type. Students who are reassigned to a K-5 school are more likely to enroll in a K-5 school in the following year. There is no clear heterogeneous patterns of compliers by grade in 2005. These results indicate that the school reassignment regarding school type is a strong predictor of the type of schools in which the students are actually enrolled right after the school closure.

Table III.4 reports regression results based on a simplified version of the first stage

(III.2).[7] The results confirm that reassignment is predictive of the type of school a student attends after school closure, although there are students who do not end up following their assigned paths. Overall being assigned to a K-5 school increases the probability of being actually enrollment in a K-5 school by about 60%, which does not depend on the grade in 2005 much. Therefore, the TOT results can be interpreted as local average treatment estimates.

Again, enrolling in a K-5/6-8 school in 2006 is the treatment and the results are shown in Figure III.3 and Appendix Table A.8. Therefore, a negative effect would suggest that there are adverse effects from enrolling a K-5/6-8 school relative to a K-8 school and would provide support for a K-8 policy. The opposite would be true for positive effects. As with the case for the ITT analyses, we observe positive effect on both of math and reading in grades 3 to 5. For math we find some evidence of a negative transition effect to stand-alone middle schools. For reading the transition effect is much smaller in magnitude and statistically insignificant. However, like the ITT analyses, we observe positive effects by grade 8, especially in reading. Again, these results are in contrast to the results from previous literature using IV approach, which showed long-term negative effects from enrolling in a middle school.

Overall, while these results provide some evidence of adverse transition effects for those students transitioning to a stand-alone middle school, our analyses suggests that students perform better in K-5 schools at the elementary level and do not show long-term adverse effects from attending a middle school. Therefore, when controlling for the possible endogenous selection at the entry grade of the school system as well as for switching between schools after third grade, our geographic quasi-experimental approach shows less support for K-8 schools than the previous research which accounted only for the endoge-

[7] The simplified first stage is a cross-sectional version of equation (III.2) without all grade dummies, time dummies and their interactions with reassignment or actual enrollment. Individual and pair fixed effects are also omitted. In the actual first stage there are eight endogenous variables: six interactions of grade and actual enrollment (grade 3-8), one interaction of time and actual enrollment, and one interaction of grade 4, time and actual enrollment. As a result there are eight first stage equations. Results from the actual first stage regressions are available upon request.

Table III.4: Effect of Assignment to K-5 Schools on Enrolling in a K-5 School

	Bandwidth	
	0.5 mile	0.3 mile
Overall	0.628 ***	0.620 ***
	(0.021)	(0.026)
R-squared	0.400	0.393
F statistic	918.0	577.9
Sample size	1390	932
Grade in 2005 (immediately prior to school closures):		
Grade 2	0.614 ***	0.594 ***
	(0.037)	(0.046)
Grade 3	0.643 ***	0.651 ***
	(0.035)	(0.043)
Grade 4	0.626 ***	0.611 ***
	(0.037)	(0.046)

Notes: Clustered standard errors by student and pair are reported in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% confidence level respectively. The specification is the simplified first-stage regression described in footnote 6.

nous switching after third grade. In fact, our results suggest that such a policy would have adverse effects for elementary students.

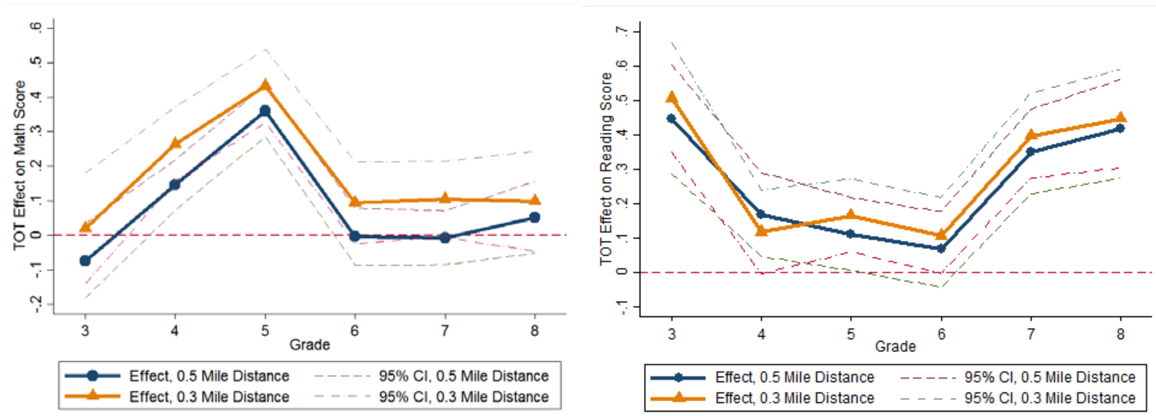
III.4 Robustness Check

III.4.1 Identifying Pair by Actual Enrollment

In our analysis so far, we identify the pair of students by choosing the students who were assigned to the same school before the school closure but were reassigned to different

Figure III.3: Impact on Achievement of Being Actually Enrolled in K-5/6-8 School

(A) Math (B) Reading



Notes: The figure shows the coefficient and 95% confidence interval for TOT effect of being reassigned to a K-5 School on test scores of math (panel (A)) and reading (panel (B)). The specification is the TOT regression described in equations III.2.

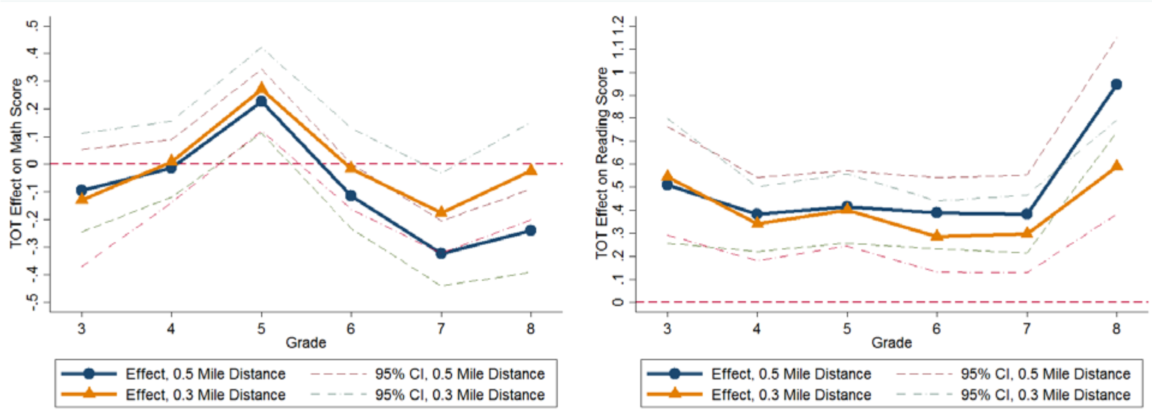
types of schools after the school closure. There are alternative ways to identify pairs. For example, those students assigned to the same school in 2005-06 school year were not necessarily enrolled in the same school. The students living close to each other and assigned to the same school are very likely to be similar in both of observed and unobserved characteristics. However, this may not be the case if one student does not enroll in the assigned school, exercising choice to enroll in another school instead. In Figure III.4 below (for details see appendix Table A.9), we re-define the pairs of students by choosing those who indeed were enrolled in the same school before the school closure and keeping all of the other filters. Despite of different magnitude, the overall pattern of the estimated effects is consistent with our previous findings.

III.4.2 Mixed Effect of Moving and School Type

One weakness of the existing literature is that the estimated effect during the transition is a mixture of effects of transition generally and effects of different school types (e.g.,

Figure III.4: Impact on Achievement of Being Actually Enrolled in K-5/6-8 School, Paired by Prior Actual Enrollment

(A) Math (B) Reading



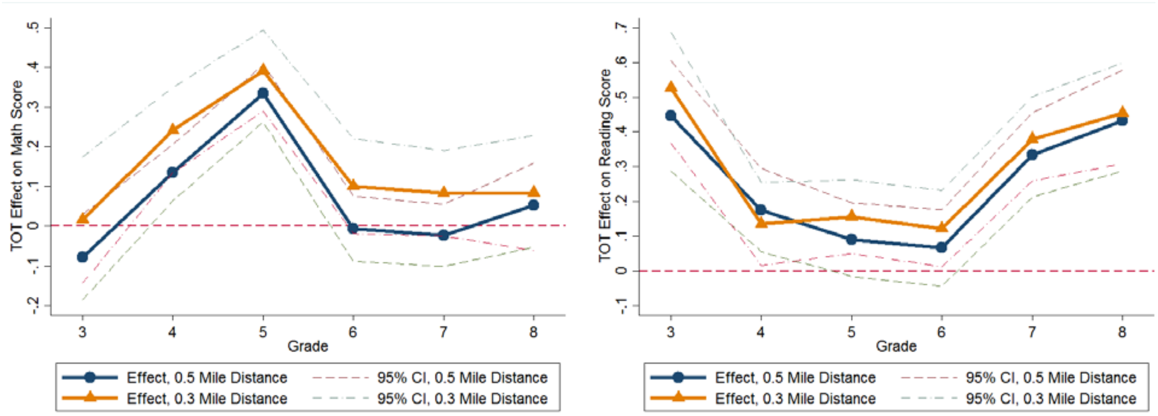
Notes: The figure shows the coefficient and 95% confidence interval for TOT effect of being reassigned to a K-5 School on test scores of math (panel (A)) and reading (panel (B)). The specification is the TOT regression described in equations III.2.

poor learning environment in middle school). In the previous studies exogenous school transitions are collinear with attendance in middle schools. In this paper the students are force to move between schools because of the school closure, independently of the new schools grade configuration. In the previous balance check (Table III.2) we do not find significant difference in the probability of moving to another school between the students rezoned to different types of schools. This indicates that our main results are not biased due to forced school transition. Nevertheless, the school closure program provides us an opportunity to separate the effect of transition from the effect of attending different types of schools. In particular, we estimate the following variation of equations III.2:

$$\begin{cases} E5_{ip} = f_i + l_p + S^{gt} + S^{gtA5} + M^{gtM} + u_{ip} \\ y_{igtp} = a_i + d_p + S^{gt} + S^{gtE5} + M^{gtM} + e_{igtp}, \end{cases} \tag{III.3}$$

Figure III.5: Impact on Achievement of Being Assigned to K-5/6-8 School Conditional on Transition

(A) Math | (B) Reading



Notes: The figure shows the coefficient and 95% confidence interval for TOT effect of being reassigned to a K-5 School on test scores of math (panel (A)) and reading (panel (B)). The specification is the conditional TOT regression described in equations III.3.

where M^{gtM} is a fixed effect for students moved to another school due to the school closure, interacted with grade. By conditioning on the indicator of school transition, we identify the effect of attending different types of schools after accounting for the transition shock from the school closure.

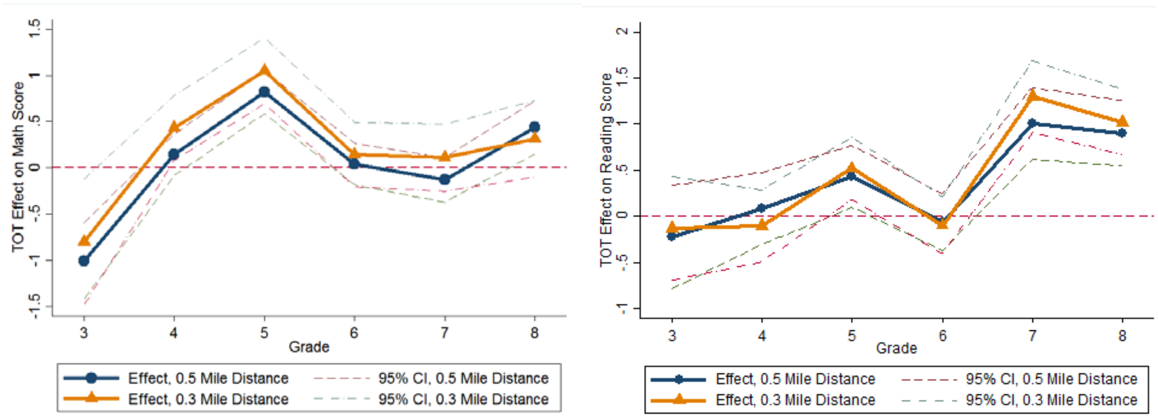
Figure III.5 shows the conditional effects of being reassigned to a K-5 school after school closure (for details see appendix Table A.10). The effects conditional on whether moving to another school are quite similar with the main results presented in Figure III.3. Our effects are not driven by unbalanced school transition between the treatment and control groups.

III.4.3 Effect of Actual Moving to Middle School

We estimate the effect of being reassigned to a K-5 school after school closure, including the TOT effect which takes into account the possibility that students may not follow the

Figure III.6: Impact on Achievement of Moving to Middle School

(A) Math (B) Reading



Notes: The figure shows the coefficient and 95% confidence interval for TOT effect of moving to middle schools after grade 5 on test scores of math (panel (A)) and reading (panel (B)). The specification is the TOT regression described in equations III.2, replacing the treatment of being enrolled in a K-5 school in 2006 with moving to a middle school after grade 5.

reassignment. For the middle school grade, there is another type of non-compliers: students follow the reassignment to a K-5 school can move to a K-8 school later, e.g., in 2007, to avoid the transition to middle schools. We estimate the effects of moving to middle school after grade 5 on subsequent student achievement by equation III.2, replacing the treatment of being enrolled in a K-5 school in 2006 with moving to a middle school after grade 5. The results are presented in Figure III.6 (for details see appendix Table A.11).

We should note, we are not able to interpret the effects at elementary grades 3 to 5 as causal effects as they emerge before the treatment. Focusing on the effects after the transition to middle school, we find negative transitional effect on both subjects after grade 5. In other words, moving to middle school indeed harm student' performance. However, as we discovered before, such negative effects of moving do not linger, especially for reading.

III.4.4 Non-moving K-8 Student

One possible critique of our geographic approach for the middle school analysis is that inferences regarding the impact of attending a K-8 versus a stand-alone middle school may be limited to situations in which many existing schools are closed and capacity in K-8 schools is expanded. The shift to K-8 schools in the study district was part of a larger reform of closing schools, which required many students to transition to a new school. Therefore, our geographic analysis may only have implications for a policy that causes students to transition to new schools. Moreover, for our findings there are two possible alternative explanations other than the ineffectiveness of K-8 schools. First, one strength of K-8 schools is that they eliminate the transfer to middle school. However, under the school closure policy many of the students who were assigned to K-8 schools also moved. Since they miss a key benefit of K-8 schools, they do not outperform the students entering stand-alone middle schools. Second, since many new K-8 schools are expanded from K-5 schools, those K-8 schools may serve students in grade 6-8 schools worse than a K-8 schools which has been established for a long time.

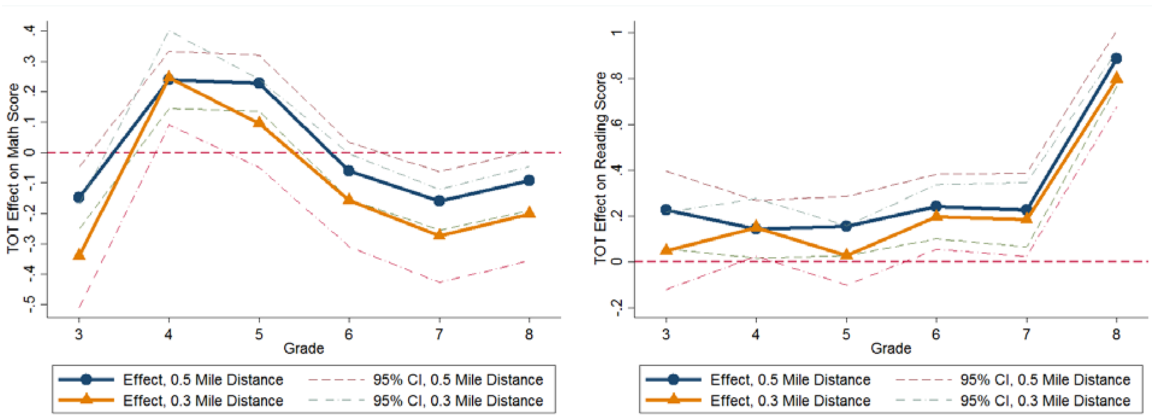
To address those two alternative explanations, we redo our main analyses using newly defined pairs of students. In each pair one student stayed in the same K-8 school after the reassignment and the other was reassigned to a K-5 school. Since both students were enrolled in the same pre-existing K-8 school and the student who was assigned to K-8 school did not move, the results from these pairs should provide more general evidence about the effectiveness of K-8 school.

Figure III.7 shows the effect of being assigned (ITT) in a stand-alone middle school using the pairs in which one student stayed in the same K-8 school after the school closure (for details see Appendix Table A.12).[8] Overall the pictures of math scores show similar

[8] With the restricted sample we do not observe enough compliers, thus are not able to do the TOT analysis reliably. However, we know that TOT effect should have the same sign as ITT effect as long as on average the assignment leads to higher probability of treatment, which is the case here. Thus the TOT effects of enrolling in K-5 schools should show similar patterns as the ITT effects.

Figure III.7: Impact of Being Assigned in a K-5 School on Achievement, Nonmoving K-8 Student

(A) Math (B) Reading



Notes: The figure shows the coefficient and 95% confidence interval for TOT effect of moving to middle schools after grade 5 on test scores of math (panel (A)) and reading (panel (B)). The specification is the TOT regression described in equations III.2, replacing the treatment of being enrolled in a K-5 school in 2006 with moving to a middle school after grade 5.

patterns as our main results. Students who were assigned to K-5 schools experienced a significant drop in math score in grade 6, but then such negative effect attenuates and either vanish or almost vanish by the end of middle school. For math, the effect is no longer statistically significant in the cases with the tightest bounds. For reading scores we find no negative shocks to achievement in grade 6, but again there is no evidence showing that the students who stayed in the same K-8 school outperforming the counterparts assigned to K-5 schools in grade 8.[9]

Finally, even though our primary results do include students who move to new schools removing one possible benefit of a K-8 policy, it should be noted that if a district or state adopted a K-8 policy, many students would transition to new schools. Therefore, our geo-

[9]A better solution to the problem of mixed transitional effects is to compare students who were in early grade (ideally, who were about to start kindergarten) at the time of school closure and reassignment. Because of the data limitation, we do not have sufficient observations who were about to start kindergarten before the school closure. Moreover, these students were still in elementary school in the last year when the data is available.

graphic IV estimates do have important implications for students that transition to a different school, at least in the short run.

III.4.5 Are the differences in results to the previous literature because different locations?

Our analysis so far has not only estimated an effect for elementary students that have previously not been examined, it has drawn somewhat different conclusions for middle schools students than the three previous that claimed causal estimates (Dhuey, 2013; Rockoff and Lockwood, 2010; Schwerdt and West, 2013). However, one could argue the differences in results may be the results of differences locations examined. To address this concern, we could either employ our geographic approach to the data used for these studies in New York, Florida or British Columbia or employ the IV approach of the three previous studies to our anonymous district. Because the same opportunities through schools closures do not exist with these other studies (not to mention the fact that we do not have access to these data sets), we employ the IV approach to our data sets.[10]

As a reminder, the IV approach introduced by Rockoff and Lockwood (2010) takes into account the possible endogeneity of students first attending a K-5 school and then, after grade 3, switching to a K-8. Such moves may be correlated with unobserved contributions to achievement such as a bad school experience or a residential move. In the IV approach, the authors define treatment as attending a stand-alone middle school and then instrument for middle school entry in grade 6 or 7 using the terminal grade of the school a student attended in grade 3. Specifically, they instrument for entering middle school in grade 6 with an indicator for whether the school that the student attended in grade 3 had a terminal grade of 5. To carry out the grade 3 IV approach, we restricted the sample by dropping students with missing enrollment information at grade 3.[11] [12] Using these data, we estimate the

[10]We focus on the IV approach used by Rockoff and Lockwood (2010) and Schwerdt and West (2013). Dhuey (2013) uses the terminal grade of the school a student attended in grade 4 as the instrument. Dhuey (2013) does not have student achievement in grade 5 and 6. Moreover, the institutional settings are different in Dhuey (2013), which focuses on Canada.

[11]This is in addition to the five restrictions we list out in the data description section.

[12]Those students may have missing information about the ending grade of the school they were enrolled at

same treatment effect as the earlier studies, i.e., the effect of transitioning to a stand-alone middle school as compared to a K-8 school on student achievement, by using the following equation:

$$y_{ig} = a_i + b_g + c_g m_i^6 + e_{ig}, \tag{III.4}$$

where y_{ig} is the achievement of student i in grade g, a_i is the student-specific fixed effect, b_g is the grade-specific fixed effect, m_i^6 is the indicator of whether student i attends a middle school in grade 6 multiplied by a grade-specific parameter c_g, and e_{ig} is the student-grade-specific unobserved error term. We omit the g subscript from the middle school indicator to denote that it does not vary by grade for a given student. By including a grade-specific coefficient we can examine the treatment effect by grade, including grades prior to attending a middle school. As is in the previous studies, we set $c_g = 0$ for $g = 3$, imposing the restriction that the school's grade configuration does not have an impact on student achievement in grade 3. The estimates of c_g, $g > 3$, indicate the achievement difference in grades 4-8 between students who will be enrolled in a middle school and students who will be enrolled in a K-8, relative to their achievement in grade 3.

The first stage of the grade 3 IV approach involves predicting middle school enrollment in grade 6 based on the terminal grade of the school attended in third grade. Like the earlier studies, we find that terminal grade is strongly predictive of attending a middle schoolthe probability of attending a stand-alone middle school in grade 6 is approximately 56 percent greater for the third-grade students who were enrolled in an elementary school with a terminal grade of grade 5.

Table 5 shows the estimates of the effect of middle school enrollment on achievement using this IV method. We display the estimates not only during middle school grades, but pre-middle school grades to examine trends. In examining these pre-middle school trends, much like Rockoff and Lockwood as well at Schwerdt and West, we find that the students who later attend stand-alone middle schools perform at least on par with students in K-8

grade 3, or they may had already finished grade 3 by 1999, the first year in our data set.

schools in math. In reading, these elementary students actually outperform K-8 students in grade 5 by 0.157 standard deviations. In discussing their own similar results,[13] Schwerdt and West (2013) suggest that the positive pre-middle school trends may be a reflection of either better school quality[14] in these schools or a selection into K-5 and K-6 (versus a K-8 school) that is correlated with learning trajectories. In other words, if the second explanation is true, the IV approach may not completely remove the selection bias as a "bias-free" approach would have similar (and insignificant) achievement trajectories prior to attending a middle school.[15]

In examining the middle school grades from our employment of the IV analysis, there is at least some evidence across the estimates that their achievement levels drop both in the transition grade of sixth grade and later in grades 7 and 8. In some cases, these effects are not trivial. For instance, in the transitional sixth grade, student reading achievement drops from an achievement level of 0.157 in fifth grade to -0.010 in sixth grade - a drop of over 0.16 standard deviations. Furthermore, the achievement in grade 8 for students attending a separate middle school lags the achievement of students in K-8 schools by 0.223 and 0.209 standard deviations in math and reading respectively.[16] This suggests that while we

[13]In Schwerdt and West (2013), the authors found positive math achievement trajectory estimates for students in fourth and fifth grade of 0.060 and 0.040 standard deviations, respectively. Similarly, the authors found a positive reading achievement trajectory of 0.058 for fourth graders. Similar prior achievement trajectories are found for students entering a middle school in seventh grade. Rockoff and Lockwood (2010) found for students entering a middle school in sixth grade a positive and statistically significant math and reading estimate in fifth grade of 0.053 and 0.080 standard deviations, respectively. Again, similar results are found for students entering a middle school in seventh grade.

[14]Rockoff and Lockwood (2010) suggested the same possibility.

[15]While Schwerdt and West argue that there is no plausible selection into K-5 or K-6 schools that would explain the drop in performance in the entry middle school grade, we argue that there is at least a plausible argument that families make an endogenous choice at the entry point of Kindergarten, which would lead to unobserved difference in students assigned to K-8 and separate middle schools during the middle school grades. This argument is supported by the number of observable differences of students assigned to K-8 and separate middle schools at grade 3. At the very least, it raises the question of whether there may be an alternative estimation approach that be more effective at dealing with selection not only as a result of students switching once in the school system, but also from students' original decisions about the type of school to attend in Kindergarten.

[16]While the main purpose of the analysis above is to replicate the results of the previous studies, we did also examine whether the results hold true when using information from earlier grades, which the previous studies did not have. More specifically, we the instrumental variable based on earlier enrollment. In our data set we have enrollment information in grade 1 and 2, which allows us to re-estimate the effect using instrumental variables based on grade 1 and 2, respectively. Because of smaller sample sizes, we lose sta-

Table III.5: IV Achievement Estimates of Enrolling in a Stand-Alone Middle School in Grade 6

	Math Score		Reading Score	
Panel A: Estimate of Effects on Test Score Levels				
Grade 4	0.030		0.038	
	(0.034)		(0.035)	
Grade 5	0.024		0.157	***
	(0.037)		(0.037)	
Grade 6	0.032		-0.010	
	(0.039)		(0.039)	
Grade 7	-0.079	*	-0.074	
	(0.046)		(0.046)	
Grade 8	-0.223	***	-0.209	***
	(0.062)		(0.063)	
Panel B: Gains based on the Estimated Coefficient in Panel A				
Grades 4 to 5	-0.006		0.119	***
	(0.029)		(0.032)	
Grades 5 to 6	0.008		-0.166	***
	(0.030)		(0.034)	
Grades 6 to 7	-0.111	***	-0.065	
	(0.035)		(0.041)	
Grades 7 to 8	-0.144	***	-0.134	**
	(0.052)		(0.057)	
No. of students	4769		4769	
Sample Size	24442		24388	

Notes: Standard errors clustered by student are reported in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% confidence level respectively. The specification for Panel A is the IV regression described in equation III.4. Panel B reports differences between estimated coefficients in panel A. Significance levels are based on tests with the null hypothesis that estimated coefficients for consecutive grades are the same.

do not observe exactly the same results (especially in math), there is some evidence that

tistical power for analysis using earlier grades. However, across the different approaches, we gain similar substantive conclusions of the changes in achievement levels for the transition year (grade 5 to 6) as well as long-term achievement levels by grade 8. Therefore, our results from the IV approach are robust to the choice of instrumental variable based on different grades. Nevertheless, using IV based on earlier grades may still not completely solve the problem of endogeneity because the original choice attending a separate elementary/middle school or K-8 school is still self-selected. The results are available upon request.

students experience a dip in reading performance when entering a separate middle school, just as Rockoff and Lockwood (2010) and Schwerdt and West (2013) demonstrated.[17] [18] Therefore, when we employ the same approach as the previous studies, we do come to similar conclusions. It is when we employ a different approach that we come to different conclusions suggesting that the differing conclusions may be the result of the differences in approach, not locations.

However, when comparing the IV approach to our geographic approach, we cannot necessarily claim one is right and one is wrong. They could have different results because of differences in local average treatment effects. The local average treatment effect for the IV estimator is the average effect for all students who attend schools of the same configuration of the school they attended in third grade. The local average treatment effect for the geographic quasi-experimental estimator is the average effect only for students affected by the school closures. Therefore, we cannot definitely conclude whether the differences in our results are because one approach removes a bias that affects the other or because the average effect is different for the two groups. However, by robustness check using the students who stayed in the same K-8 school after school closure, to a large extent we exclude the possibility that our findings are driven by the difference between counterfactuals. Another way to address this concern would be to employ the IV approach to the sample of students in the geographic quasi-experimental approach. However, there is limited overlap between the two samples (which reduces the power to detect effects). The use of only students affected by school closures will limit the ability of the instrument used in the previous IV approach (i.e., terminal grade of school attended in third grade) to predict middle school attendance, so we would not be able to distinguish student selection of school con-

[17] It should be noted that Schwerdt and West (2013) extended the analysis beyond 8th grade to 10th grade to examine whether students who have undergone a middle school transition are more prepared for high school, relative to students who attended K-8. They show that the effects for separate middle schools do not disappear.

[18] Rockoff and Lockwood (2010) find that similar adverse transitional effects in grade 6 for kids enrolled (in grade 3) in a K-5 school, in grade 7 for kids enrolled (in grade 3) in a K-6 school, and in grade 5 for kids enrolled (in grade 3) in a K-4 school.

figuration from school assignments due to the new policy. Therefore, we have to rely on the conceptual argument that our geographic approach is a preferred approach as it deals with endogeneity due both to selection prior to initial enrollment and to selection following third grade and the fact that our approach can actually produce an estimate for elementary grades.

III.5 Conclusion

Grade configuration has been a controversial topic among policymakers and district leaders (Hough, 2005). It was traditionally believed that the separation between elementary and middle schools around grade 6 is a good model for the education of adolescents because middle schools can be a "bridge" between elementary and high schools catering to the special needs of children aged 11 – 13. However, the traditional use of stand-alone middle schools have been challenged fueled by recent research suggesting a move to middle schools can have both short- and long-term adverse effects on students (Dhuey, 2013; Rockoff and Lockwood, 2010; Schwerdt and West, 2013). These papers used an IV approach using the terminal grade of the school attended in third grade as instrument to account for endogenous moves between schools after third grade. However, we argue that these papers do not consider the effects of a K-8 policy on elementary students as the approach cannot produce causal estimates for elementary students as it does not take into account the possibility of students endogenously choosing to attend a K-8 or separate elementary and middle schools at the point of entry into the school system. This not only eliminates their ability to produce causal estimates for elementary students, but may reduce their ability to produce causal estimates for middle school students as well. Our study uses an alternative identification strategy arising from school closures in a midsize district that created new exogenous geographic boundaries for schools and caused some students to be reassigned to K-8 or separate elementary and middle schools. With these new boundaries came an opportunity to create a quasi-experimental approach comparing students who live close to each other but on either side of the boundaries between different configurations of schools. Using this approach, we account not only for the endogenous moves between schools, but endogenous choice of school configuration at entry point which produces causal estimates for both middle and elementary grades.

Our findings from the geographic quasi-experimental design provide some evidence of adverse effects for middle school students in the transition year of grade 6, which is consistent with results from the Rockoff and Lockwood (2010) and Schwerdt and West (2013) papers. However, unlike these earlier papers, we do not find any lingering effects for middle school students in grades beyond the transition grade. Furthermore, we find adverse effects for elementary students enrolled in K-8 schools. The adverse effects for elementary students in K-8 schools combined with the lack of long-term adverse effects for students attending stand-alone middle schools does not provide support for K-8 schools as the previous research suggest. In fact, our results provide some evidence against K-8 schools as a policy. While further research needs to be conducted in a larger set of districts before definitive conclusions can be drawn, we argue that future studies should not only account for students switching to K-8 school after entering a school, but account for the configuration of school in which a student initially enrolls so both the effects for middle and elementary students can be estimated.

Chapter IV

Understanding the Mechanisms Linking Socioemotional Skills and College Education with Longevity

There is a long history of research studying the effect of education on health and longevity (see Grossman, 2006, 2015, and Grossman and Kaestner, 1997, for surveys). A more recent literature documents sizable differences in health and longevity by socioemotional skills. However, we still do not understand the mechanisms behind these effects. To our knowledge, this paper is the first to decompose the effects of socioemotional skills and subsequent college choice on longevity with respect to behavioral and other socioeconomic mediators using a causal framework. Our results support the claim that skills and education causally affect longevity, contribute to a better understanding of the available policy options, shed light on puzzling gender differences in the education-mortality gradient, and allow for predicting the effects of skills and education in different economic environments.

The causal effect of education on longevity is a major question in both theory and policy (Galama and Kippersluis, 2015; Grossman, 1972, 2000). Despite the importance of this question and the long history of research, the evidence is controversial and incomplete. Studies exploiting changes in compulsory schooling laws as exogenous variation in educational attainment report different estimates of the effect: some of them strong, some weak and statistically insignificant (Lochner, 2011). In contrast, few authors have investigated the effect of post-compulsory education on longevity (Buckles et al., 2014; Savelyev, 2015), and additional evidence would strengthen these studies. By employing a unique data set containing measures of multiple plausible mediators, we provide evidence of key mechanisms explaining the effect of education on longevity to buttress causal claims.

Our paper therefore also adds to the economics literature that uncovers the mechanisms behind the education-longevity gradient.[1] Buckles et al. (2014) support their claim that

[1] Psychologists and epidemiologists have performed a number of related mediation analyses, but not in a

college education negatively affects male mortality by showing that education also affects a number of health behaviors, but they stop short of quantifying the contribution of each of these behaviors to the effect on mortality. Cutler, Lleras-Muney, and Vogl (2011) regress 10-year mortality dummies on education and find that the inclusion of health behaviors as controls attenuates the education coefficient by 40%. Balia and Jones (2008) use British data to decompose the Gini coefficient for mortality with respect to determinants as diverse as age, sex, height, ethnicity, social class, education, lifestyles, and health behaviors and find that accounting for lifestyles and health behaviors decreases education's contribution to the total Gini coefficient by 72%.[2]

Unlike previous research, we jointly model and explain the mechanisms behind interrelated effects of socioemotional skills and college education on longevity. In doing so we establish the role of not only aggregates, such as health behaviors, but also specific mediators, such as smoking, earnings, and marriage.

While the education-health gradient has traditionally been the focus of health economists, an emerging literature in the economics of human development builds on a broader concept of human capital that emphasizes socioemotional skills in addition to cognitive skills and education (e.g., Conti et al., 2010; Heckman et al., 2014, 2013b, 2006; Mendolia and Walker, 2014; Savelyev and Tan, 2015). This literature provides evidence for a skill-health gradient in tandem with the education-health gradient and suggests that it is important to account for socioemotional skills as potential confounders of education's effect on health. Our paper integrates a model of longevity with a model of socioemotional skills,

causal framework. These studies typically do their analyses only for one particular cause, such as Conscientiousness (e.g., Lodi-Smith et al., 2010), or one particular health behavior such as smoking (e.g., Turiano et al., 2012). Such approaches make any causal interpretation difficult for a number of reasons, including a lack of control for other correlated skills and mediators. Van Oort et al. (2005) perform a decomposition with multiple mediators for education as a primary cause but do not account for likely confounders, such as IQ and socioemotional skills in early life, family background, and unobserved heterogeneity.

[2]Contoyannis and Jones (2004) and Brunello et al. (2015) explore the mediating role of health behaviors in the effect of education on self-reported general health. We support the claim made in these papers that health behaviors are important mediators linking education with health, but we concentrate on explaining longevity. Although self-reported general health is an informative and widely used summary of health status, longevity has the advantage of being an objective measure.

education choice, and essential life outcomes—an approach that leads to novel results. Our findings reinforce the claim of a causal link between socioemotional skills and longevity.

We use the Wisconsin Longitudinal Survey (WLS), which is based on a sample of about 10,000 high school graduates from the state of Wisconsin, USA. Subjects were first surveyed in 1957 when they graduated from high school and have been followed ever since. The WLS is well suited for studying determinants of longevity due to its long panel and rich information on health behaviors, health status, and mortality. Furthermore, the WLS contains measures of IQ, school achievement, personality, and college education of respondents. Finally, it captures detailed family background variables.

Results of this paper are based on a full information maximum likelihood estimation of a system of equations, which model educational attainment, health behaviors, lifestyles, income, work conditions, health stock, and mortality. We take advantage of natural exclusion restrictions leading to recursive equations: earlier life outcomes may affect later life outcomes but not the reverse.

We use academic achievement in high school conditional on IQ, to account for pre-college socioemotional skills. IQ is primarily a measure of fluid intelligence, which represents the ability to quickly solve new puzzles that do not require specific acquired knowledge.[3] In contrast to IQ scores, school achievement represents the ability to pass tests by using acquired skills and knowledge, called crystallized intelligence (see Heckman and Kautz (2014) for a survey). As a result, school achievement depends not only on IQ but also on socioemotional skills that affect the learning process such as perseverance, ability to delay gratification, locus of control, emotional stability, and motivation. Borghans et al. (2011) show a major contribution of socioemotional skills to school achievement on top of the fluid intelligence. In line with this reasoning, we view school achievement conditional on IQ as a one-dimensional measure of socioemotional skills, aggregated with weights representing the relative roles of those skills for academic success.

[3]Clearly, basic knowledge, such as reading or arithmetic, is still essential for completing the IQ test, but such knowledge should be a common background of students who are graduating from high school.

We acknowledge that socioemotional skills are complex and multidimensional. While it would be ideal to have access to direct measures of socioemotional skills in early life, the WLS collected no such measures. However, our method allows us to estimate the effect of a highly predictive aggregate of the socioemotional skills that are relevant to educational success and, as we show, to health. The advantage of this approach, when applied to the WLS data, is that socioemotional skills are measured prior to college education and observations of health-related outcomes, adding to the plausibility of our exclusion restrictions and a causal interpretation of the estimated effects. Moreover, we estimate conditional associations between early socioemotional skills, as we define them, and the Big Five personality factors at age 53 to gain a better understanding of what our one-dimensional measure of socioemotional skills captures: it is positively related to Conscientiousness and Agreeableness but is negatively related to Neuroticism. This relationship is consistent with our expectation that students with high socioemotional skills should be organized, thoughtful about the future, compliant with norms and rules, cooperative, and emotionally stable.

Controlling for socioemotional skills should capture usual confounders, such as perseverance and the ability to delay gratification. Additionally, we control for IQ and detailed background variables. We combine continuous latent factors that represent skills and health stock with discrete latent factors that account for unobserved heterogeneity. As in Heckman et al. (2006), we identify the effects of our continuous latent factors since they are of substantive interest, and control for them, as they are potential confounders of effects of education and other mediators of longevity. As in Mroz (1999), through finite mixtures we flexibly account for unobserved heterogeneity.[4] Sources of unobserved heterogeneity may include omitted controls such as health and genetic endowments as well as possible measurement error in variables. We allow for non-linear effects of unobserved heterogeneity on outcomes.

One advantage of our model is that it allows us to adjust the effects of skills and ed-

[4]The finite mixtures model is also called latent class model, or semiparametric heterogeneity model, or a model with discrete factor random effects (e.g., Cameron and Trivedi, 2005).

ucation on mortality for different environments: hypothetical or actual (for later cohorts). Adjusting to hypothetical environments is beyond the reach of a traditional treatment effect approach when a natural (or randomized) experiment is used to calculate causal effects without explicitly modeling the mechanisms that underlie them (Heckman and Vytlacil, 2007). The only way to adjust such traditional estimates to later cohorts is to re-estimate them in the future when new data become available, by which time policy implications regarding the cohort of interest may become less relevant.[5] Also, such re-estimation requires new data and new research efforts, assuming a natural experiment is available to identify effects for future cohorts.[6]

Based on our findings, we reach three main conclusions. First, we show that a large part of the estimated effect of socioemotional skills and college education on longevity can be explained by plausible mediators, giving us greater confidence about the causal relationship between skills, education, and longevity. Second, we show that the positive effect of skills and education on longevity through earnings and other mediators is partially masked by the negative effect of education on marriage for females of this cohort. We adjust the effect of education on marriage to reflect more recent cohorts and obtain an effect of education on longevity that is 14–23% stronger. Finally, our estimates show that skills and education act through multiple channels. Targeting socioemotional skills and education may be an advantageous policy that will simultaneously improve a range of health behaviors–resulting in increased health and longevity.

[5]Adjusting to actual environment is especially relevant in longevity research due to a long lag between essential health-related decisions and mortality. By the time we see substantial mortality of a cohort and are able to estimate the effects using traditional methods, it is too late to develop a policy that prevents mortality for this cohort.

[6]Compulsory schooling, which is by far the most popular instrument for identifying the effect of education on health and longevity, is unlikely to show much further variation in the developed world. Another popular natural experiment is a military draft, which no longer affects the civilian population in the US because of the transition to a volunteer army.

IV.1 Data

We use the Wisconsin Longitudinal Study (WLS) (Hauser and Sewell, 2005), which follows about $10,000$ individuals from Wisconsin, who were first surveyed at high school graduation in 1957 with follow-ups conducted in 1975, 1992, 2004, and 2011. The study represents white, non-Hispanic high school graduates.[7] Subjects were born between 1937 and 1940, with 78% of them born in 1939 and 16% born in 1938.

The WLS is well suited for the study of the developmental origins of health and longevity because it is a long panel with a relatively early start and a unique combination of measurements. At high school graduation in 1957, we observe measures of IQ and high school achievement. We also observe important background variables, including number of siblings, order of birth, degree of urbanization, and parental education, occupation, and income. We include parental longevity observed in the period from 1957 to 2011 in the set of background controls to better account for possible genetic influences on life outcomes and longevity of the subject (see Table IV.1 for details about background variables). Later surveys contain information about post-compulsory education. Starting in 1992, the WLS collected detailed information about health behaviors, such as smoking tobacco and exercising, as well as Big Five personality measures.

Furthermore, the WLS tracks the death of respondents through multiple sources to maximize accuracy. First, death status is updated when the WLS tries to contact the respondent. Second, death status is updated periodically with the Social Security Administration's Death Index. Third, information from the National Death Index is also used.

Figure IV.1 documents the structure of observations starting in 1992, at which time the first set of potential post-college mortality mediators $\boldsymbol{B}_1$ is measured. Constrained by the data on mediators, we perform our longevity mediation study conditional on survival

[7] There are only 30 minority respondents in the sample, whom we exclude from our statistical analysis because this subsample is too small to reliably study the minority population, for whom effects and mechanisms may differ. In 1940, which is the birth year of the youngest individuals in the WLS cohort, the share of whites in Wisconsin was 99.2%.

Table IV.1: Description of Background Variables

Variable	Year of measurement	Males mean	Males s.d.	Females mean	Females s.d.
IQ[a]	1957	101.5	15.2	101.0	14.4
Other background variables, ***X***					
Father is a farmer or a farm manager	1957	0.200	0.400	0.198	0.398
Father is a white collar employee	1957	0.301	0.459	0.299	0.458
Father has attended college[b]	1957	0.160	0.367	0.145	0.352
Mother has attended college[b]	1957	0.144	0.351	0.150	0.357
Parental income (log)[c]	1957	8.536	0.695	8.532	0.666
Attended high school in a rural area	1957	0.187	0.390	0.187	0.390
Resided in a metropolitan area[d]	1957	0.337	0.473	0.366	0.482
Respondent's number of siblings	1975	3.170	2.522	3.284	2.561
First-born or the only child	1975	0.407	0.491	0.375	0.484
Second-born	1975	0.261	0.439	0.274	0.446
Third-born	1975	0.144	0.351	0.142	0.349
Fourth-born or above	1975	0.188	0.391	0.208	0.406
Respondent has abnormal weight[e]	1957	0.284	0.451	0.313	0.464
Father's age at death	1957–2011	75.07	14.34	74.93	14.29
Mother's age at death	1957–2011	81.88	14.62	81.97	14.54
Respondent's childhood household had a smoker[f]	1957	0.757	0.429	0.734	0.442
Ever lived with a problem drinker when growing up	1992	0.184	0.387	0.220	0.415
Birth year 1937–38	1957	0.216	0.411	0.144	0.351
Birth year 1939	1957	0.746	0.435	0.804	0.397
Birth year 1940	1957	0.038	0.192	0.053	0.223
Sample size[g]		3961		4491	

Notes: [a]Henmon-Nelson test score. [b]At least some college coursework or above. [c]Calculated as log(1+parental income). [d]Includes Madison, and Milwaukee, as well as Brown, Kenosha, Racine, and Douglas counties. [e]BMI in 1957 is below 10th percentile or above 80th percentile. [f]Up until 16 years old. [g]Estimation sample size. Calculations are performed for the WLS respondents conditional on survival to January, 1993.

Figure IV.1: Mediators and Mortality: Timing of Observations

Mediators[(a)]	B_1												B_2							B_3		
Year[(b)]	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13
Age, 1940 cohort	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
Age, 1939 cohort	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74
Age, 1938 cohort	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
Age, 1937 cohort	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76

Notes: [(a)]Mediators, including health behaviors, lifestyles, earnings, and job conditions, measured in three follow-ups in 1992, 2004, and 2011. [(b)]Calendar years corresponding to mortality observations starting from the 1992 follow-up.

to January 1993. We exclude individuals who either have missing education information or attrited before January 1993. Altogether we exclude 12% of the initial sample, among whom 3.7% died before 1993. However, the difference between our estimation sample and the initial sample is either weak or nonexistent: we find no statistically significant difference between the two samples with respect to the key variables available for both samples: school achievement, IQ, and background variables.[8] The most recent death status data are available for 2013, at which time surviving respondents reach up to 76 years of age. About 20% of men and 15% of women died over the period 1993–2013.

As Figure IV.1 shows, we observe a vector of potential mediators $\boldsymbol{B}_j$ three times, $j = 1, 2, 3$. At the same points of time, we observe measures $\boldsymbol{H}_j$ of health stock. As we show later, a model that accounts only for 1992 observations, $\boldsymbol{B}_1$ and Θ_1^H, is about as informative as a more complex model that also accounts for $\boldsymbol{B}_2$ and Θ_2^H. (Information on $\boldsymbol{B}_3$ and Θ_3^H is of limited practical use at this point given a short period of mortality observations after 2011.) We also account for potential early behavioral confounders $\boldsymbol{B}_0$: smoking tobacco and marriage before the median college graduation age. By doing so we address the concern that early smoking may confound the effect of schooling on smoking (Farrell and Fuchs, 1982) and that early marriage may increase the probability of dropping out of

[8]See Table A.14 of the Web Appendix.

college, especially for women (Goldin, 1997).[9] Table IV.2 documents descriptive statistics of potential mediators, health measures, early behaviors, education, and mortality.

We expect our study of high school graduates to be salient for a substantial share of the U.S. white population. Hauser and Willis (2005) document that roughly 75% of students in Wisconsin graduated from high school during the late 50's, while the corresponding national statistic for whites is even higher (83%) according to Fischer and Hout (2008). Furthermore, high school graduation rates for later cohorts are even higher. For example, in the 1980 cohort, the graduation rate of whites is 91% in Wisconsin and 88% at the national level.[10]

The WLS contains standard measures of the Big Five taxonomy of personality which we use to understand our early life construct of socioemotional skills. The Big Five is an established contemporary categorization of personality consisting of: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. In short, Openness is a propensity to be intellectual, creative, and open to new ideas and actions; Conscientiousness is about following rules and being organized and thoughtful about the future; Extraversion implies an energetic and outgoing character; Agreeableness distinguishes people who are trusting and cooperative; and Neuroticism refers to emotional instability. The Big Five taxonomy is widely used and is considered to be a comprehensive representation of human personality across languages, geographic regions, and cultures (Goldberg et al., 2006; John and Srivastava, 1999). There is growing evidence that personality is malleable in the period from childhood to young adulthood (see Heckman and Kautz (2014) for a survey). Table IV.3 describes established measures of the Big Five along with academic achievement in high school.

[9] We do not have access to other pre-college health behaviors and lifestyles, but those two are especially important, given the key role that smoking and marriage play in mediating longevity.

[10] 1980 cohort statistics are based on the American Community Survey for respondents aged 25–34 in 2010 (http://www.higheredinfo.org).

Table IV.2: Education and Health-Related Outcomes

Variable	Year of measurement	Males mean	Males s.d.	Females mean	Females s.d.
Education, *D*					
Bachelor's degree or above	1975–1992	0.306	0.461	0.178	0.382
Potential mediators, $\boldsymbol{B}_1$					
Smoking tobacco[a]	1992	0.175	0.380	0.178	0.383
Risky drinking of alcohol[b]	1992	0.082	0.274	0.061	0.239
Physical exercise[c]	1992	0.830	0.376	0.788	0.409
Overweight[d]	1992	0.201	0.401	0.166	0.372
Marriage[e]	1992	0.851	0.356	0.798	0.402
Social activity[f]	1992	0.887	0.316	0.798	0.401
Household per capita income (log)[g]	1992	3.240	0.726	3.061	0.832
Dangerous working conditions[h]	1992	0.227	0.419	0.117	0.321
Measures of physical health, $\boldsymbol{H}_1$					
Major illness[i]	1992	0.552	0.497	0.615	0.487
Stayed in bed at least once last year[j]	1992	0.297	0.457	0.399	0.490
Hospitalization at least once last year[k]	1992	0.078	0.269	0.084	0.278
General health[l]	1992	4.137	0.666	4.163	0.679
Early behaviors, $\boldsymbol{B}_0$					
Smoking tobacco[m]	1992	0.225	0.418	0.183	0.386
Marriage[n]	1992	0.547	0.498	0.679	0.467
Mortality					
Died	1992–2013	0.189	0.392	0.145	0.352
Age of death	1992–2013	65.98	5.501	66.48	5.701
Sample size[o]		3961		4491	

Notes: [a]Currently a tobacco smoker. [b]Either (1) consumed more than 4 alcoholic drinks on average occasion last month if male or more than 3 if female, or (2) exceeded 14 drinks per average week (in the last month) if male or 7 per week if female (above the http://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/moderate-binge-drinking threshold for low-risk drinking according the National Institute of Alcohol Abuse and Alcoholism). [c]Respondent does light or heavy exercises at least once per week. [d]Respondent's BMI is above 25. [e]Respondent is currently married. [f]Current participation in at least one social organization, such as a club or a church. [g]Calculated as log(1+household income/current household size). [h]Working conditions are classified by the WLS as "extremely dangerous." [i]Major illness includes cancer, diabetes, heart disease, anemia, asthma, arthritis or rheumatism, bronchitis or emphysema, chronic liver trouble, serious back trouble, high blood pressure, circulation problems, kidney or bladder problems, ulcer, allergies, multiple sclerosis, colitis, and other major conditions mentioned by a medical professional. [j]Stayed in bed at least once last year for more than half of the day because of illness or injury. [k]Has been hospitalized at least once last year for at least one night. [l]Self-reported health ranges from 1 (very poor) to 5 (excellent). [m]The respondent first started smoking before the median age of college graduation for whites in 1950s–1960s: https://www.census.gov/hhes/socdemo/education/data/cps/1960/p23-09.pdf 23.4 for men and https://www.census.gov/hhes/socdemo/education/data/cps/1960/p23-09.pdf 22.2 for women. [n]The first marriage occured before the median age of college graduation (defined as above). [o]Estimation sample size. Calculations are performed for the WLS respondents conditional on survival to January, 1993.

Table IV.3: Measures of School Achievement and Big Five Personality

Variable	Year of measurement	Males mean	Males s.d.	Females mean	Females s.d.
Academic achievement at high school					
Standardized academic achievement[(a)]	1957	97.9	14.5	104.5	14.3
Member of an honor society[(b)]	1957	0.044	0.204	0.063	0.243
Outstanding student[(c)]	1957	0.100	0.301	0.118	0.323
Big Five[(d)]					
Conscientiousness					
Thorough	1992	5.466	0.774	5.576	0.705
Reliable	1992	5.794	0.596	5.863	0.511
Disorganized	1992	4.545	1.416	4.408	1.548
Lazy	1992	4.134	1.438	3.974	1.472
Efficient	1992	5.164	0.862	5.301	0.832
Easily Distracted	1992	4.095	1.368	4.064	1.417
Openness					
Artisitic	1992	4.256	1.291	4.767	1.185
Imaginative	1992	4.836	1.069	4.595	1.248
Sophisticated	1992	2.944	1.494	3.569	1.506
Extraversion					
Talkative	1992	4.130	1.392	4.472	1.305
Reserved	1992	2.765	1.314	2.798	1.373
Quiet	1992	3.207	1.440	3.277	1.474
Shy	1992	3.587	1.424	3.443	1.462
Enthusiastic	1992	4.509	1.045	4.702	1.059
Agreeableness					
Fault-Finding	1992	3.847	1.397	3.982	1.396
Rude	1992	4.281	1.392	4.747	1.340
Aloof	1992	4.075	1.407	4.442	1.446
Considerate	1992	5.029	0.965	5.379	0.794
Cooperative	1992	5.089	0.830	5.411	0.703
Neuroticism					
Tense	1992	4.178	1.303	4.366	1.242
Stable	1992	2.236	1.168	2.220	1.189
Worries	1992	3.547	1.508	3.933	1.553
Calm	1992	2.378	1.133	2.536	1.262
Nervous	1992	3.081	1.442	3.546	1.528
Sample[(e)]		3961		4491	

Notes: [(a)]Standardized high school grades percentile rank. [(d)]Member of at least one honor society. [(c)]Teacher's evaluation of the high school graduate as "outstanding." [(d)]Each Big Five measure ranges from 1 to 6. [(e)]Estimation sample size. Calculations are based on the WLS data.

IV.2 Methodology

Below we model recursive relationships between earlier and later life outcomes. Latent skills Θ are proxied by measures of achievement at high school. Since we condition on IQ, we interpret the effect of Θ as the effect of socioemotional skills based on the logic presented in the introduction. Skills Θ affect health behaviors of young adults B_{0q}, $q = 1,...,Q$, before the age of college graduation. To model binary variables B_{0q}, we use the logit model, in which latent indices B^*_{0q} determine the outcomes in a standard way: $B_{0q} = 1$ if $B^*_{0q} \geq 0$, and $B_{0q} = 0$ otherwise.

Next, early behaviors and skills affect latent index D^*, which determines the college education outcome D, defined as bachelor's degree or above. In turn, education, early behaviors, and skills affect index B^*_{1k}, which determines the choice of potential mediators of longevity at midlife, B_{1k}, where $k = 1,...,K$ denotes the type of the mediator. All mediators are either binary or continuous. For continuous mediators, $B^*_{0k} = B_{0k}$, and for binary mediators we use the logit model.

Another mediator is the latent general health stock at midlife, Θ^H_1, which we relate to its determinants in the same way as we do for B^*_{1k}. Finally, potential mediators B_{1k} and Θ^H_1 affect $\lambda(t)$, which is the conditional instantaneous probability of a person's death at time t (event $T = t$) given that a person has survived to that age (event $T > t$), where time t is continuous.

We estimate a parsimonious system of recursive equations (IV.1–IV.6) that models relationships among variables over the life cycle as described above. The parsimonious model is a test-based simplification of a more general model documented in Web Appendix E. The general model accounts for (a) interactions between socioemotional skills and IQ, (b) dependence of regression coefficients on education level, and (c) time-dependence of (c.1) coefficients and (c.2) regressors in the mortality model. In the parsimonious model presented below, we dispense with features (b) and (c.2), but maintain (a) for equations (IV.2) and (IV.3) and (c.1) for a subset of coefficients. The main motivation for the parsimonious model is practical: the general model is very complex and requires a long time for estimation and specification tests. Additionally, elimination of redundant degrees of freedom is known to improve the efficiency of estimators.

The parsimonious recursive system estimated conditional on IQ and $\boldsymbol{X}$ is the following:[11]

$$\Theta = \tau_\Theta + \varepsilon_\Theta \tag{IV.1}$$

$$B^*_{0q} = a_{1q}\Theta + b_{2q}\Theta \cdot IQ + \tau_{B0q} + \varepsilon_{B0q},\ q = 1,...,Q \tag{IV.2}$$

$$D^* = b_1\Theta + b_2\Theta \cdot IQ + \sum_q b_{3q}B_{0q} + \tau_D + \varepsilon_D \tag{IV.3}$$

$$B^*_{1k} = c_{1k}\Theta + \sum_q c_{2kq}B_{0q} + c_{3k}D + \tau_{B1k} + \varepsilon_{B1k},\ k = 1,...,K \tag{IV.4}$$

$$\Theta^H_1 = d_1\Theta + \sum_q d_{2q}B_{0q} + d_3D + \tau_{H1} + \varepsilon_{H1} \tag{IV.5}$$

$$\ln(\lambda(t)) = e_{1j}\Theta + \sum_q e_{2q}B_{0q} + e_{3j}D + \sum_k e_{4k}B_{1k} + e_5\Theta^H_1 + \tau_{\lambda j} + \ln(\lambda_0(t)),\ j = 1,...,J, \tag{IV.6}$$

where ε_Θ, ε_{B0q}, ε_D, ε_{B1k}, and ε_{H1} are uncorrelated error terms with zero means. Intercorrelated terms τ_Θ, τ_{B0q}, τ_D, τ_{B1k}, τ_{H1}, and $\tau_{\lambda j}$ account for unobserved heterogeneity that is modeled using finite mixtures (Aitken and Rubin, 1985; Heckman and Singer, 1984; Mroz, 1999). Heckman and Singer (1984) have shown that finite mixtures with a small number of points of support (2–5) are sufficiently flexible for the purpose of approximating unobserved heterogeneity in duration models. This method leads to accurate estimates of structural parameters and accurate predictions of durations despite less accurate estimates of the mixing distribution. We use a model with four points of support, a setting that minimizes both Akaike and Bayesian Information Criteria. Also, due to empirical considerations described in sections IV.1 and IV.3, we set $Q = 2$, $K = 8$, and $J = 2$.

[11] We omit $\boldsymbol{X}$ and IQ from equations to simplify notation. IQ becomes rank-stable by about age 10, and so we view IQ as one of the determinants of achievement by the end of high school together with other family background variables $\boldsymbol{X}$.

Equation (IV.6) represents a generalization of the continuous-time mixed proportional hazard (MPH) model, in which coefficients are allowed to change in discrete time j, while baseline hazard λ_0 is a function of continuous time t.[12] Discrete time points naturally correspond to ages of survey follow-ups. The hazard function depends on potential mediators $\boldsymbol{B}_1$ measured at the beginning of the risk period ($t = 0$). We also allow for direct effects of early behaviors, $\boldsymbol{B}_0$, education, D, and skills, Θ, which may be determinants of the unobserved mediators, just as they are for the observed ones.

Identification of equations (IV.1–IV.6) comes from a natural exclusion restriction, which makes the model recursive: earlier outcomes affect later outcomes, but not vice versa. We assume that error terms ϵ are mutually uncorrelated, conditional on a rich set of background variables, latent socioemotional skills, and unobserved heterogeneity. Terms $\boldsymbol{\tau}$ correlate with skills and outcomes but not with background variables.[13] Unobserved heterogeneity terms $\boldsymbol{\tau}$ are correlated across equations. Such modeling of common unobserved heterogeneity justifies a joint estimation of system (IV.1–IV.6) as opposed to equation-by-equation estimation. The use of common latent factors Θ and Θ_1^H is another reason for joint estimation. Identification of nonlinear recursive models involving endogenous binary variables is shown in Maddala (1983), *pp*.120–123. Mroz (1999) shows how using finite mixtures leads to accurate estimates of structural parameters in a recursive model with a binary endogenous variable.

Equations (IV.4)–(IV.5) are in reduced form since B_{1k}, $k = 1, ..., K$, and Θ_1^H are measured simultaneously and we cannot identify effects of these variables on each other. The same consideration is true for equations (IV.2). However, this limitation does not hurt our ability to make decompositions of the effect of Θ and D on λ, which is the main aim of this paper.

In order to identify system (IV.1–IV.6), we link the latent skills and the health stock to

[12] See Asparouhov et al. (2006) for a detailed discussion of this generalization.

[13] Orthogonality of τ to background variables is a standard assumption (e.g., Van Den Berg, 2001; Mroz, 1999).

their observed measures $\boldsymbol{M}_l$ and $\boldsymbol{H}_{1m}$ while accounting for measurement error. For this purpose, we estimate the following factor model (or "measurement system") conditional on $\boldsymbol{X}$ simultaneously with model (IV.1–IV.6):

$$M_l{}^* = \alpha_{1l}\Theta + \alpha_{2l} + \eta_{Ml},\ l \in 1,...,L \tag{IV.7}$$

$$H_{1m}{}^* = \beta_{1m}\Theta_1{}^H + \beta_{2m} + \eta_{H1m},\ m \in 1,...,M, \tag{IV.8}$$

where η_{Ml} and η_{H1m} are zero-mean error terms, α_{1l} and β_{1m} are factor loadings, and α_{2l} and β_{2m} are intercepts. Variables M_l^* determine measures of performance in high school: achievement percentile in high school, M_1; teacher's nomination as an outstanding student, M_2; and membership in at least one honor society, M_3, so that $L = 3$.[14] Since M_1 is continuous, $M_1^* = M_1$, while for binary M_2 and M_3 we use the logit model.[15] Similarly, variables H_{1m}^* correspond to $M = 4$ continuous and binary measures of general health documented in Table IV.2.

We use a standard assumption that error terms are mutually uncorrelated and also uncorrelated with latent factors, but latent factors may correlate among themselves. We make further standard normalizations to avoid indeterminacy: $(\tau_{H1}|C=1) = (\tau_{\lambda j}|C=1) = 0$ for $j = 1,2$, where latent class $C = 1$ is the class with the largest probability, and $\mathrm{Var}(\varepsilon_\Theta|C = c) = \mathrm{Var}(\varepsilon_{H1}|C = c) = 1$ for $c = 1,2,3,4$. Finally, coefficients α_{11} and β_{11} are set to be positive. These two coefficients link latent socioemotional skills to standardized academic achievement in high school and link latent health stock to self-reported general health at age 53. Assigning the coefficients positive values makes the latent factors interpretable as skills and health, rather than negative skills and sickness, without changing any decompositions or substantive conclusions. Our measurement system satisfies the sufficient identification

[14] One can argue that achievement may depend on the type of school: for the same level of IQ and socioemotional skills it may be harder to be on the same achievement percentile in a more prestigious school. However, controlling for predictions of the school type should minimize the issue. We control for parental earnings, education, and occupation; rural area, town, and metropolitan area, and others.

[15] See Table IV.3 for details of these three measures of achievement.

condition since it has at least three dedicated measures per latent factor.[16]

For estimation and inference, we use the MCMC multiple imputation of missing data (Rubin, 1987; Schafer, 1997) and full information maximum likelihood estimation. We use multiple random starting values to ensure that we reach the global maximum of the likelihood function.

IV.2.1 Decomposition of the Effect of Education on the Hazard of Death

Performing a decomposition of the effect of binary variable D on the hazard of death with respect to its mediators is complicated by our inability to take derivatives with respect to D. To circumvent this problem, we use an approximation based on the Taylor expansion. For the case in which education changes from some level d to another level $d+\Delta$, we can write:[17]

$$\frac{\Delta\lambda}{\lambda} = \frac{(\lambda|D=d+\Delta)-(\lambda|D=d)}{(\lambda|D=d)} \approx ln(\lambda|D=d+\Delta)-ln(\lambda|D=d) \equiv \psi^d. \qquad \text{(IV.9)}$$

Then, using formulas (IV.4–IV.6) and (IV.9), we obtain the following decomposition for $d=0$:

$$\psi^d = \underbrace{(\sum_k e_{4k}\tilde{c}_{3k} + e_5 d_3)\Delta}_{\substack{\text{effect through}\\ \text{observed mediators}}} + \underbrace{e_3\Delta}_{\substack{\text{effect through}\\ \text{unobserved mediators}}}, \qquad \text{(IV.10)}$$

where $\tilde{c}_{3k}$ is the marginal effect of education on mediator B_k, and e_3 is a weighted average of e_{30} and e_{31}.[18] Equation (IV.10) is most accurate for a small change in the share of educated

[16] For details about identification of standard factor models, see Anderson and Rubin (1956). A case with binary measures is analyzed in Muthen (1983).

[17] Decomposition analysis is conditional on $\boldsymbol{X}$, Θ, IQ, and τ.

[18] Here and above, $\tilde{c}_{3k} = c_{3k}$ for continuous mediators. For binary mediators, $\tilde{c}_{3k}$ is the weighted average of logit marginal effects for each latent class with probabilities of classes used as weights. To facilitate presentation of the main results, we use e_3, which is the weighted average of e_{30} and e_{31}, with weights proportional to corresponding time intervals. When calculating life expectancy and the value of remaining life in Appendix D, we fully account for the estimated time-dependence of the MPH model coefficients.

people in the population Δ. Case $\Delta = 1$ approximates a counterfactual for a person who is induced to get a college degree. We present results for $\Delta = 1$, but the reader can easily multiply these results by any specific Δ of interest.

IV.2.2 Decomposition of the Effect of Socioemotional Skills on the Hazard of Death

Consider a partial derivative, ψ, of the logarithm of the hazard of death (IV.6) with respect to skill θ. The interpretation of ψ is the relative change in the hazard of death in response to a one standard deviation increase in socioemotional skills. By applying the chain rule to the system (IV.2–IV.6) while evaluating marginal effects at the average levels of background variables and skills, we obtain that $\psi = \psi_A + \psi_B$, where ψ_A is the effect explained through mid-life mediators B_1 and Θ^H, and ψ_B is the residual effect. Formulas for ψ_A and ψ_B are the following:

$$\begin{aligned}\psi_A = & \underbrace{\sum_k e_{4k}\{\tilde{c}_{1k} + \sum_q \tilde{c}_{2kq}\tilde{a}_{1q} + \tilde{c}_{3k}(\tilde{b}_1 + \sum_q \tilde{b}_{3q}\tilde{a}_{1q})\}}_{\text{effect through mid-life mediators}} \\ & \underbrace{+ e_5(d_1 + \sum_q d_{2q}\tilde{a}_{1q} + d_3(\tilde{b}_1 + \sum_q \tilde{b}_{3q}\tilde{a}_{1q}))}_{\text{effect through mid-life health stock}}, \end{aligned} \tag{IV.11}$$

$$\psi_B = \underbrace{e_1}_{\substack{\text{unexplained}\\ \text{effect}}} + \underbrace{e_3(\tilde{b}_1 + \sum_q \tilde{b}_{3q}\tilde{a}_{1q})}_{\substack{\text{effect through education,}\\ \text{but not through observable}\\ \text{mid-life behaviors}}} + \underbrace{\sum_q e_{2q}\tilde{a}_{1q}}_{\substack{\text{effect through early behaviors,}\\ \text{but not through observable}\\ \text{mid-life behaviors}}}, \tag{IV.12}$$

where coefficients with tildes denote the marginal effects implied by the corresponding logit models.[19]

By regrouping these terms, we obtain the following decomposition that allows us to

[19] e_1 is a weighted average of coefficients e_{10} and e_{11}. Similarly, e_3 is a weighted average of e_{30} and e_{31}. Weights are proportional to corresponding durations.

investigate the role of education as a mediator: $\psi = \psi_C + \psi_D$, where ψ_C is the effect mediated by college education, and ψ_D is the effect mediated by all other possible channels. We calculate ψ_D as $\psi_A + \psi_B - \psi_C$, where

$$\psi_C = \underbrace{(\tilde{b}_1 + \sum_q \tilde{b}_{3q}\tilde{a}_{1q})}_{\substack{\text{effect of socioemotional skills} \\ \text{on education}}} \underbrace{(\sum_k e_{4k}\tilde{c}_{3k} + e_5 d_3 + e_3)}_{\substack{\text{effect of education} \\ \text{on longevity}}}. \tag{IV.13}$$

To summarize, we propose a number of informative decompositions. We identify the mediating role of each health-related outcome, as well as the unexplained portion of the effect through unobserved channels. In addition, we can separate the effects of skills into those that operate through education and those that operate through other channels.

IV.3 Empirical Results

Below we first discuss the empirical evidence supporting our interpretation of socioemotional skills. Next, we analyze two components of mediation analysis: (a) the effects of causes (education and skills) on potential mediators and (b) the effects of potential mediators on the outcome of interest (longevity). We then combine the two components to obtain a decomposition of the effects of skills on longevity with respect to mediators.

IV.3.1 Linking Socioemotional Skills at High School with Personality in Late Adulthood

We interpret school achievement conditional on IQ as a one-dimensional measure of socioemotional skills, which we expect to capture important dimensions of character that are needed to succeed in high school. These dimensions include perseverance, low discount rate, locus of control, propensity to be organized and cooperative, and emotional stability.

Our results demonstrate that the data are consistent with our interpretation. We find that, conditional on background controls, socioemotional skills at high school graduation as we define them positively correlate with Conscientiousness and Agreeableness and negatively correlate with Neuroticism for both genders at age 53 (see Table IV.4).

Table IV.4: Conditional Associations Between Socioemotional Skills at High School and the Big Five Personality Factors at Age 53

	Conscien-tiousness	Extraversion	Openness	Agreeableness	Neuroticism
Males					
Socioemotional skills	0.125 ***	-0.077 ***	-0.015	0.120 ***	-0.064 **
	(0.030)	*(0.028)*	*(0.032)*	*(0.030)*	*(0.029)*
Females					
Socioemotional skills	0.156 ***	0.014	-0.051	0.109 ***	-0.053 *
	(0.030)	*(0.028)*	*(0.031)*	*(0.030)*	*(0.028)*

Notes: Each latent Big Five factor is regressed on IQ, other background controls, and Θ. Coefficients correspond to a standardized latent factor Θ. Huber-White standard errors are reported in parentheses. Asterisks denote the level of statistical significance: ***, **, and * represent $p < 0.01$, 0.05, and 0.10. Results are based on the WLS data.

For men, socioemotional skills also negatively correlate with Extraversion, which is not surprising given the well-documented fact that school achievement and extraversion are negatively correlated: intense socialization is not conducive to academic success (de Raad and Schouwenburg, 1996; Goff and Ackerman, 1992).

IV.3.2 Determinants of Education and Potential Mediators in Mid-Life

The potential mediators at the start of the risk period (age 53) are health behaviors and other health-related socioeconomic outcomes that we select based on prior evidence from the literature: (1) health-related behaviors, such as smoking tobacco or engaging in physical exercise (Cawley and Ruhm, 2012); (2) lifestyles, such as marriage or intensity of social life (Holt-Lunstad et al., 2010); (3) income (Snyder and Evans, 2006); and (4) dangerous working conditions (Viscusi, 2013). In total, we model $K = 8$ such mediators. An additional mediator is health stock at the start of the risk period. Health stock accounts for influences of mediators of longevity in earlier life.

Table IV.5: Marginal Effects of Skills and Education on Potential Mediators at Midlife

	Bachelor's degree or above (1)	Smoking tobacco (2)	Risky drinking of alcohol (3)	Physical exercise (4)	Overweight (5)	Marriage (6)	Social activity (7)	Household per capita income (log) (8)	Dangerous working conditions (9)	Health stock (10)
Males										
Bachelor's degree	-	-0.038 **	-0.045 ***	0.037	-0.052	-0.006	0.062 ***	0.269 ***	-0.131 ***	0.411 ***
or above	-	*(0.017)*	*(0.014)*	*(0.028)*	*(0.093)*	*(0.017)*	*(0.017)*	*(0.035)*	*(0.021)*	*(0.156)*
Socioemotional skills	0.255 ***	-0.037 ***	-0.019 ***	0.029 **	-0.011	0.000	0.025 **	0.074 ***	-0.026 **	0.041
	(0.025)	*(0.011)*	*(0.007)*	*(0.014)*	*(0.046)*	*(0.011)*	*(0.010)*	*(0.022)*	*(0.012)*	*(0.046)*
Females										
Bachelor's degree	-	-0.027	-0.007	0.066 **	-0.025	-0.067 ***	0.210 ***	0.167 ***	-0.008	0.156
or above	-	*(0.019)*	*(0.007)*	*(0.034)*	*(0.032)*	*(0.020)*	*(0.027)*	*(0.049)*	*(0.017)*	*(0.186)*
Socioemotional skills	0.098 ***	-0.055 ***	-0.009 **	0.046 *	-0.020	0.033 ***	0.039 ***	0.104 ***	-0.016	0.138
	(0.010)	*(0.013)*	*(0.004)*	*(0.025)*	*(0.025)*	*(0.013)*	*(0.013)*	*(0.030)*	*(0.010)*	*(0.089)*

Notes: Column (1) shows estimates of $\tilde{b}_1$, a marginal effect corresponding to b_1 from equation (IV.3). Columns (2–9) show the effect of education ($\tilde{c}_{3k}$) and the effect of standardized skills (the sum of direct and indirect effects of skills, $\tilde{c}_{1k}+\sum_q \tilde{c}_{2kq}\tilde{a}_{1q}+\tilde{c}_{3k}(\tilde{b}_1+\sum_q \tilde{b}_{3q}\tilde{a}_{1q})$). Column (10) shows estimates of d_3 for education and $d_1+\sum_q d_{2q}\tilde{a}_{1q}+d_3(\tilde{b}_1+\sum_q \tilde{b}_{3q}\tilde{a}_{1q})$ for skills, calculated for the standardized latent health factor. Health behaviors, lifestyles, income, and work conditions are measured in 1992. Models for each behavior are conditional on background variables documented in Table IV.1. Sample size is 3961 for men and 4491 for women. Huber-White standard errors are reported in parentheses. Asterisks denote the level of statistical significance: ***, **, and * represent $p < 0.01$, 0.05, and 0.10. See Table IV.2 for definitions of mediators. Calculations are based on the WLS.

Table IV.5 investigates the effects of skills on education as well as the effects of education and skills on potential mediators among behavioral, socioeconomic, and health outcomes, all measured simultaneously at the start of the risk period in 1992 (about age 53). Column 1 of Table IV.5 shows that one standard deviation in socioemotional skills increases the likelihood of completing college by 26 percentage points (PP) in men and by 10 PP in women. Thus, socioemotional skills strongly affect college education, which itself affects many mediators of longevity, as we show below.

From columns 2–10 we see that college graduates of both genders have higher incomes and are more likely to engage in social activity. Additionally, educated men enjoy superior health and are less likely to smoke tobacco, participate in risky drinking of alcohol, and experience dangerous working conditions. Educated women are more likely to exercise.

The only exception to this pattern of health-beneficial effects of college education is marriage for women, which is negatively affected by a college degree. This result is in line with a well-documented historical pattern (e.g., England and Bearak, 2012), which is largely due to societal views regarding the role of women during that period. Since the 1970s, the marriage gap between educated and uneducated women has closed, and possibly even reversed sign. The change is largely related to greater marriage stability for educated women. In Section IV.3.5 we adjust our estimates for more recent cohorts that are no longer subject to this historical adverse effect.

Socioemotional skills also show multiple statistically significant effects, all of which are health-beneficial. For both genders, those with superior socioemotional skills are less likely to smoke tobacco and to drink alcohol in excess. They are more likely to exercise, be socially active, and have high income. In addition, superior socioemotional skills of men decrease the exposure to dangerous working conditions. Socioemotional skills in women lead to an increased likelihood of marriage. The effect on marriage is positive because the positive direct effect on marriage dominates the negative indirect effect on marriage through education.

Our results for the effects of socioemotional skills and college education on essential health behaviors, lifestyles, health stock, and earnings are auxiliary estimates that serve as building blocks for our decompositions. Quantitatively, these results are not directly comparable to other papers due to a combination of differences: definitions of variables, age group, population type, and the effect type (in this paper, the average treatment effect). Qualitatively, our results for effects of college education are consistent with the literature on tobacco smoking (Buckles et al., 2014; de Walque, 2007; Heckman et al., 2014), risky drinking of alcohol (Buckles et al., 2014), income (Heckman et al., 2014), social activity (Huang et al., 2009), and health stock (Heckman et al., 2014). Our estimates of the effects of education on overweight status and physical exercise are not precisely determined in men, but these estimates have signs consistent with findings in the literature (Buckles et al., 2014) and contribute to our decomposition by boosting the estimated total effect (in women, we find statistically significant effect of education on exercise). Finally, strong effects of socioemotional skills on essential health-related outcomes that we find are in accordance with Heckman et al. (2006) and Heckman et al. (2014).

IV.3.3 Effects of Potential Mediators on Longevity

Table IV.6 presents estimated parameters of the MPH model, showing the effects of mediators $\boldsymbol{B}_1$ and health stock Θ_1^H at the start of the risk period (1992) on the hazard of death. These estimates represent another building block for our decompositions. By controlling for latent health stock, we control for a potential confounder that may affect both behaviors and longevity and for health consequences of behaviors in the past.

We can see that several behaviors show effects on mortality for both men and women. As expected, smoking increases the hazard of death. In contrast, health stock, marriage, and income decrease the hazard of death. In men, we also observe statistically significant harmful effects of risky drinking of alcohol and dangerous working conditions. Overweight and physical exercise show sizable coefficients in the expected direction, but these estimates

Table IV.6: Effects of Potential Mediators on the Hazard of Death for Age 54–74, MPH Model Coefficients

	Males			Females		
	Estimates		PH test p-value[a]	Estimates		PH test p-value[a]
Smoking tobacco[b]	0.348	***	0.710	0.271	**	0.910
	(0.124)			*(0.132)*		
Risky drinking of alcohol	0.224	*	0.996	-0.009		0.383
	(0.129)			*(0.169)*		
Physical exercise[b]	-0.135		0.447	-0.077		0.665
	(0.117)			*(0.141)*		
Overweight[b]	0.183		0.494	0.118		0.378
	(0.448)			*(0.188)*		
Marriage	-0.481	***	0.744	-0.416	***	0.915
	(0.104)			*(0.093)*		
Social activity	-0.028		0.986	0.036		0.337
	(0.142)			*(0.116)*		
Household per capita income[b]	-0.104	*	0.992	-0.207	***	0.474
	(0.075)			*(0.089)*		
Dangerous working conditions[b]	0.142	*	0.994	0.017		0.997
	(0.090)			*(0.127)*		
Health stock[b]	-0.360	***	0.985	-0.336	**	0.637
	(0.128)			*(0.160)*		
Other controls[c]	Yes			Yes		
Joint test p-value[d]	0.000		0.954	0.004		0.884
Sample size	3961			4491		

Notes: Estimates of e_{4k}, $k = 1,...,8$ and of e_5 from equation (IV.6) are shown. The dependent variable is the hazard of death conditional on survival to January 1993, at which time the median age is 54. (We control for belonging to older and younger cohorts.) Mediators are measured in year 1992. Standard errors are shown in parentheses. Asterisks denote the level of statistical significance: ***, **, and * represent $p < 0.01$, 0.05 and 0.10. Coefficients for health stock represent the effect of a one standard deviation change in health stock. Calculations are based on the WLS. [a]We test the proportional hazard (PH) assumption by allowing the MPH model coefficients to differ by age and testing whether they are the same over ages. We perform this procedure both variable-by-variable for individual tests and overall for the joint PH test. [b] For these mediators, asterisks correspond to one-sided tests, chosen due to evidence in the literature about the direction of the effect. [c]"Other controls" include the cognitive and socioemotional skills, college education, and background variables listed in Table IV.1. [d]We test and reject the hypothesis that all MPH model structural coefficients are jointly zero, but do not reject the joint PH test.

are not precisely determined.

These results are in line with a long history of longevity research in the health and medical literatures. Studies find effects of earnings (Snyder and Evans, 2006), marriage (Rendall et al., 2011), smoking, obesity (Cutler and Lleras-Muney, 2008), and physical activity (Manini et al., 2006).

The effect of marriage on mortality that we estimate is large and similar across genders. Similar sizes of the effect for men and women are consistent with the meta-analysis by Manzoli et al. (2007), who summarize 54 estimates from the literature on the effect of marriage on longevity for the elderly. The relative risk derived from the meta-analysis is 0.88. Our average estimate of the relative risk is 0.64, which is in the 13th percentile of the distribution of estimates in Figure 1 from Manzoli et al. (2007), but the estimates are not directly comparable since we control for behaviors, lifestyles, earnings, working conditions, health, education, socioemotional skills, IQ, and rich background variables. In contrast, the literature summarized in the meta-analysis at most controls for age, gender, education, and health.

Table IV.6 also provides the results of specification tests. We test and do not reject the proportional hazard (PH) hypothesis that regression coefficients are constant over time for the risk period (see p-values for the individual and joint PH tests).

Among our list of mediators, health stock is special since it is a cumulative statistic that accounts for health behaviors and lifestyles in the past. To better understand the health stock, in Table IV.7 we document associations between health stock and other mediators at age 53 conditional on background controls, skills, and education. The correlations reflect what we would expect. Health stock positively correlates with beneficial health behaviors and negatively correlates with adverse health behaviors. Our interpretation is that health stock at age 53 is influenced by unobserved health behaviors in the past, which can be expected to correlate with health behaviors at age 53 due to addiction or persistence. One exception from this pattern is risky drinking of alcohol in females, which has a small but

Table IV.7: Conditional Associations between Health Stock and Other Mediators at Age 53

	Smoking tobacco	Risky drinking of alcohol	Physical exercise	Overweight	Marriage	Social activity	Household per capita income	Dangerous working conditions
Males								
Health stock	-0.034 ***	-0.001	0.066 ***	-0.088 ***	0.034 ***	0.009	0.076 ***	-0.008
	(0.011)	*(0.006)*	*(0.010)*	*(0.010)*	*(0.011)*	*(0.009)*	*(0.021)*	*(0.010)*
Females								
Health stock	-0.001	0.012 **	0.093 ***	-0.105 ***	0.013	0.002	0.085 ***	-0.018 **
	(0.009)	*(0.005)*	*(0.010)*	*(0.010)*	*(0.010)*	*(0.011)*	*(0.019)*	*(0.007)*

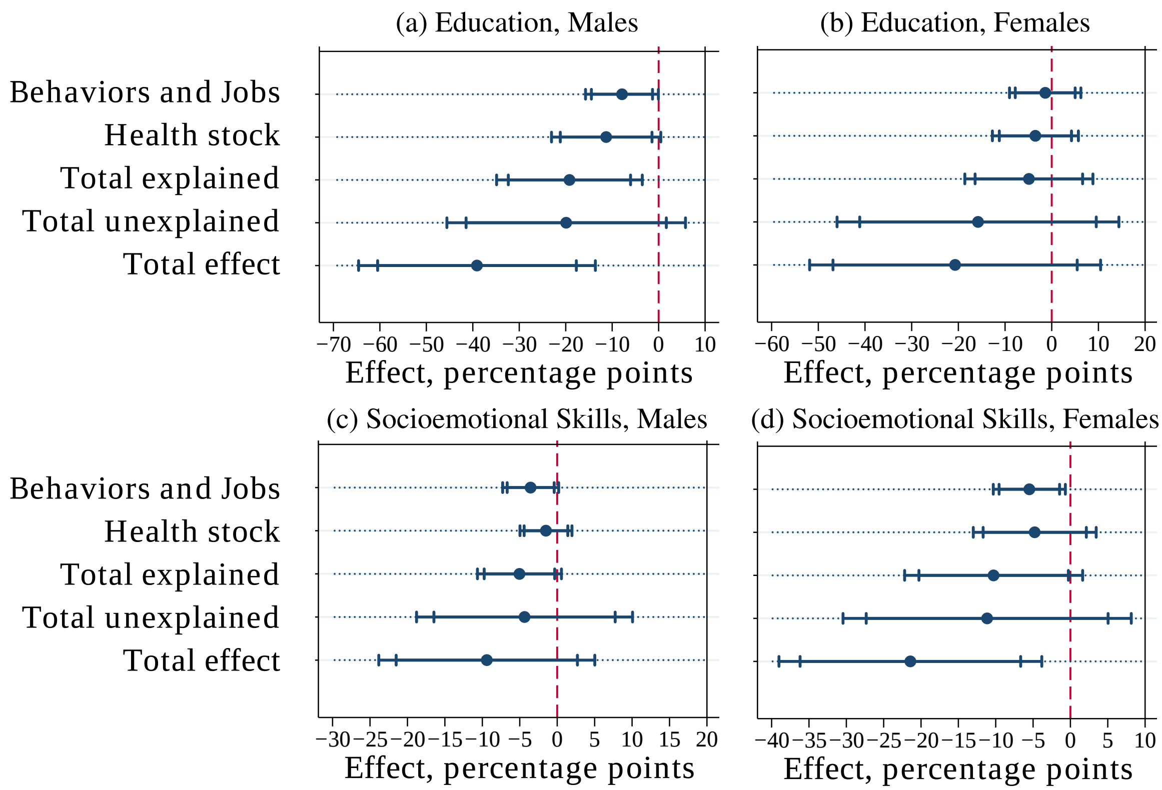
Notes: We regress each mediator on health stock, skills, education, and background variables. Standard errors are shown in parentheses. Coefficients correspond to a standardized health stock factor. Asterisks denote the level of statistical significance: ***, **, and * represent $p < 0.01$, 0.05, and 0.10. Calculations are based on the WLS.

positive and statistically significant correlation with heath stock. This association could be due to a reverse-causal mechanism: those who feel healthier in 1992 may also feel that it is safe for them to consume more alcohol.

IV.3.4 Decomposition of Effects of Education and Socioemotional Skills on Longevity

Figure IV.2 shows both explained and unexplained components of the effects of education and socioemotional skills on longevity using decompositions (IV.10–IV.12). Components due to behaviors, job-related outcomes, and health stock sum up to the total explained effect. We add the total explained effect to the total unexplained effect to obtain the total effect. We tend to explain sizable portions of the total effect across panels. Mediators related to behaviors and jobs do not explain effects of education in women (see panel (b)), a feature that we explain below. The large contribution of health stock to total effects as shown in panel (a) suggests that by age 53 men have differences in health stock by education, which lead to major differences in the hazard of death (11%). The *total explained* effect is statistically significant in all panels but (b). The *total* effect is statistically significant in all panels but (c) if a one-sided test is used for the effect of education.[20]

Figure IV.2: Explained, Unexplained, and Total Effects of Education and Skills on the Hazard of Death



Notes: Panels (a) and (b) share a common scale and represent decomposition (IV.10), in which the effect through observed mediators is referred as "total explained"; Panels (c) and (d) share another common scale and show decompositions (IV.11) and (IV.12) for one standard deviation change in socioemotional skills. "Total explained" refers to ψ_A defined by (IV.11). Total effect is $\psi_A + \psi_B$. "Total unexplained" refers to ψ_B, defined by (IV.12), and represents the portion of the total effect that is explained neither by mid-life behaviors nor by mid-life health stock. Inner and outer vertical bars represent the 90% and 95% Huber-White robust confidence intervals calculated using the delta method. See also Table A.13 for tabulated results. Calculations are based on the WLS.

We evaluate our decompositions not only in terms of relative change in the hazard of death, but also in terms of years of expected life at age 53 and the corresponding value of remaining life measured in 2012 US dollars.[21] In order to compute the expected longevity

[20] One-sided test for the total effect is motivated by abundant evidence from the literature that the total effect of education on longevity is nonnegative (Grossman, 2006; Grossman and Kaestner, 1997; Lochner, 2011). See Table A.13 for estimates, standard errors, and p-values.

[21] As Murphy and Topel (2006), we calculate changes in the value of remaining life induced by greater

gain, we use our MPH model to compute the baseline survival function up to the maximum observed uncensored data point. To avoid having a defective distribution, we use survival data of white men and women from the Census (Arias, 2012) to extrapolate the baseline survival function up to age 100. We use the methodology developed by Murphy and Topel (2006) to calculate the value of remaining life. To calibrate the model, we use a value of statistical life of 9.1 million USD adopted by the US Department of Transportation, which is in line with recent economic research (Viscusi, 2013). We find that the total effect of education on life expectancy at age 53 for men and women corresponds to 4.7 and 1.8 additional years of life worth 385,000 and 144,000 USD, respectively. Socioemotional skills add 1.4 and 3.1 years of life to men and women, worth 117,000 and 238,000 USD. See the Appendix for more details about these calculations.[22]

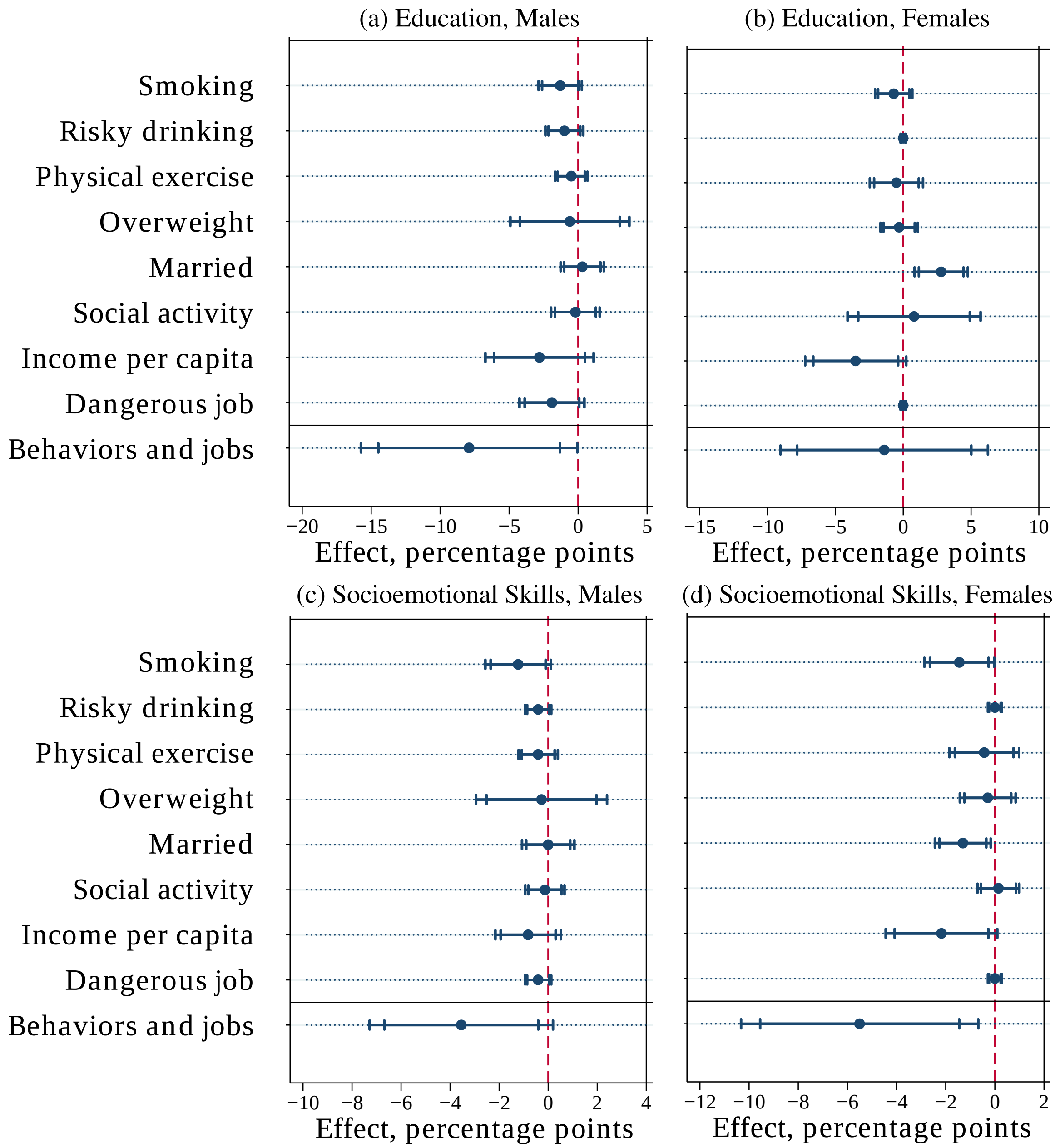
Figure IV.3 zooms in to the contribution of each specific mediator, such as smoking tobacco, in explaining the education- and skill-longevity gradient. Contributions of specific mediators are aggregated into "behaviors and jobs," a component that links Figure IV.3 to Figure IV.2. Specific components in the upper part of each panel allow us to better understand the mechanisms behind the aggregated mediation results. In particular, we see a statistically significant "behaviors and jobs" aggregate effect in panels (a), (c), and (d). In panels (a) and (c), representing effects of education and socioemotional skills for men, the major contributors to the total decrease in mortality include a reduction in smoking, risky drinking, and dangerous job exposure, as well as an increase in income. Some of these components are borderline statistically significant but they sum up to a statistically significant aggregate. For women, the negative effect of socioemotional skills on mortality is explained by lower smoking, higher likelihood of marriage, and greater income.

The zero aggregate effect for females in panel (b) representing the effect of education

longevity under constant quality of life. Evaluation of change in the quality of life induced by education and skills is beyond the scope of this paper, but since we show that education enhances not only longevity but also health, we can expect the full effects of education and skills on the value of life to be even greater than our estimates of their components related to greater longevity.

[22] Table A.13 evaluates all mediation effects in terms of life expectancy and the value of remaining life.

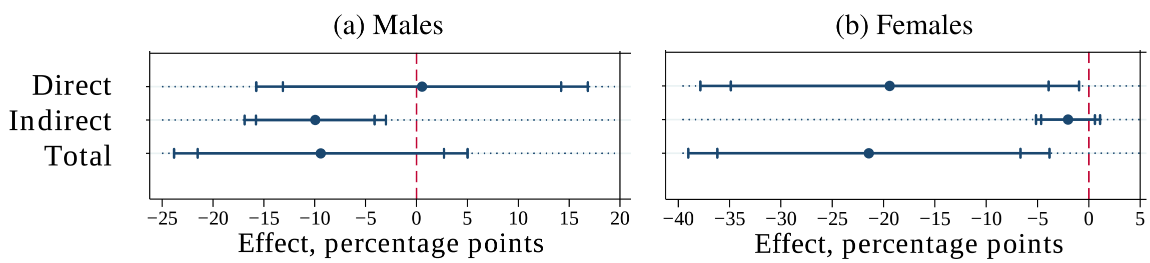
Figure IV.3: Decompositions of the Effect of Education and Socioemotional Skills on the Hazard of Death with Respect to Behavioral and Job-Related Mediators



Notes: Panels (a) and (b) share a common scale and represent decomposition (IV.10); Panels (c) and (d) share another common scale and show decomposition (IV.11) for one standard deviation change in socioemotional skills. "Behaviors and jobs" component refers to a sum of all individual components presented in the same panel. Inner and outer vertical bars represent the 90% and 95% Huber-White confidence intervals calculated using the delta method. See Table IV.2 for definitions of mediators and Table A.13 for tabulated results. Calculations are based on the WLS.

is not due to a lack of specific effects, but rather a result of effects canceling each other out. The beneficial effect of education on longevity acting through increase in income (about 3.5 PP) is negated by the adverse effect acting through the decline in the likelihood of marriage (2.8 PP). In contrast, in panel (d) we see a strong beneficial contribution of socioemotional skills through marriage. The reason for the difference between panels (b) and (d) is that there is a strong direct positive effect of socioemotional skills on marriage that is larger than the negative indirect effect through education, making the total contribution of marriage in panel (d) beneficial for longevity.[23]

Figure IV.4: Decomposition of the Total Effect of Socioemotional Skills on the Hazard of Death with Respect to Direct and Indirect Components



Notes: Decomposition (IV.13) is shown: ψ_C is the indirect effect. $\psi = \psi_A + \psi_B$, defined by (IV.11–IV.12), is the total effect. The direct effect is $\psi_D = \psi - \psi_C$. Effects correspond to one standard deviation increase in socioemotional skills. Panels share a common scale. Inner and outer vertical bars represent the 90% and 95% Huber-White robust confidence intervals calculated using the delta method. Calculations are based on the WLS.

Finally, Figure IV.4 shows the role of college education as a mediator of skills in the total effect (see equation (IV.13) representing the indirect effect ψ_C). For men, the indirect channel is strong, while the direct channel cannot be distinguished from zero. Skills strongly affect education, while education strongly affects many mediators of longevity. For women, the indirect effect is not statistically significant partly because effects of education through observed mediators of longevity, mainly marriage and earnings, cancel each

[23]Since decompositions in Figures IV.2 and IV.3 are evaluated at average levels of the standardized IQ, they do not show the contribution of the interaction between IQ and socioemotional skills (see equations (IV.2) and (IV.3)). We investigate this interaction and find that it plays a moderate role. When cognition is above average, a number of mediators of socioemotional skills become statistically significant and play a larger part. These results are documented in Figure A.1 of the Web Appendix.

other out. As a result, the explained effects of socioemotional skills for women are almost entirely driven by the direct effect.

IV.3.5 Adjustments for Different Environments and Implications for External Validity and Health Policy

Below we provide examples of adjustments to different environments, both actual (that have been realized for more recent cohorts) and hypothetical.

While our sample of females born around 1939 experienced negative effects of education on the probability of being married, more recent cohorts have had different experiences due to social changes in the role of women. Lefgren and McIntyre (2006) present OLS evidence that for cohorts born between 1955 and 1970, female college graduates have a 4.5 PP higher probability of being married in 2000 compared to high school graduates.[24] The authors find that this effect is driven mainly by greater stability of marriage rather than a higher probability of ever being married. The authors' alternative approach is based on using birth quarter as an instrument for education, which leads to a statistically insignificant effect.

We provide a simple simulation of the total effect of education on the hazard of death for a population in which the marginal effect of education on marriage for women is either zero or +4.5 PP, as opposed to -6.7 PP, as we find for our cohort of females. All other parameters of the model stay the same. We also adjust the variance-covariance matrix and re-estimate standard errors using the delta method.[25]

Once we apply these estimates to our model, we find that the total estimated effect of college education on the hazard of death for women increases from 21% to 25% if a 4.5 PP effect is assumed and to 24% if a zero effect is assumed (see Table IV.8). The

[24] We sum up the changes in the probability of being married in Figure 1 of Lefgren and McIntyre (2006): those with some college credits relative to high school graduates (about 0.6 PP), and college graduates relative to those with some college credits (about 3.9 PP).

[25] The adjusted standard errors do not account for errors generated by assumptions of this simulation.

corresponding change in the effect of socioemotional skills is small since only the indirect effect of skills through education is affected.[26] Our finding, therefore, suggests that the education-mortality gradient for more recent cohorts of women, who are no longer exposed to a negative relationship between marriage and education, is likely to be larger than the estimated effects for older cohorts.

The main limitation of this adjustment results from the simultaneous measurement of longevity mediators, which prevents us from modeling how changes in one mediator cause changes in others. Marriage has been linked to multiple health behaviors and lifestyles, although findings are mixed as to whether it is beneficial or harmful.[27] Overall, our adjustment is probably an underestimate since marriage is linked to higher household income for women, as shown by Lefgren and McIntyre (2006), and income is among the strongest observable mediators, as we see in Figure IV.3.

Table IV.8: A Simulation of the Effect of Education on the Hazard of Death for Females: Correcting the Total Effect Through Correcting its Component due to Marriage

	WLS cohort, -6.7 PP			Simulated cohort, 4.5 PP			Simulated cohort, 0 PP		
	Estimate	standard error	*p*-value[(a)]	Estimate	standard error	*p*-value[(a)]	Estimate	standard error	*p*-value[(a)]
Contribution of marriage	0.028 ***	0.010	0.006	-0.019 ***	0.004	0.000	0	-	-
Total effect[(b)]	-0.207 *	0.159	0.096	-0.254 *	0.159	0.055	-0.235 *	0.159	0.069

Notes: Huber-White standard errors are shown. [(a)]Asterisks denote the level of statistical significance: ***, **, and * represent $p < 0.01$, 0.05, and 0.10. [(b)]One-sided test is used for the total effect due to abundant evidence in the literature about the non-negative total effect of education on longevity (Grossman, 2006; Grossman and Kaestner, 1997; Lochner, 2011). Sample size is 4491.

Our estimates also allow us to analyze how hypothetical environments change estimated effects. Since smoking is a major mediator, we provide an example of a hypothetical additional tax on cigarettes and its role as a policy intervention aimed at reducing the education-health gradient as a major contributor to health inequality among other policy

[26]See Table A.15 of the Web Appendix.

[27]Wood et al. (2007) provide a detailed survey of the most statistically-rigorous studies relating marriage and health-related outcomes. They report that marriage is linked with lower cost of health care among older adults but also with reduced physical activity and modest weight gain. Evidence for the effect on smoking is mixed.

goals.

While our model does not estimate price elasticities, we can make predictions for different tax regimes by combining estimates from our model with estimates of price elasticities from the literature.

Our model allows us to counterfactually predict changes in life expectancy and the value of remaining life resulting from changes in skills and education.[28] We can also predict specific behaviors, such as smoking by education level. The counterfactual smoking probability for an average person can be viewed as the counterfactual average smoking rate in the population.

We draw from the meta-analysis in Gallet and List (2003) to obtain the median estimated price elasticity of smoking participation in the literature (-0.5).[29] A change in taxes of cigarettes will change average smoking levels by (price elasticity) $\times$ (% tax) $\times$ (average smoking level). We apply this change to counterfactual average levels of smoking of college graduates and non-graduates.

We find that an increase in the tax rate of cigarettes as large as 50% would reduce the college education-longevity gradient by two weeks for men and one week for women.[30] This change is small compared to the total college education-longevity gap of 4.3 years for men and 2.1 years for women (using the total for women adjusted in the previous section).[31]

The poor health of disadvantaged populations, as well as the related health gap between people with advantaged and disadvantaged backgrounds, are major policy concerns. Low parental socio-economic status often translates into low skills and low education levels of children and results in major health inequality, among other problems (Heckman, 2008). As our cigarette tax example shows, this gap is not easy to close, even with a heavy-handed approach aimed at a major observed mediator of mortality.

[28] See Appendix D for technical details.

[29] It is unlikely that price elasticities differ much by education level (Gruber and Koszegi, 2004).

[30] The health inequality reduction effect would be stronger if we would tax the uneducated more per cigarette than we tax the educated, but such a surtax is regressive and politically infeasible.

[31] 1.8 years for women before correction for the education-marriage relationship.

To complement traditional health policies, the gap between populations with affluent and disadvantaged backgrounds can be decreased through enhancing the following developmental causes of health in disadvantaged populations: (1) socioemotional skills during childhood and adolescence (Heckman and Kautz, 2014);[32] and (2) college education attainment in the case of market failures leading to sub-optimal investment levels in education (Cutler and Lleras-Muney, 2008). This paper quantifies effects of such policies and shows that many of them operate through healthier lifestyles and superior earnings. This paper shows that the effects of skills and education on longevity act through many mediators, so it may be more efficacious to target skills and education in disadvantaged populations early in life compared to traditional remedial policies targeting specific health-related outcomes once they are developed.

IV.4 Conclusions

We explore a number of mediators, including health behaviors, lifestyles, income, work conditions, and health stock, that together explain a sizable portion of the effects of cognitive skills, socioemotional skills, and college education on longevity. The role of job-related mediators is substantial, but behavioral mediators show comparable importance.

Our results are heterogenous by both skill and gender. Additionally, some mediation effects cancel each other out. Such conflicting mediation effects of education for women through marriage on the one hand and smoking on the other contribute to our understanding of the historic gender gap in the effect of education on longevity.

Uncovering these mechanisms reinforces the claims of causal links that are debated in the literature. The non-dominant role of income and job conditions highlights the importance of health behavior choices. Finally, identifying these mechanisms opens the door for modeling and predicting the effects of education and skills for different economic environ-

[32] Socioemotional skills are also malleable in young adulthood, but our main model only quantifies effects of socioemotional skills as developed by high school graduation.

ments.

We offer a methodology that can deliver more detailed results if applied to a superior data source. Therefore, it is important to improve the design of new longitudinal surveys aimed at the study of life cycle health by making multifaceted measurements of both cognitive and socioemotional skills in childhood and adolescence, as well as detailed measurements of health, health behaviors, earnings, lifestyles, and working conditions over the whole life cycle.

Abdulkadiroglu, A., J. D. Angrist, and P. A. Pathak 2014. The Elite Illusion: Achievement Effects at Boston and New York Exam Schools. *Econometrica 82*(1), 137–196.

Aiayi, K. F. 2014. Does School Quality Improve Student Performance? New Evidencce from Ghana. Unpublished manuscript, Boston University.

Aitken, M. and D. B. Rubin 1985. Estimation and hypothesis testing in finite mixture models. *Journal of Royal Statistical Society, Series B 47 (1)*.

Anderson, T. W. and H. Rubin 1956. Statistical inference in factor analysis. In J. Neyman (Ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 5*, pp. 111–150. Berkeley: University of California Press.

Angrist, J. D. and M. Rokkanen 2015. Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away from the Cutof. *Journal of American Statistical Association 110*(512), 1331–1344.

Arias, E. 2012, September. United States life tables, 2008. *National Vital Statistics Reports 61*(3).

Asparouhov, T., K. Masyn, and B. Muthen 2006. Proceedings of the Joint Statistical Meeting in Seattle, August 2006. In *ASA Section on Biometrics*, pp. 180–187.

Balia, S. and A. Jones 2008. Mortality, lifestyle, and socio-economic status. *Journal of Health Economics 27*, 1–26.

Bargain, O. and K. Doorley 2013. Putting structure on the rd design: Social transfers and youth inactivity in france. Discussion Paper Series 7508, IZA.

Bedard, K. and C. Do 2005. Are Middle Schools More Effective? The Impact of School Structure on Student Outcomes. *Journal of Human Resources 40*(3), 660–682.

Beijing Municipal Commission of Education 2013. Proposal of 2014-2016 Reform of High School Entrance Exam and Admission. Accessed April 30, 2016. http://www.bjedu.gov.cn/publish/portal0/tab67/info28951.htm.

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

Blincoe, J. M. 2009. *The Age and Condition of Texas High School as Related to Student Academic Achievement*. Ph. D. thesis, University of Texas at Austin.

Bollen, K. A. 1989. *Structural Equations with Latent Variables*. New York: Wiley.

Borghans, L., B. H. H. Golsteyn, J. J. Heckman, and J. E. Humphries 2011. Identification problems in personality psychology. *Personality and Individual Differences 51*(3: Special Issue on Personality and Economics), 315–320.

Brunello, G., M. Fort, N. Schneeweis, and R. Winter-Ebmer 2015. The causal effect of education on health: What is the role of health behaviors. *Health Economics*.

Buckles, K., A. Hagemann, O. Malamud, M. Morrill, and A. Wozniak 2014. The Effect of College Education on Mortality. Unpublished manuscript, University of Notre Dame.

Calonico, S., M. Cattaneo, and R. Titiunik 2014. Robust Nonparametric Confidence Intervals for Regression-discontinuity Designs. *Econometrica 82*(6), 2295–2326.

Cameron, A. C. and P. K. Trivedi 2005. *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Cattaneo, M. D., B. R. Frandsen, and R. Titiunik 2015. Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the u.s. Senate. *Journal of Causal Inference 3*(1), 1–24.

Cawley, J. and C. J. Ruhm 2012. The economics of risky health behaviors. *Handbook of Health Economics 2*, 95–199.

Cellini, S. R., F. Ferreria, and J. Rothstein 2010. The Value of School Facility Investments: Evidence from a Dynamic Regression Discontinuity Design. *The Quarterly Journal of Economics 125*(1), 215–261.

Clark, D. 2010. Selective Schools and Academic Achievement. *The B.E. Journal of Economic Analysis & Policy 10*(1), Article 9.

Clark, D. and D. B. Emilia 2016. The Long-Run Effects of Attending an Elite School: Evidence from the United Kingdom. *American Economic Journal: Applied Economics 8*(1), 150–176.

Conti, G., J. J. Heckman, and S. Urzúa 2010, May. The education-health gradient. *American Economic Review: Papers and Proceedings 100*(2), 1–5.

Contoyannis, P. and A. M. Jones 2004. Socio-economic status, health, and lifestyle. *Journal of Health Economics 23*, 965–995.

Cook, P., R. MacCoun, C. Muschkin, and J. Vigdor 2008. The Negative Impacts of Starting Middle School in Sixth Grade. *Journal of Policy Analysis and Management 27*(1), 104–121.

Cullen, J., B. Jacob, and S. Levitt 2006. The Effect of School Choice on Student Outcomes: Evidence from Randomized Lotteries. *Econometrica 74*(5), 1191–1230.

Cunha, F., J. J. Heckman, and S. Navarro 2005, April. Separating uncertainty from heterogeneity in life cycle earnings, The 2004 Hicks Lecture. *Oxford Economic Papers 57*(2), 191–261.

Cutler, D. M. and A. Lleras-Muney 2008. Education and health: Evaluating theories and evidence. In R. F. Schoeni, J. S. House, G. A. Kaplan, and H. Pollack (Eds.), *Making Americans Healthier: Social and Economic Policy as Health Policy*, pp. 29–60. New York: Russell Sage Foundation.

Cutler, D. M., A. Lleras-Muney, and T. Vogl 2011. Socioeconomic status and health: Dimensions and mechanisms. *In S. Glied and P. Smith, eds the Oxford Handbook of Health Economics*.

de Raad, B. and H. C. Schouwenburg 1996. Personality in learning and education: a review. *European Journal of Personality 10*, 303–336.

de Walque, D. 2007. Does education affect smoking behaviors?: Evidence using the Vietnam draft as an instrument for college education. *Journal of Health Economics 26*(5), 877 – 895.

Dejong, W. and T. Glover 2003, Feb. Urban school facilities. *School Planning & Management*.

Dhuey, E. 2013. Middle School or Junior High? How Grade-Level Configurations Affect Academic Achievement. *Canadian Journal of Economics 46*(2), 469–496.

Dobbie, W. and R. G. Fryer Jr. 2014. The Impact of Attending a School with High-Acheving Peers: Evidence from the New York City Exam Schools. *American Economic Joournal: Applied Economics 6*(3), 58–75.

Dong, Y. and A. Lewbel 2015. Identifying the Effect of Changing the Policy Threshold in Regression Discontinuity Models. *Review of Economics and Statistics 97*(5), 1081–1092.

Duflo, E., P. Dupas, and M. Kremer 2011. Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review 101*(5), 1739–1774.

Eccles, J., C. Midgley, and T. Alder 1984. *Advances in Motivation and Achievement, Vol. 3*, Chapter Grade Related Changes in the School Environment: Effects on Achievement Motivation, pp. 283–331. JAI Press.

Eccles, J. S. and C. Midgley 1989. *Research on Motivation in Education, Vol. 3*, Chapter Stage/Environment Fit: Developmentally Appropriate Classrooms for Younger Adolescents, pp. 139–186. Academic Press.

Elias, M. J., M. Gara, and M. Ubriaco 1985. Sources of Stress and Support in Children's Transition to Middle School: An Empirical Analysis. *Journal of Clinical Child Psychology 14*(2), 112–118.

Elsner, B. and I. E. Isphording 2015. A Big Fish in a Small Pond: Ability Rank and Human Capital Investment. Discussion Paper Series 9121, IZA.

England, P. and J. Bearak 2012. Women's education and their likelihood of marriage: A historic reversal. A fact sheet prepared for the council on contemporary families. Technical report, New York University.

Farrell, P. and V. R. Fuchs 1982. Schooling and health: the cigarette connection. *Journal of Health Economics 1*(3), 217–230.

Feldlaufer, H., C. Midgley, and J. Eccles 1988. Student, Teacher, and Observer Perceptions of the Classroom Environment before and after the Transition to Junior High School. *Journal of Early Adolescence 8*(2), 133–156.

Fischer, C. S. and M. Hout 2008. *Century of Difference: How America Changed in the Last One Hundred Years*. Russell Sage Foundation.

Galama, T. and H. v. Kippersluis 2015. A theory of education and health. CESR-Schaeffer Working Paper No 2015-001.

Gallet, C. A. and J. A. List 2003. Cigarette demand: a meta-analysis of elasticities. *Health Economics 12*, 821–835.

Goff, M. and P. L. Ackerman 1992. Personality-intelligence relation: Assessment of typical intellectual engagement. *Journal of Educational Phychology 84*(4), 537–552.

Goldberg, L. R., J., E. John A., H. Herbert W., A. Robert, C. Michael C., R. C., and H. G. Gough 2006. The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality 40*, 84–96.

Goldin, C. 1997. Career and family: College women look to the past. In F. Blau and R. Ehrenberg (Eds.), *Gender and Family Issues in the Workplace*. Russell Sage Press, New York.

Grossman, M. 1972, March-April. On the concept of health capital and the demand for health. *Journal of Political Economy 80*(2), 223–255.

Grossman, M. 2000. The human capital model. In A. J. Culyer and J. P. Newhouse (Eds.), *Handbook of Health Economics*, Volume 1, pp. 347–408. Amsterdam: Elsevier.

Grossman, M. 2006. Education and nonmarket outcomes. In E. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, Volume 1, Chapter 10, pp. 577–633. Amsterdam: Elsevier.

Grossman, M. 2015, October. The relationship between health and schooling: What's new? Working Paper 21609, National Bureau of Economic Research.

Grossman, M. and R. Kaestner 1997. Effects of education on health. In J. R. Behrman and N. Stacey (Eds.), *The Social Benefits of Education*, pp. 69–124. Ann Arbor, MI: University of Michigan Press.

Gruber, J. and B. Koszegi 2004. Tax incidence when individuals are time-inconsistent: the case of cigarette excise taxes. *Journal of Public Economics 88*, 1959–1987.

Gunter, W. D. and N. W. Bakken 2010. Transitioning to Middle School in the Sixth Grade: A Hierarchical Linear Modeling (HLM) Analysis of Substance Use, Violence, and Suicidal Thoughts. *The Journal of Early Adolescence 30*(6), 895–915.

Hahn, J., P. E. Todd, and W. Van der Klaauw 2001, January. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica 69*(1), 201–209.

Hainmueller, J., A. B. Hall, and J. M. Snyder 2014. Assessing the External Validity of Election RD Estimates: An Investigation of the Incumbency Advantage. Stanford University Graduate School of Business Research Paper 14–02, Stanford University Graduate School of Business.

Hanushek, E. A. 1979. Conceptual And Empirical Issues in The Estimation of Educational Production Functions. *Journal of Human Resources 14*(3), 351–388.

Hanushek, E. A., J. F. Kain, and S. G. Rivkin 2004. Disruption vs. Tiebout Improvement: the Costs and Benefits of Switching Schools. *Journal of Public Economics 88*(9–10), 1721–1746.

Hastings, J. S. and J. M. Weinstein 2008. Information, School Choice, and Academic Achievement: Evidence from Two Experiments. *The Quarterly Journal of Economics 123*(4), 1373–1414.

Hauser, R. M. and W. H. Sewell 1957-2005. Wisconsin Longitudinal Study (WLS) [graduates, siblings, and spouses]: 1957-2005 Version 13.01.

Hauser, R. M. and R. J. Willis 2005. Survey Design and Methodology in the Health and Retirement Study and the Wisconsin Longitudinal Study. In L. J. Waite (Ed.), *Aging, Health, and Public Policy: Demographic and Economic Perspectives*, pp. 209–235. New York: Population Council.

Heckman, J., R. Pinto, and P. Savelyev 2013a. Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review 103*(6), 2052–2086.

Heckman, J. J. 2008, July. Schools, skills and synapses. *Economic Inquiry 46*(3), 289–324.

Heckman, J. J., J. E. Humphries, and G. Veramendi 2014. Education, health and wages. Unpublished manuscript, University of Chicago, Department of Economics.

Heckman, J. J. and T. Kautz 2014. Fostering and Measuring Skills: Interventions That Improve Character and Cognition. In *The Myth of Achievement Tests: The GED and the Role of Character in American Life*. Chicago, IL: University of Chicago Press.

Heckman, J. J., R. Pinto, and P. A. Savelyev 2013b. Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review 103*(6), 1–35.

Heckman, J. J. and B. S. Singer 1984, March. A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica 52*(2), 271–320.

Heckman, J. J., J. Stixrud, and S. Urzúa 2006, July. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics 24*(3), 411–482.

Heckman, J. J. and E. J. Vytlacil 2005, May. Structural equations, treatment effects and econometric policy evaluation. *Econometrica 73*(3), 669–738.

Heckman, J. J. and E. J. Vytlacil 2007. Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, pp. 4779–4874. Amsterdam: Elsevier.

Holt-Lunstad, J., T. B. Smith, and J. B. Layton 2010, 07. Social Relationships and Mortality Risk: A Meta-analytic Review. *PLoS Med 7*(7), e1000316.

Hong, K., P. Savelyev, and K. Tan 2014. Understanding the mechanisms linking personality and education with longevity. Unpublished manuscript, Vanderbilt University, Department of Economics.

Hong, K. and R. Zimmer 2015. Does Investing in School Capital Infrastructure Improve Student Achievement? *SSRN eLibrary*. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2672102.

Hough, D. L. 2005. The Rise of the "Elemiddle" School. *School Administrator 62*(3), 10–14.

Hoxby, C. and W. Gretchen 2006. Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects. Unpublished manuscript, Harvard University.

Huang, J., H. M. van den Brink, and W. Groot 2009. A meta-analysis of the effect of education on social capital. *Economics of Education Review 28*, 454–464.

Imbens, G. and K. Kalyanaraman 2012. Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *Review of Economic Studies 79*(3), 933–959.

Imbens, G. and T. Lemieux 2008. Regression discontinuity designs: A guide to practice. *Journal of Econometrics 142*(2), 615–635.

Jackson, C. K. 2010. Do Student Benefit from Attending Better Schools? Evidence from Rule-Based Student Assignments in Trinidad and Tobago. *The Economic Journal 120*(549), 1399–1429.

Jackson, C. K. 2013. Can Higher-Achieving Peers Explain the Benefits to Attending Selective Schools? Evidence from Trinidad and Tobago. *Journal of Public Economics 108*, 63–77.

John, O. P. and S. Srivastava 1999. *Handbook of Personality: Theory and Research* (2 ed.)., Chapter The Big Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives, pp. 102–138. New York: The Guilford Press.

Jones, J. T. and R. Zimmer 2001. Examining the Impact of Capital on Academic Achievement. *Economics of Education Review 20*(6), 577–588.

Juvonen, J., V. Le, T. Kaganoff, C. Augustine, and L. Constant 2004. Focus on the Wonder Years: Challenges Facing the American Middle School. Technical report, RAND Corporation.

Krueger, A. B. 2003, February. Economic considerations and class size. *Economic Journal 113*(485), F34–F63.

Lai, F., E. Sadouolet, and A. de Janvry 2011. The Contributions of School Quality and Teacher Qualifications to Student Peformance: Evidence from a Natural Experiment in Beijing Middle Schools. *Journal of Human Resources 46*(1), 123–153.

Lee, D. S. 2008. Randomized Experiments from Non-random Selection in u.s. House Elections. *Journal of Econometrics 142*(2), 675–697.

Lee, D. S. and T. Lemieux 2010. Regression Discontinuity Designs in Economics. *Journal of Economic Literature 48*(2), 281–355.

Lefgren, L. and F. McIntyre 2006. The Relationship between Women's Education and Marriage Outcomes. *Journal of Labor Economics 24*, 787–830.

Lippold, M. A., C. J. Powers, A. K. Syvertsen, M. E. Feinberg, and M. T. Greenberg 2013. The Timing of School Transitions and Early Adolescent Problem Behavior. *The Journal of Early Adolescence 33*(6), 821–844.

Lochner, L. 2011, January. Non-production benefits of education: Crime, health, and good citizenship. Working Paper 16722, National Bureau of Economic Research.

Lodi-Smith, J., J. Jackson, T. Bogg, K. Walton, D. Wood, P. Harms, and B. W. Roberts 2010. Mechanisms of health: Education and health-related behaviours partially mediate the relationship between conscientiousness and self-reported physical health. *Psychology & Health 25*(3), 305–319.

Lucas, A. M. and I. M. Mbiti 2014. Effects of School Quality on Student Achievement: Discontinuity Evidence from Kenya. *American Economic Journal: Applied Economics 6*(3), 234–263.

Ludwig, J. and D. L. Miller 2007. Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design. *The Quaterly Journal of Economics 122*(1), 159–208.

Maddala, G. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press.

Manini, T. M., J. E. Everhart, K. V. Patel, D. A. Schoeller, L. H. Colbert, M. Visser, F. Tylavsky, D. C. Bauer, B. H. Goodpaster, and T. B. Harris 2006. Daily Activity Energy Expenditure and Mortality among Older Adults. *Journal of the American Medical Association 296*, 171–179.

Manzoli, L., P. Villari, G. M. Pirone, and A. Bocca 2007. Marital status and mortality in the elderly: A systematic review and meta-analysis. *Social Science and Medicine 64*, 77–94.

Marsh, H. W., D. Chessor, R. Craven, and L. Roche 1995. The Effects of Gifted and Talented Programs on Academic Self-Concept: The Big Fish Strikes Again. *American Educational Research Journal 32*(2), 285–319.

Martorell, P., I. McFarlin Jr, and S. Kevin 2015. Investing in Schools: Capital Spending, Facility Conditions, and Student Achievement. NBER Working Paper 21515, National Bureau of Economic Research.

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

Mendolia, S. and I. Walker 2014. The Effect of Noncognitive Traits on Health Behaviour in Adolescence. *Health Economics 23*, 1146–1158.

Michigan Department of Treasury 2015. Bond Qualification Process Overview. Accessed October 28, 2015. https://www.michigan.gov/documents/3160_2815_7.pdf.

Michigan Department of Treasury AND Office of Revenue and Tax Analysis 2009. The Effects of an Aging Population on the Components of Michigan Income. Accessed October 28, 2015. https://www.michigan.gov/documents/treasury/Components_of_Michigan_Income-Effects_of_An_Aging_Population_340170_7.pdf.

Midgley, C., H. Feldlaufer, and J. S. Eccles 1989. Student/teacher Relations and Attitudes toward Mathematics before and after the Transition to Junior High School. *Child Development 60*(4), 981–992.

Miller, D. 1970. Adolescence and the High School System. *Community Mental Health Journal 6*(6), 483–491.

Mroz, T. A. 1999. Discrete factor approximation in simultaneous equation models: Estimating the impact of a dummy endogenous variable on a continuous outcome. *Journal of Econometrics 92*(2), 233–274.

Murphy, K. M. and R. H. Topel 2006, October. The value of health and longevity. *Journal of Political Economy 114*(5), 871–904.

Muthen, B. 1983. Latent variable structural equation modeling with categorical data. *Journal of Econometrics 22*, 43–65.

National Center for Education Statistics 2015. Digest of Education Statistics, 2013. Accessed October 28, 2015. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2015011.

Nielson, C. and S. Zimmerman 2014. The Effect of School Construction on Test Scores, School Enrollment, and Home Prices. *Journal of Public Economics 120*, 18–31.

Park, A., X. Shi, C.-t. Hsien, and X. An 2015. Magnet High Schools and Academic Performance in China: A Regression Discontinuity Design. HKUST IEMS Working Paper 2015-07, HKUST.

Picus, L., S. Marion, N. Calvo, and W. Glenn 2005. Understanding the Relationship between Student Achievement and the Quality of Educational Facilities: Evidence from Wyoming. *Peabody Journal of Education 80*(3), 71–95.

Pop-Eleches, C. and M. Urquiola 2013. Going to a Better School: Effects and Behavioral Responses. *American Economic Review 103*(4), 1289–1324.

Pratt, E., C. S. Mock, D. Zin, and C. Diehr 2012. School Capital Expenditure Fiance in Michigan Issues and Alternatives. Discussion paper, Senate Fiscal Agency. Accessed October 28, 2015. http://www.senate.michigan.gov/sfa/Publications/Issues/SchoolCapital/SchoolCapital.pdf.

Rendall, M. S., M. M. Weden, M. M. Favreault, and H. Waldron 2011. The Protective Effect of Marriage for Survival: A Review and Update. *Demography 48*, 481–506.

Rockoff, J. E. and B. Lockwood 2010. Stuck in the Middle: Impacts of Grade Configuration in Public Schools. *Journal of Public Economics 94*(11–12), 1051–1061.

Rokkanen, M. 2015. Exam Schools, Ability, and the Effects of Affirmative Action: Latent Factor Extrapolation in the Regression Discontinuity Design. Unpublished manuscript, Columbia University, Department of Economics.

Rubin, D. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: J. Wiley & Sons.

Rudolph, K. D., S. F. Lambert, A. G. Clark, and K. D. Kurlakowsky 2001. Negotiating the Transition to Middle School: The Role of Self-Regulatory Processes. *Child Development 72*(3), 929–946.

Salam, R. 2014. Close Stuyvesant High School: Why Super-Elite Public Magnet Schools Aren't Necessary Anymore. *Slate*. Accessed April 30, 2016. http://www.slate.com/articles/news_and_politics/politics/2014/07/the_case_for_shutting_down_stuyvesant_high_school_the_best_public_school.html.

Savelyev, P. 2014, May. Psychological skills, education, and longevity of high-ability individuals. Working paper.

Savelyev, P. A. 2015. Socioemotional Skills, Education, and Longevity of High-Ability Individuals. Unpublished manuscript, Vanderbilt University, Department of Economics.

Savelyev, P. A. and K. Tan 2015. Socioemotional Skills, Education and Health-Related Outcomes of High-Ability Individuals. Unpublished manuscript, Vanderbilt University, Department of Economics.

Schafer, J. 1997. *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall.

Schneider, M. 2002. Do School Facilities Affect Academic Outcomes? Information analyses, National Clearinghouse for Educational Facilities. Accessed October 28, 2015. http://www.ncef.org/pubs/outcomes.pdf.

Schochet, P. Z. 2009. Statistical Power for Regression Discontinuity Designs in Education Evaluations. *Journal of Educational and Behavioral Statistics 34*(2), 238–266.

Schwartz, A., L. Stiefel, R. Rubenstein, and J. Zabel 2011. The Path Not Taken: How Does School Organization Affect Eighth-Grade Achievement? *Education Evaluation and Policy Analysis 33*(3), 293–317.

Schwerdt, G. and M. R. West 2013. The Impact of Alternative Grade Configurations on Student Outcomes through Middle and High School. *Journal of Pubic Economics 97*, 308–326.

Snyder, S. E. and W. N. Evans 2006. The Effect of Income on Mortality: Evidence from the Social Security Notch. *The Review of Economics and Statistics 88*, 482–495.

South, S. J., D. Haynie, and S. Bose 2007. Student Mobility and School Dropout. *Social Science Research 36*(1), 68–94.

Turiano, N. A., P. L. Hill, B. W. Roberts, A. Spiro, and D. K. Mroczek 2012. Smoking mediates the effect of conscientiousness on mortality: The veterans affairs normative aging study. *Journal of Research in Personality 46*(6), 719–724.

Van Den Berg, G. J. 2001. Duration models: Specification, identicification, and multiple durations. *Handbook of Econometrics 5*, 3381–3460.

Van Oort, F. V. A., F. J. van Lenthe, and J. P. Mackenbach 2005. Material, psychosocial, and behavioural factors in the explanation of educational inequalities in mortality in the Netherlands. *Journal of Epidemiology and Community Health (1979-) 59*(3), 214–220.

Viscusi, W. K. 2013, October. Estimating the value of a statistical life using census of fatal occupational injuries (CFOI) data. *Monthly Labor Review*, 1–17.

Wang, W. and W. D. Duncombe 2009. School Facilities Funding and Capital-Outlay Distribution in the States. *Journal of Education Finance 34*(3), 324–350.

Weiss, C. C. and L. Kipnes 2006. Reexamining Middle School Effects: A Comparison of Middle Grade Students in Middle Schools and K-8 Schools. *American Journal of Education 112*(2), 239–272.

Welsh, W., E. Coghlan, B. Fuller, and L. Dauter 2012. New Schools, Overcrowding Relief, and Achievement Gains in Los Angeles — Strong Returns from a $19.5 Billon Investment. Policy Brief 12-2, PACE. http://edpolicyinca.org/sites/default/files/pace_pb_08.pdf.

Wing, C. and T. D. Cook 2013. Strengthening the Regression Discontinuity Design Using Additional Design Elements: A Within-Study Comparison. *Journal of Policy Analysis and Management 32*(4), 853–877.

Wood, R. G., B. Goesling, and S. Avellar 2007. The effects of marriage on health: A synthesis of recent research evidence. Technical report, Mathematica Policy Research, Inc.

Xu, Z., J. Hannaway, and S. D'Souza 2009. Student Transience in North Carolina: The Effect of Mobility on Student Outcomes Using Longitudinal Data. Working Paper 22, National Center for Analysis of Longitudinal Data in Education Research.

Zhang, H. 2014. The Mirage of Elite Schools: Evidence from Lottery-Based School Admissions in China. Working Paper, Chinese University of Hong Kong.

Appendices

A Supplemental Tables and Figures

Table A.1: School Bond Summary Statistics

				Average vote share (%)		Average bond amount per pupil ($)[a]	
Year	Number of bonds	Percentage passed	Number of voters	mean	standard deviation	mean	standard deviation
1996	164	51%	2536	49.4	11.9	7312	5423
1997	149	43%	2497	48.3	11.2	7764	5670
1998	107	41%	2644	48.5	10.7	9472	7965
1999	117	48%	2097	49.8	11.7	8348	6388
2000	117	49%	2364	49.8	12.9	7694	5835
2001	108	63%	2469	53.3	12.6	7487	6093
2002	83	59%	2560	52.3	14.8	7882	6440
2003	70	39%	3361	46.7	14.7	9820	10590
2004	71	63%	3034	53.9	15.3	9243	6661
2005	58	40%	2558	48.9	11.8	9981	10131
2006	59	44%	3740	48.0	11.5	7771	6827
2007	68	47%	2660	48.2	13.4	8033	8179
2008	44	57%	2320	51.6	12.6	7598	4791
2009	50	70%	5386	54.7	12.2	6087	6375
Total	1265	50%	2722	50.0	12.7	8123	6907

Notes: The sample includes bonds with non-missing values in both passage and vote share. [a]Average bond amount per pupil is measured in constant 2000 dollars.

Table A.2: District Descriptive Statistics of Expenditure, Demography and Achievement

	All school district		Never proposed an election		Ever proposed an election	
	mean	standard deviation	mean	standard deviation	mean	standard deviation
Expenditure per pupil[(a)]						
Total	8849	3981	8898	4083	8835	3953
Current	7319	3038	7804	3579	7183	2855
Teacher salary	3194	1400	3428	1544	3129	1351
Capital[(b)]	714.7	1713	399.6	1157	802.7	1828
Construction	625.3	1644	337.3	1103	705.6	1757
Land and structure	89.48	448.6	62.24	332.8	97.08	475.7
Instructional equipment[(c)]	46.31	87.50	61.53	151.2	42.06	57.73
Demography						
Enrollment	2937	7249	1845	2818	3243	8036
White students[(d)]	87.76	18.12	87.23	19.53	87.91	17.70
Free lunch[(d)(e)]	32.71	18.89	33.05	20.95	32.62	18.26
Achievement (proficiency)[(d)]						
4th grade reading	70.15	18.54	68.72	19.98	70.49	18.18
7th grade reading	61.83	18.85	61.67	19.79	61.87	18.62
Number of districts	577		144		433	
Sample size	7751		1693		6058	

Notes: Estimation sample includes districts with no missing information in proficiency, election time and vote share. [(a)]All expenditures per pupil are measured in constant year 2000 dollars. [(b)]Capital is defined as the sum of construction, land, and structure. [(c)]Instruction equipment includes expenditures for all instructional equipment recorded in general and operating funds under "instruction". [(d)]White students, free lunch and proficiency are measured in percentage point. [(e)]A student can receive free lunch if the households income is within the limits on the Federal Income Eligibility Guidelines or one of several other conditions is satisfied. For details see http://www.michigan.gov/mde/0,4615,7-140-66254_50144-194552--,00.html

Table A.3: ITT Effect of Bond Passage on Achievement, Subsequent Bond Passage and Capital Expenditure

Relative year	4th grade (1)	7th grade (2)	Subsequent bond passage (3)	Capital expenditure (4)
1	-0.459	-0.505	0.083	734.4 ***
	(0.553)	*(0.578)*	*(0.158)*	*(95.19)*
2	-0.089	-0.867	-0.259 **	2935 ***
	(0.593)	*(0.614)*	*(0.128)*	*(193.8)*
3	-0.136	0.168	-0.169	1007 ***
	(0.632)	*(0.616)*	*(0.119)*	*(187.0)*
4	0.043	0.629	-0.120	-878.1 ***
	(0.654)	*(0.632)*	*(0.127)*	*(163.8)*
5	-0.128	-0.679	-0.287 ***	-1156 ***
	(0.694)	*(0.655)*	*(0.110)*	*(158.3)*
6	-0.550	-0.290	-0.191	-733.5 ***
	(0.685)	*(0.747)*	*(0.133)*	*(187.2)*
7	-0.665	-0.003	-0.178	-524.4 ***
	(0.692)	*(0.804)*	*(0.135)*	*(147.0)*
8	-0.494	-0.217	0.306 **	-501.6 ***
	(0.701)	*(0.755)*	*(0.144)*	*(143.1)*
9	-1.291 *	-0.595	0.070	-164.4
	(0.731)	*(0.774)*	*(0.146)*	*(140.4)*
10	-0.828	-0.926	0.012	*61.00*
	(0.761)	*(0.825)*	*(0.176)*	*(154.6)*
11	-1.062	-0.810	-0.189	*10.33*
	(0.893)	*(0.967)*	*(0.235)*	*(177.6)*
12	-0.643	-1.972 *	-0.421	-260.3
	(1.117)	*(1.013)*	*(0.304)*	*(229.9)*
Sample size	9665	9663	2665	9829

Notes: The table shows the coefficients and standard errors for ITT effects of bond passage on subsequent 4th grade reading proficiency (Column (1)), 7th grade reading proficiency (Column (2)), subsequent bond passage (Column (3)) and subsequent capital investment (Column (4)). The specification is the ITT regression described in Equation (I.6). Clustered standard errors by school district are shown in parentheses. *, ** and *** indicate the statistical significance at 10%, 5% and 1% levels, respectively.

Table A.4: TOT Effect of Bond Passage on Achievement and Subsequent Expenditure on Construction and Land and Structure

Relative year	4th grade (1)	7th grade (2)	Construction (3)	Land and Structure (4)
1	-0.450	-0.468	596.1 ***	81.58 ***
	(0.480)	*(0.508)*	*(70.59)*	*(17.04)*
2	0.390	-0.375	2796 ***	279.1 ***
	(0.477)	*(0.532)*	*(174.5)*	*(47.18)*
3	0.768	0.389	1897 ***	39.82
	(0.521)	*(0.491)*	*(157.3)*	*(55.37)*
4	0.608	0.996	381.9 ***	-26.65
	(0.467)	*(0.535)*	*(133.6)*	*(38.08)*
5	0.493	-0.283	-179.2	-32.39
	(0.486)	*(0.497)*	*(124.1)*	*(44.37)*
6	0.244	0.515	-167.0	28.68
	(0.491)	*(0.597)*	*(140.4)*	*(26.63)*
7	0.227	0.947 *	-129.0	-0.295
	(0.486)	*(0.567)*	*(114.1)*	*(16.18)*
8	0.774	1.118 **	-275.9 ***	-27.45 **
	(0.501)	*(0.551)*	*(99.12)*	*(13.97)*
9	-0.067	0.537	-67.12	-28.18 *
	(0.511)	*(0.587)*	*(85.42)*	*(15.04)*
10	0.963 *	0.431	-115.5	19.85
	(0.539)	*(0.622)*	*(103.8)*	*(31.45)*
11	1.181 *	0.287	-131.8	85.23
	(0.637)	*(0.750)*	*(108.6)*	*(61.88)*
12	1.368 *	-0.311	-139.0	-18.32
	(0.771)	*(0.790)*	*(188.7)*	*(71.51)*
Sample size	7244	7219	7746	7746

Notes: The table shows the coefficients and standard errors for TOT effects of bond passage on subsequent 4th grade reading proficiency(Column (1)), 7th grade reading proficiency (Column (2)), subsequent expenditure on construction (Column (3)) and land and structure (Column (4)). The specification is the TOT regression described in Equation (I.7). Clustered standard errors by school district are shown in parentheses. *, ** and *** indicate the statistical significance at 10%, 5% and 1% levels, respectively.

Table A.5: ITT and TOT Effects of Bond Passage on 4th Grade Reading Proficiency Using Alternative Specifications

	ITT			TOT		
	Pooled OLS 1	Pooled OLS 2	Fixed Effect	Pooled OLS 1	Pooled OLS 2	Fixed Effect
Relative year	(1)	(2)	(3)	(4)	(5)	(6)
1	2.451 **	1.788 **	-0.868	-0.613	1.819 ***	1.838 ***
	(1.243)	*(0.733)*	*(0.622)*	*(0.997)*	*(0.679)*	*(0.686)*
2	3.618 ***	2.149 ***	-0.644	0.418	2.477 ***	2.407 ***
	(1.165)	*(0.752)*	*(0.682)*	*(1.017)*	*(0.732)*	*(0.708)*
3	3.866 ***	2.144 ***	-0.813	2.078 **	2.939 ***	2.776 ***
	(1.112)	*(0.724)*	*(0.743)*	*(0.986)*	*(0.711)*	*(0.695)*
4	4.005 ***	2.253 ***	-0.753	3.287 ***	3.026 ***	2.816 ***
	(1.145)	*(0.696)*	*(0.765)*	*(1.028)*	*(0.700)*	*(0.679)*
5	3.752 ***	2.257 ***	-0.993	4.291 ***	3.014 ***	2.897 ***
	(1.071)	*(0.678)*	*(0.776)*	*(1.044)*	*(0.704)*	*(0.676)*
6	2.246 **	1.998 ***	-1.376	5.094 ***	2.480 ***	2.326 ***
	(1.028)	*(0.739)*	*(0.841)*	*(1.118)*	*(0.742)*	*(0.733)*
7	1.638 **	1.614 **	-1.580 *	6.531 ***	1.974 ***	1.860 ***
	(0.719)	*(0.654)*	*(0.863)*	*(0.977)*	*(0.693)*	*(0.692)*
8	1.932 ***	1.982 ***	-1.512	10.26 ***	2.561 ***	2.421 ***
	(0.671)	*(0.659)*	*(0.923)*	*(0.962)*	*(0.655)*	*(0.651)*
9	1.071	1.115	-2.282 **	11.27 ***	1.497 **	1.392 **
	(0.716)	*(0.719)*	*(0.965)*	*(1.078)*	*(0.684)*	*(0.700)*
10	1.264 *	1.228 *	-1.631	13.06 ***	1.701 **	1.640 **
	(0.705)	*(0.699)*	*(0.998)*	*(1.196)*	*(0.720)*	*(0.754)*
11	2.123 **	2.055 **	-1.976 *	13.68 ***	2.152 **	1.964 **
	(0.915)	*(0.918)*	*(1.139)*	*(1.371)*	*(0.872)*	*(0.909)*
12	1.716	1.763 *	-1.809	12.07 ***	2.149 **	2.097 **
	(1.056)	*(1.053)*	*(1.239)*	*(1.575)*	*(1.040)*	*(1.051)*
Year fixed effect	No	Yes	Yes	No	Yes	Yes
Bond fixed effect	No	No	Yes	-	-	-
District fixed effect	-	-	-	No	No	Yes
Latent preference	No	No	No	No	No	No
Sample size	9665	9665	9665	7244	7244	7244

Notes: The table shows the ITT and TOT effects of bond passage on reading proficiency. In Columns (1) to (3) the specification is based on the ITT regression described in Equation (I.6). In Columns (4) to (6) the specification is based on the TOT regression described in Equation (I.7). In all columns I do not control for the latent preferences for educational investment. In addition, Columns (1) and (4) do not include year fixed effect, bond fixed effect or district fixed effect. Columns (2) and (4) do not include bond fixed effect or district fixed effect. Clustered standard errors by school district are shown in parentheses. *, ** and *** indicate the statistical significance at 10%, 5% and 1% levels, respectively.

Table A.6: Bandwidth Choice

	AS	BS	CS	DS	ES	FS	GS	HS	IS	JS
CCT	0.292	0.321	0.394	0.545	0.58	0.563	0.416	1.025	0.683	0.756
IK	0.487	0.618	0.416	0.418	0.516	0.497	0.509	1.386	1.438	1.460
CV	0.542	0.504	1.315	2.514	2.602	0.824	2.346	1.568	1.579	1.589

Table A.7: ITT Achievement Estimates of Being Assigned to K-5/6-8 School

	Math		Reading	
	0.5 mile	0.3 mile	0.5 mile	0.3 mile
	Panel A: Estimate of Effects on Test Score Level			
Grade 3	-0.076 ***	-0.075 ***	0.088 ***	0.112 ***
	(0.019)	(0.029)	(0.020)	(0.028)
Grade 4	0.006	0.047 **	0.103 ***	0.108 ***
	(0.015)	(0.022)	(0.016)	(0.023)
Grade 5	0.120 ***	0.154 ***	0.046 ***	0.067 ***
	(0.014)	(0.021)	(0.015)	(0.022)
Grade 6	-0.015	0.022	0.028 *	0.040 *
	(0.016)	(0.024)	(0.015)	(0.022)
Grade 7	-0.013	0.029	0.125 ***	0.144 ***
	(0.015)	(0.022)	(0.017)	(0.025)
Grade 8	-0.020	0.016	0.155 ***	0.171 ***
	(0.021)	(0.031)	(0.020)	(0.030)
	Panel B: Gains based on the Estimated Coefficient in Panel A			
Grade 3 to 4	0.081 ***	0.122 ***	0.015	-0.011
	(0.019)	(0.029)	(0.020)	(0.029)
Grade 4 to 5	0.115 ***	0.106 ***	-0.057 ***	-0.040 *
	(0.012)	(0.018)	(0.015)	(0.021)
Grades 5 to 6	-0.135 ***	-0.132 ***	-0.018	-0.027
	(0.012)	(0.018)	(0.012)	(0.017)
Grades 6 to 7	0.001	0.007	0.097 ***	0.104 ***
	(0.012)	(0.017)	(0.013)	(0.018)
Grades 7 to 8	-0.007	-0.013	0.030 *	0.027
	(0.017)	(0.025)	(0.017)	(0.025)
No. of Students	1413	954	1416	956
Sample Size	82242	40534	82235	40530

Notes: Standard errors clustered by student and pair are reported in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% confidence level respectively. The specification for Panel A is the ITT regression described in equation (III.1). Panel B reports differences between estimated coefficients in panel A. Significance levels are based on tests with the null hypothesis that estimated coefficients for consecutive grades are the same.

Table A.8: TOT Achievement Estimates of Being Assigned to K-5/6-8 School

	Math		Reading	
	0.5 mile	0.3 mile	0.5 mile	0.3 mile
	Panel A: Estimate of Effects on Test Score Level			
Grade 3	-0.074	0.021	0.445 ***	0.507 ***
	(0.055)	(0.082)	(0.055)	(0.081)
Grade 4	0.146 ***	0.264 ***	0.168 ***	0.118 *
	(0.037)	(0.055)	(0.040)	(0.062)
Grade 5	0.360 ***	0.433 ***	0.111 ***	0.165 ***
	(0.038)	(0.054)	(0.039)	(0.054)
Grade 6	-0.004	0.094	0.067 *	0.107 *
	(0.042)	(0.060)	(0.040)	(0.056)
Grade 7	-0.007	0.105	0.351 ***	0.397 ***
	(0.040)	(0.056)	(0.045)	(0.063)
Grade 8	0.052	0.098	0.418 ***	0.447 ***
	(0.053)	(0.074)	(0.053)	(0.073)
	Panel B: Gains based on the Estimated Coefficient in Panel A			
Grade 3 to 4	0.219 ***	0.243 ***	-0.277 ***	-0.389 ***
	(0.052)	(0.080)	(0.048)	(0.069)
Grade 4 to 5	0.215 ***	0.169 ***	-0.058	0.046
	(0.031)	(0.046)	(0.035)	(0.052)
Grades 5 to 6	-0.364 ***	-0.339 ***	-0.044	-0.057
	(0.033)	(0.046)	(0.031)	(0.041)
Grades 6 to 7	-0.002	0.012	0.284 ***	0.289 ***
	(0.031)	(0.044)	(0.035)	(0.047)
Grades 7 to 8	0.059	-0.007	0.067	0.050
	(0.044)	(0.059)	(0.044)	(0.060)
No. of Students	1370	920	1373	922
Sample Size	80225	39672	80216	39666

Notes: Standard errors clustered by student and pair are reported in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% confidence level respectively. The specification for Panel A is the TOT regression described in equation III.2. Panel B reports differences between estimated coefficients in panel A. Significance levels are based on tests with the null hypothesis that estimated coefficients for consecutive grades are the same.

Table A.9: Impact on Achievement of Being Actually Enrolled in K-5/6-8 School, Paired by Prior Actual Enrollment

	Math		Reading	
	0.5 mile	0.3 mile	0.5 mile	0.3 mile
	Panel A: Estimate of Effects on Test Score Level			
Grade 3	-0.096	-0.129	0.510 ***	0.545 ***
	(0.076)	(0.123)	(0.081)	(0.129)
Grade 4	-0.015	0.009	0.382 ***	0.341 ***
	(0.053)	(0.075)	(0.057)	(0.082)
Grade 5	0.228 ***	0.272 ***	0.414 ***	0.401 ***
	(0.059)	(0.077)	(0.061)	(0.080)
Grade 6	-0.116 *	-0.017	0.387 ***	0.286 ***
	(0.060)	(0.075)	(0.064)	(0.079)
Grade 7	-0.323 ***	-0.177 **	0.384 ***	0.297 ***
	(0.060)	(0.074)	(0.067)	(0.086)
Grade 8	-0.241 ***	-0.025	0.945 ***	0.588 ***
	(0.077)	(0.090)	(0.096)	(0.104)
	Panel B: Gains based on the Estimated Coefficient in Panel A			
Grade 3 to 4	0.081	0.138	-0.128 *	-0.205 *
	(0.073)	(0.120)	(0.072)	(0.116)
Grade 4 to 5	0.243 ***	0.262 ***	0.031	0.060
	(0.048)	(0.064)	(0.050)	(0.068)
Grades 5 to 6	-0.344 ***	-0.289 ***	-0.027	-0.115 *
	(0.050)	(0.060)	(0.052)	(0.062)
Grades 6 to 7	-0.207 ***	-0.160 ***	-0.003	0.011
	(0.047)	(0.056)	(0.052)	(0.061)
Grades 7 to 8	0.082	0.152 **	0.561 ***	0.291 ***
	(0.063)	(0.073)	(0.079)	(0.082)
No. of Students	1620	912	1622	915
Sample Size	64322	23911	64312	23919

Notes: Standard errors clustered by student and pair are reported in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% confidence level respectively. The specification for Panel A is the TOT regression described in equation III.2. Panel B reports differences between estimated coefficients in panel A. Significance levels are based on tests with the null hypothesis that estimated coefficients for consecutive grades are the same.

Table A.10: Impact on Achievement of Being Assigned to K-5/6-8 School Conditional on Transition

	Math				Reading			
	0.5 mile		0.3 mile		0.5 mile		0.3 mile	
	Panel A: Estimate of Effects on Test Score Level							
Grade 3	-0.078		0.016		0.446	***	0.527	***
	(0.055)		(0.081)		(0.055)		(0.081)	
Grade 4	0.135	***	0.242	***	0.175	***	0.135	**
	(0.036)		(0.055)		(0.040)		(0.061)	
Grade 5	0.335	***	0.392	***	0.090	**	0.156	***
	(0.037)		(0.052)		(0.039)		(0.054)	
Grade 6	-0.006		0.101	*	0.066	*	0.122	**
	(0.042)		(0.061)		(0.040)		(0.056)	
Grade 7	-0.023		0.083		0.334	***	0.380	***
	(0.040)		(0.055)		(0.045)		(0.063)	
Grade 8	0.053		0.084		0.433	***	0.454	***
	(0.054)		(0.074)		(0.054)		(0.074)	
	Panel B: Gains based on the Estimated Coefficient in Panel A							
Grade 3 to 4	0.212	***	0.226	***	-0.271	***	-0.392	***
	(0.052)		(0.080)		(0.048)		(0.069)	
Grade 4 to 5	0.200	***	0.151	***	-0.085	**	0.021	
	(0.031)		(0.045)		(0.035)		(0.051)	
Grades 5 to 6	-0.341	***	-0.292	***	-0.024		-0.034	
	(0.033)		(0.045)		(0.030)		(0.040)	
Grades 6 to 7	-0.017		-0.018		0.268	***	0.257	***
	(0.031)		(0.044)		(0.035)		(0.046)	
Grades 7 to 8	0.076	*	0.001		0.100	**	0.074	
	(0.044)		(0.059)		(0.045)		(0.061)	
No. of Students	1370		920		1373		922	
Sample Size	80225		39672		80216		39666	

Notes: Standard errors clustered by student and pair are reported in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% confidence level respectively. The specification for Panel A is the TOT regression described in equation III.2. Panel B reports differences between estimated coefficients in panel A. Significance levels are based on tests with the null hypothesis that estimated coefficients for consecutive grades are the same.

Table A.11: Impact on Achievement of Moving to Middle School

	Math				Reading			
	0.5 mile		0.3 mile		0.5 mile		0.3 mile	
	Panel A: Estimate of Effects on Test Score Level							
Grade 3	-1.009	***	-0.801	**	-0.226		-0.131	
	(0.212)		(0.347)		(0.175)		(0.287)	
Grade 4	0.142		0.432	**	0.084		-0.104	
	(0.114)		(0.179)		(0.123)		(0.199)	
Grade 5	0.817	***	1.051	***	0.433	***	0.520	***
	(0.118)		(0.182)		(0.117)		(0.172)	
Grade 6	0.041		0.143		-0.065		-0.099	
	(0.115)		(0.178)		(0.106)		(0.157)	
Grade 7	-0.131		0.110		1.008	***	1.302	***
	(0.123)		(0.185)		(0.127)		(0.199)	
Grade 8	0.436	***	0.312		0.901	***	1.023	***
	(0.146)		(0.210)		(0.124)		(0.181)	
	Panel B: Gains based on the Estimated Coefficient in Panel A							
Grade 3 to 4	1.151	***	1.234	***	0.310	*	0.026	
	(0.206)		(0.354)		(0.163)		(0.270)	
Grade 4 to 5	0.675	***	0.619	***	0.349	***	0.624	***
	(0.099)		(0.144)		(0.106)		(0.173)	
Grades 5 to 6	-0.776	***	-0.908	***	-0.499	***	-0.619	***
	(0.098)		(0.150)		(0.089)		(0.133)	
Grades 6 to 7	-0.172	**	-0.033		1.073	***	1.401	***
	(0.084)		(0.131)		(0.109)		(0.174)	
Grades 7 to 8	0.568	***	0.203		-0.107		-0.279	*
	(0.110)		(0.147)		(0.102)		(0.147)	
No. of Students	865		607		865		607	
Sample Size	46243		23422		46232		23417	

Notes: Standard errors clustered by student and pair are reported in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% confidence level respectively. The specification for Panel A is the TOT regression described in equation III.2. Panel B reports differences between estimated coefficients in panel A. Significance levels are based on tests with the null hypothesis that estimated coefficients for consecutive grades are the same.

Table A.12: Impact of Being Assigned in a K-5 School on Achievement, Nonmoving K-8 Student

	Math		Reading	
	0.5 mile	0.3 mile	0.5 mile	0.3 mile
	Panel A: Estimate of Effects on Test Score Level			
Grade 3	-0.148 ***	-0.340 ***	0.227 ***	0.049
	(0.052)	(0.087)	(0.060)	(0.086)
Grade 4	0.239 ***	0.246 ***	0.143 ***	0.151 **
	(0.048)	(0.079)	(0.042)	(0.064)
Grade 5	0.228 ***	0.096	0.157 ***	0.029
	(0.047)	(0.074)	(0.042)	(0.066)
Grade 6	-0.061	-0.157 **	0.242 ***	0.197 ***
	(0.048)	(0.078)	(0.047)	(0.072)
Grade 7	-0.159 ***	0.274 ***	0.227 ***	0.186 **
	(0.049)	(0.078)	(0.053)	(0.082)
Grade 8	-0.092 *	-0.201 **	0.888 ***	0.799 ***
	(0.050)	(0.079)	(0.044)	(0.062)
	Panel B: Gains based on the Estimated Coefficient in Panel A			
Grade 3 to 4	0.387 ***	0.586 ***	-0.083	0.103
	(0.064)	(0.105)	(0.063)	(0.096)
Grade 4 to 5	-0.011	-0.150 **	0.014	-0.122 *
	(0.041)	(0.066)	(0.039)	(0.065)
Grades 5 to 6	-0.289 ***	-0.253 ***	0.085 ***	0.168 ***
	(0.031)	(0.045)	(0.029)	(0.045)
Grades 6 to 7	-0.098 ***	-0.118 ***	-0.015	-0.010
	(0.026)	(0.040)	(0.031)	(0.047)
Grades 7 to 8	0.066 *	0.073	0.661 ***	0.613 ***
	(0.039)	(0.058)	(0.041)	(0.059)
No. of Students	491	249	495	252
Sample Size	10514	4605	10516	4607

Notes: Standard errors clustered by student and pair are reported in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% confidence level respectively. The specification for Panel A is the ITT regression described in equation (III.1). Panel B reports differences between estimated coefficients in panel A. Significance levels are based on tests with the null hypothesis that estimated coefficients for consecutive grades are the same.

Table A.13: Decomposition of the Effects of Education and Socioemotional Skills on the Hazard of Death with Respect to Mediators

	Males					Females				
	Estimate	Standard error	*p*-value	Life expectancy[(a)]	Longevity value, thousands of USD[(b)]	Estimate	Standard error	*p*-value	Life expectancy[(a)]	Longevity value, thousands of USD[(b)]
Education										
Aggregated components and health stock										
Total behaviors and job	-0.079 **	0.040	0.047	0.661	61.6	-0.014	0.039	0.710	0.084	7.2
Health stock	-0.113 *	0.060	0.061	0.935	86.3	-0.035	0.047	0.457	0.262	21.8
Total explained	-0.192 **	0.080	0.017	1.553	142.9	-0.049	0.070	0.482	0.401	32.9
Total[(c)]	-0.391 ***	0.130	0.002	4.685	384.6	-0.207 *	0.159	0.096	1.828	143.9
Specific components										
Smoking	-0.013 *	0.008	0.085	0.108	10.0	-0.007	0.007	0.268	0.055	4.5
Risky drinking	-0.010	0.007	0.122	0.082	7.6	0.000	0.001	0.959	-0.029	-2.2
Physical exercise	-0.005	0.006	0.432	0.044	4.2	-0.005	0.010	0.624	0.012	1.3
Overweight	-0.006	0.022	0.800	0.064	6.3	-0.003	0.007	0.654	0.005	0.7
Married	0.003	0.008	0.716	-0.025	-2.3	0.028 ***	0.010	0.006	-0.282	-23.3
Social activity	-0.002	0.009	0.827	0.015	1.4	0.008	0.025	0.754	-0.075	-6.0
Income per capita	-0.028	0.020	0.170	0.241	22.4	-0.035 *	0.019	0.068	0.252	21.0
Dangerous job	-0.019	0.012	0.124	0.148	13.8	0.000	0.001	0.888	-0.027	-2.0
Socioemotional Skills										
Aggregated components and health stock										
Total behaviors and job	-0.026 *	0.014	0.061	0.137	12.7	-0.038 **	0.017	0.025	0.900	74.5
Health stock	-0.011	0.013	0.399	0.198	18.2	-0.033	0.029	0.261	0.543	45.0
Total explained	-0.037 *	0.021	0.080	0.331	30.5	-0.071 *	0.042	0.094	1.473	119.7
Total	-0.069	0.054	0.202	1.429	117.4	-0.148 **	0.062	0.018	3.113	238.1
Specific components										
Smoking	-0.009 **	0.005	0.036	0.067	6.2	-0.010 *	0.005	0.052	0.075	6.2
Risky drinking	-0.003	0.002	0.146	0.027	2.5	0.000	0.001	0.958	-0.027	-2.0
Physical exercise	-0.003	0.003	0.326	0.031	2.9	-0.003	0.005	0.604	0.035	3.3
Overweight	-0.002	0.010	0.853	0.033	4.5	-0.002	0.004	0.637	0.068	6.2
Married	0.000	0.004	0.979	-0.082	-7.7	-0.009 **	0.004	0.021	0.338	28.5
Social activity	-0.001	0.003	0.853	0.005	0.5	0.001	0.003	0.761	-0.026	-1.9
Income per capita	-0.006	0.005	0.213	0.039	3.6	-0.015 *	0.008	0.070	0.306	25.7
Dangerous job	-0.003	0.002	0.217	0.029	2.6	0.000	0.001	0.897	-0.023	-1.7
Sample size			3961					4491		

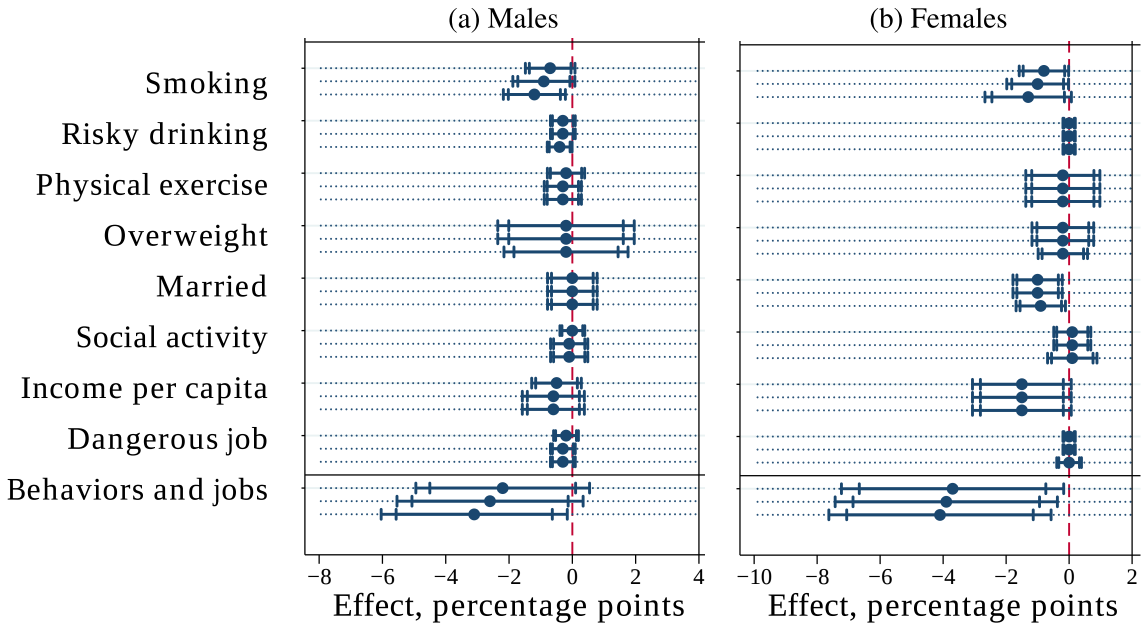
Notes: Huber-White standard errors are shown. Asterisks denote the level of statistical significance: ***, **, and * denote $p < 0.01$, 0.05, and 0.10 for the single hypothesis tests. Decompositions of the effect of socioemotional skills correspond to one standard deviation change in skills. [(a)]Effect on the life expectancy at age 53, years. [(b)]Effect on the value of remaining life at age 53 (thousands of 2012 US dollars). [(c)]One-sided test for this total effect is motivated by abundant evidence from the literature that the total effect of education on longevity is nonnegative (Grossman, 2006; Grossman and Kaestner, 1997; Lochner, 2011).

Table A.14: Difference between Samples of 1957 and 1992

Variable	Males 1957 sample	Males 1992 sample	Males Difference	Males Std. error	Females 1957 sample	Females 1992 sample	Females Difference	Females Std. error
IQ	101.0	101.5	0.503	0.330	100.6	101.0	0.395	0.296
Academic achievement at high school								
Standardized academic achievement	97.3	97.9	0.550	0.328	104.0	104.5	0.500	0.303
Member of an honor society	0.041	0.044	0.002	0.004	0.060	0.063	0.003	0.005
Outstanding student	0.096	0.100	0.004	0.006	0.114	0.118	0.004	0.007
Background variables								
Father is a farmer or a farm manager	0.197	0.200	0.003	0.009	0.194	0.198	0.004	0.009
Father is a white collar employee	0.301	0.301	0.000	0.011	0.299	0.299	-0.000	0.010
Father has attended college	0.160	0.160	0.001	0.008	0.144	0.145	0.000	0.007
Mother has attended college	0.143	0.144	0.001	0.008	0.148	0.150	0.002	0.008
Parental income (log)	8.536	8.536	0.000	0.015	8.524	8.532	0.009	0.014
Attended high school in a rural area	0.184	0.187	0.002	0.008	0.182	0.187	0.005	0.008
Resided in a metropolitan area	0.342	0.337	-0.005	0.010	0.371	0.366	-0.005	0.010
Respondent's abnormal weight	0.284	0.284	0.000	0.011	0.314	0.313	-0.002	0.010
Respondent's childhood household had a smoker	0.759	0.757	-0.002	0.011	0.735	0.734	-0.000	0.010
Birth year 1937–38	0.218	0.216	-0.002	0.009	0.147	0.144	-0.003	0.007
Birth year 1939	0.745	0.746	0.001	0.009	0.802	0.804	0.002	0.008
Birth year 1940	0.037	0.038	0.001	0.004	0.051	0.053	0.001	0.005
Sample size	4556	3961			5021	4491		

Notes: Lack of asterisks represents no statistically significant changes even at the 10% level.

Figure A.1: Decomposition of the Effect of Socioemotional Skills at Various levels of IQ



Notes: The upper bar is evaluated at the level of IQ to be one standard deviation down from the mean. The middle bar is the main result with the IQ at its average level. The lower bar is for IQ to be one standard deviation above the mean. Inner and outer vertical bars represent the 90% and 95% Huber-White confidence intervals calculated using the delta method. Calculations are based on the WLS.

Table A.15: Effects of Socioemotional Skills on the Hazard of Death for Females: Correction of the Total Through Correcting its Component

	WLS cohort, -6.7 PP			Simulated cohort, 4.5 PP			Simulated cohort, 0 PP		
	Estimate	standard error	p -value[(a)]	Estimate	standard error	p -value[(a)]	Estimate	standard error	p -value[(a)]
Socioemotional skills									
Married	-0.009 **	0.004	0.021	-0.013 ***	0.005	0.006	-0.011 **	0.004	0.011
Total	-0.148 **	0.062	0.018	-0.151 **	0.062	0.015	-0.150 **	0.062	0.016

Notes: Huber-White standard errors are shown. [(a)]Asterisks denote the level of statistical significance: ***, **, and * represent $p < 0.01$, 0.05 and 0.10. Sample size is 4491.

Table A.16: Comparisons of the Main Model 1 with Biased Models 2–5, Males

	Model 1			Model 2			Model 3			Model 4			Model 5		
	Estimate	std. error	*p*-value	Estimate	std. error	*p*-value	Estimate	std. error	*p*-value	Estimate	std. error	*p*-value	Estimate	std. error	*p*-value
Aggregated components and health stock															
Total behaviors and job	-0.079 **	0.040	0.047	-0.090 ***	0.026	0.000	-0.087 **	0.034	0.011	-0.105 ***	0.038	0.006	-0.113 ***	0.033	0.001
Health stock	-0.113 *	0.060	0.061	-0.206 ***	0.046	0.000	-0.224 ***	0.047	0.000	-0.237 ***	0.048	0.000	-0.261 ***	0.046	0.000
Total explained	-0.192 **	0.080	0.017	-0.204 *	0.119	0.085	-0.184 *	0.110	0.094	-0.155	0.105	0.138	-0.115	0.099	0.248
Total[c]	-0.391 ***	0.130	0.002	-0.410 ***	0.117	0.000	-0.408 ***	0.109	0.000	-0.392 ***	0.103	0.000	-0.375 ***	0.093	0.000
Specific components															
Smoking	-0.013 *	0.008	0.085	-0.014 *	0.008	0.069	-0.004	0.005	0.348	-0.005	0.005	0.334	-0.012 *	0.006	0.058
Risky drinking	-0.010	0.007	0.122	-0.010	0.006	0.124	-0.006	0.006	0.316	-0.006	0.006	0.313	-0.009	0.006	0.138
Physical exercise	-0.005	0.006	0.432	-0.006	0.005	0.224	-0.011	0.010	0.236	-0.015	0.013	0.220	-0.012	0.009	0.206
Overweight	-0.006	0.022	0.800	-0.011	0.007	0.135	-0.008	0.007	0.237	-0.008	0.007	0.229	-0.015 *	0.009	0.094
Married	0.003	0.008	0.716	0.003	0.008	0.715	0.007	0.011	0.553	0.001	0.010	0.892	-0.001	0.006	0.915
Social activity	-0.002	0.009	0.827	-0.003	0.008	0.710	-0.006	0.019	0.738	-0.007	0.021	0.755	-0.003	0.010	0.752
Income per capita	-0.028	0.020	0.170	-0.029 *	0.015	0.056	-0.028 *	0.016	0.068	-0.034 *	0.020	0.082	-0.040	0.025	0.106
Dangerous job	-0.019	0.012	0.124	-0.020	0.012	0.103	-0.029	0.018	0.107	-0.032	0.020	0.115	-0.022	0.015	0.146
Unobserved heterogeneity		Yes			No			No			No			No	
Socioemotional skills		Yes			Yes			No			No			No	
Cognitive skills		Yes			Yes			Yes			No			No	
Background variables		Yes			Yes			Yes			Yes			No	
Sample size		3961			3961			3961			3961			3961	

Notes: The table compares our main model that controls for background variables, cognitive skills, socioemotional skills, and unobserved heterogeneity, with models that lack one or several of these controls, as specified in the bottom of the table.

Table A.17: Comparisons of the Main Model 1 with Biased Models 2–5, Females

	Model 1			Model 2			Model 3			Model 4			Model 5		
	Estimate	std. error	*p*-value	Estimate	std. error	*p*-value	Estimate	std. error	*p*-value	Estimate	std. error	*p*-value	Estimate	std. error	*p*-value
Aggregated components and health stock															
Total behaviors and job	-0.014	0.039	0.710	-0.025	0.032	0.440	-0.056 *	0.034	0.097	-0.077 **	0.036	0.034	-0.124 ***	0.042	0.003
Health stock	-0.035	0.047	0.457	-0.078 *	0.043	0.068	-0.137 ***	0.044	0.002	-0.168 ***	0.046	0.000	-0.242 ***	0.050	0.000
Total explained	-0.049	0.070	0.482	-0.151	0.150	0.314	-0.149	0.146	0.308	-0.159	0.142	0.265	-0.178	0.137	0.193
Total[c]	-0.207 *	0.159	0.096	-0.229 *	0.150	0.064	-0.286 **	0.145	0.024	-0.327 ***	0.140	0.010	-0.421 ***	0.133	0.001
Specific components															
Smoking	-0.007	0.007	0.268	-0.008	0.007	0.241	-0.011	0.008	0.139	-0.013	0.008	0.104	-0.015 *	0.009	0.073
Risky drinking	0.000	0.001	0.959	0.000	0.002	0.949	0.000	0.002	0.987	0.000	0.002	0.987	0.000	0.001	0.958
Physical exercise	-0.005	0.010	0.624	-0.009	0.009	0.319	-0.012	0.011	0.282	-0.012	0.011	0.277	-0.018	0.014	0.210
Overweight	-0.003	0.007	0.654	-0.007	0.006	0.250	-0.009	0.007	0.196	-0.008	0.006	0.215	-0.015	0.010	0.117
Married	0.028 ***	0.010	0.006	0.028 ***	0.010	0.005	0.019 **	0.009	0.023	0.022 ***	0.009	0.010	0.023 ***	0.008	0.005
Social activity	0.008	0.025	0.754	0.003	0.024	0.915	0.002	0.025	0.922	0.002	0.027	0.937	-0.002	0.030	0.953
Income per capita	-0.035 *	0.019	0.068	-0.031 ***	0.011	0.005	-0.044 ***	0.013	0.001	-0.067 ***	0.018	0.000	-0.094 ***	0.023	0.000
Dangerous job	0.000	0.001	0.888	0.000	0.002	0.757	-0.001	0.002	0.704	-0.001	0.003	0.686	-0.002	0.006	0.702
Unobserved heterogeneity	Yes			No			No			No			No		
Socioemotional skills	Yes			Yes			No			No			No		
Cognitive skills	Yes			Yes			Yes			No			No		
Background variables	Yes			Yes			Yes			Yes			No		
Sample size	4491			4491			4491			4491			4491		

Notes: The table compares our main model that controls for background variables, cognitive skills, socioemotional skills, and unobserved heterogeneity, with models that lack one or several of these controls, as specified in the bottom of the table.

Table A.18: Average Bias Induced by Omitting Essential Controls

	Comparison 1	Comparison 2	Comparison 3	Comparison 4
OLS and Logit, marginal effects	5%	40%	57%	48%
COX	15%	17%	17%	19%
Total explained effects	6%	4%	19%	40%
Total effects	8%	21%	29%	54%
Unobserved heterogeneity	no	no	no	no
Socioemotional Skills	yes	no	no	no
IQ	yes	yes	no	no
Background controlls	yes	yes	yes	no

Notes: The table compares our main model that controls for background variables, cognitive skills, socioemotional skills, and unobserved heterogeneity, with models that lack one or several of these controls, as specified in the bottom of the table. For each statistically significant estimate, the bias is calculated in % relative to the main model counterpart. (Statistically insignificant results are excluded as leading to less reliable estimates of the bias.) Then absolute values of the these biases are averaged to form a measure of an average bias to either direction.

Table A.19: Estimates of the Measurement System

	Males		Females	
Socioemotional skills				
Standardized academic achievement	0.870	***	0.887	***
	(0.034)		*(0.028)*	
Member of an honor society	0.733	***	0.716	***
	(0.033)		*(0.025)*	
Outstanding student	0.714	***	0.757	***
	(0.029)		*(0.024)*	
Health stock				
General health	0.477	***	0.544	***
	(0.024)		*(0.025)*	
Major illness	-0.593	***	-0.667	***
	(0.027)		*(0.027)*	
Stayed in bed at least once last year	-0.600	***	-0.497	***
	(0.029)		*(0.028)*	
Hospitalization at least once last year	-0.832	***	-0.632	***
	(0.033)		*(0.042)*	
RMSEA[(a)]	0.071		0.061	
CFI[(b)]	0.918		0.945	
TLI[(c)]	0.868		0.911	
Chi-square test[(d)]	275.642		228.595	
Degrees of freedom	13		13	
p -value	0.000		0.000	
Sample size	3961		4491	

Notes: Model estimated based on the WLS data.

Table A.20: Robustness of the Total Effect to Omission of Individual Behaviors in The Model

	College				Socioemotional skills			
	male		female		male		female	
Omitted mediators								
No omission	-0.391	***	-0.207		-0.069		-0.148	**
	(0.130)		*(0.159)*		*(0.054)*		*(0.062)*	
Smoking	-0.408	***	-0.212		-0.081		-0.161	***
	(0.135)		*(0.158)*		*(0.056)*		*(0.061)*	
Risky drinking	-0.400	**	-0.213		-0.071		-0.151	**
	(0.170)		*(0.160)*		*(0.084)*		*(0.063)*	
Exercise	-0.390	***	-0.189		-0.074		-0.147	
	(0.127)		*(1.258)*		*(0.050)*		*(2.267)*	
Overweight	-0.374	***	-0.223		-0.070		-0.154	**
	(0.131)		*(0.170)*		*(0.054)*		*(0.061)*	
Marriage	-0.388		-0.205		-0.068		-0.142	**
	(0.759)		*(0.160)*		*(0.088)*		*(0.063)*	
Social activity	-0.410	***	-0.206		-0.081		-0.152	**
	(0.122)		*(0.160)*		*(0.053)*		*(0.061)*	
Income	-0.365		-0.253		-0.068		-0.092	
	(1.325)		*(15.00)*		*(1.803)*		*(38.00)*	
Dangerous job	-0.401		-0.200		-0.071		-0.144	**
	(0.291)		*(0.163)*		*(0.054)*		*(0.064)*	
Average deviation [(a)]	0.26%		2.72%		5.80%		-3.46%	
Average deviation excluding income[(b)]	1.24%		-0.07%		6.83%		1.45%	

Table A.21: Log Likelihood and Information Criteria by the Number of Points of Support in Finite Mixture Model

Number of classes	No classes	2	3	4	5
Males					
AIC	90046	89696	89490	89339	89295
BIC	92830	92580	92475	92425	92481
Change in AIC		-350	-206	-151	-44
Change in BIC		-249	-105	-50	57
Probabilities of latent classes					
P_1	-	0.983	0.935	0.773	0.514
P_2	-	0.017	0.050	0.191	0.447
P_3	-	-	0.015	0.022	0.023
P_4	-	-	-	0.014	0.014
P_5	-	-	-	-	0.002
Females					
AIC	98588	98084	97671	97446	97568
BIC	101427	101026	100716	100593	100818
Change in AIC		-504	-413	-225	123
Change in BIC		-401	-310	-123	225
Probabilities of latent classes					
P_1	-	0.954	0.723	0.556	0.510
P_2	-	0.046	0.238	0.387	0.308
P_3	-	-	0.039	0.038	0.143
P_4	-	-	-	0.019	0.037
P_5	-	-	-	-	0.003

B Elite and Non-Elite Schools

Elite high schools in Beijing are officially called model schools. In 2008 there were 68 model schools in Beijing. This kind of high school originated from the key schools policy started in the 1950s, when the government allocated more resources to certain so-called key schools in hopes of getting higher education quality and better outcomes. Model schools replaced key schools in the 1990s. Unlike key schools, model schools focus on multiple outcomes rather than exam scores. The admission to college largely depends on exam scores, and the pedagogy of model schools is restricted to a large extent. All schools other than model schools are classified as regular schools. The differences between the two types of schools can be summarized as follows. First, model schools are favorites of public education finance; they receive more government financial support. Moreover, they can collect more money from external funding because of their excellent reputations. With more funding, model schools can afford better learning conditions, higher-quality teachers and so on. Model schools have much appeal for middle school students, and thus many excellent students choose them. In the anonymous district studied in this paper there were 2 model schools in 2008, named by AS and BS schools.

AS High School was founded in 1956 and appointed as the only key school of the district in 1978. In 2002 it was selected into the first batch of model schools of Beijing City. AS High School is equipped with modern educational facilities, such as a standard stadium and a library with hundreds of thousands of books. More than 20% percent of the teachers have at least a master's degree. Therefore it is not surprising that AS High School has great outcomes; on average 95% of students are admitted to colleges, and 30% of them go to key universities. BS High School was a normal school after its establishment in 1949 and changed to a high school in 1997. It was selected as a model school of Beijing City (the second model school in the district) in 2005. Although it operated as a high school for only 10 years, BS High School also did well on the outcomes of interest. In recent years 90% of the students at BS High School were able to enter college after graduation.

C Exam, Admission and Track Rules

Students in middle schools can participate in the entrance exam if they satisfy certain criteria, which we do not discuss in detail because they are irrelevant to our study. The entrance exam to high schools is held once per year, in late June. The score serves as the only criterion for enrollment in high schools in most cases with two exceptions. One is that students with excellent awards, such as "Jin Fan" and "Yin Fan" awards, are able to gain admission without taking the exams. The other is that students who satisfy several conditions, such as minority race or being children of a martyr, can obtain additional scores. The first case does not matter as we drop observations whose SEEH is missing. The second case may lead to more or less mismatching between the score and ability, but since we do not care about conditions before high school, such potential mismatching should not undermine the results of this study. The exam consists of six sub-exams covering Chinese, Mathematics, English, Physics, Chemistry and Physical Education. The full scores are 120, 120, 120, 100, 80 and 30 respectively, which makes the total SEEH 570.

In a typical admission problem under the Boston mechanism there are a number of students who are willing to be assigned one seat at a high school. Each student reports a strict preference ordering over schools before the entrance exam. After the exam each school has a strict preference ranking of all students by exam score. The outcome of the mechanism is determined in several rounds as follows.

Round 1. Schools only consider the students who listed them as the first choice. Each school admits those students by SEEH until there is no student left who listed it as the first choice or its capacity is filled.

Round 2. Schools with available seats consider the unassigned students who listed them as the second choice. Each school admits those students by the order of SEEH until there is no student left who listed it as the second choice or its capacity is filled.

Round k. Schools whose capacity is not filled admit unassigned students who listed them as the k-th choice by SEEH.

Each student in the district can list at most eight schools, and it is possible that after round eight some schools still have unfilled seats. These schools can contact any unassigned students to see whether they would like to attend this school. After all schools are filled, the students who are still unassigned have to try other options, such as attending private or professional schools or leaving school for the labor market. The cutoff score for admission for each school is the exam score of the last student who filled the capacity.

The entrance exam of colleges in Beijing is also held in each June. The subjects covered in the exam are different between two tracks. The exam for the art track includes Chinese, Mathematics for Art, Foreign Language and Integrated Art, while the exam for the science track includes Chinese, Mathematics for Science, Foreign Language and Integrated Science. Integrated Art combines History, Politics and Geography while Integrated Science combines Physics, Chemistry and Biology. The full scores of Chinese, Mathematics for Art or Science and Foreign Language are all 150. Full scores of the two integrated subjects are both 300. Thus the full score of the exam for both tracks is 750. Students satisfying certain conditions, such as minority race, can obtain additional scores in the admission phase. However this does not affect the SEEC, so we do not focus on these extra conditions. The score serves as the only criterion for enrollment in colleges in most cases. There is also a complicated admission process for enrollment in colleges. However, this process has little to do with the effect of Model Schools, so we do not examine its impact on our outcomes of interest.

D Tabulated Decompositions, Life Expectancy, and the Value of Remaining Life

Table A.13 documents the decompositions of the effects of education and socioemotional skills on longevity with respect to mediators. We evaluate the decompositions not only in terms of the hazard of death but also in terms of the expected years of additional life at age 53 and the corresponding value of remaining life. We use the following formulas for this evaluation. The life expectancy at the start of the risk period t_1 is

$$e = \int_{t_1}^{\infty} S(t)dt, \tag{D.1}$$

where $S(t)$ is the survival function conditional on survival to time t_1 calculated as

$$S(t) = \begin{cases} \sum_c p_c S_{01}(t)^{\exp(g_{1c})}, & \text{for } t_1 \leq t < t_2 \\ \sum_c p_c S_{02}(t)^{\exp(g_{2c})}, & \text{for } t_2 \leq t < \infty. \end{cases} \tag{D.2}$$

Here we use the following notation: p_c is the probability of latent class c; $S_{0j}(t)$ is the baseline survival function for the period $[t_j, t_{j+1})$. We extrapolate function $S_{02}(t)$ from age 75 to 100 using the survival function for this cohort of white men and women from the US Census Bureau (Arias, 2012). Finally,

$$g_{jc} = e_{1j}\Theta + \sum_q e_{2q}B_{0q} + e_{3j}D + \sum_k e_{4k}B_{1k} + e_5\Theta_1^H + \tau_{\lambda jc} + BG, \tag{D.3}$$

where $\tau_{\lambda jc}$ is the intercept for class $c = 1, \ldots, 4$ and time period $[t_j, t_{j+1})$, $j = 1, 2$, and BG represents terms that account for contributions of background variables, which are part of the model but are not explicitly written to simplify notation.

Following Murphy and Topel (2006), we calculate the value of remaining life V_R at age t_1 as

$$V_R = \int_{t_1}^{\infty} S(t)v(t)e^{-r(t-t_1)}dt, \tag{D.4}$$

where $v(t)$ is the value of a life-year, and r is the discount factor. We use $v(t)$ from Murphy and Topel (2006), adjusted to the updated value of life of 9.1 mln USD adopted by the US Department of Transportation, which is in line with recent economic research (Viscusi, 2013). We also use a discount factor of 0.035 as in Murphy and Topel (2006).

We calculate our decompositions in terms of life expectancy and the value of remaining life by calculating counterfactual changes in function $S(t)$ (see equations (D.2) and (D.3)) induced by exogenous changes in education D and skills Θ.[33] Counterfactual $\Delta S(t)$ generates counterfactuals Δe and ΔV_R using formulas (D.1) and (D.4).[34] Note that the calculation of ΔV_R only accounts for induced changes in the survival function $S(t)$, but not in the quality of life $v(t)$. We leave estimation of $\Delta v(t)$ to future research based on data better fitted to address this question. Our current estimates of ΔV_R should be viewed as evaluations of the additional longevity. Total effects on the value of life are likely even larger since skills and education tend to increase not only longevity but also health.

E A More General Model Specification

Before adopting our main (parsimonious) model (IV.1–IV.6) described in detail in the main text, we consider the more general model presented here. Our analysis shows that many degrees of freedom of the more general model are redundant.

Following the same notation as in the main text, we can write the general model as

[33] We evaluate the simulation at average values of background variables and skills.

[34] Note that unlike the decomposition for $\frac{\Delta\lambda}{\lambda}$ based on calculus, in which elements sum up to totals by construction (see columns named "estimate" in Table A.13), elements of decompositions for Δe and ΔV_R (see columns named "life expectancy" and "longevity value") are not required to exactly sum up to "totals" because of nonlinearity, but we find the estimated totals to be remarkably close to the sum of their estimated components.

follows:

$$\Theta = \tau_\Theta + \varepsilon_\Theta \tag{E.5}$$

$$B_{0q}^* = a_{1q}\Theta + a_{2q}\Theta \cdot IQ + \tau_{B0q} + \varepsilon_{B0q},\ q = 1,...,Q \tag{E.6}$$

$$D^* = b_1\Theta + b_2\Theta \cdot IQ + \sum_q b_{3q}B_{0q} + \tau_D + \varepsilon_D \tag{E.7}$$

$$B_{1kj}^* = c_{1kj}\Theta + \sum_q c_{2kqj}B_{0q} + c_{3kj}D + c_{4kj}\Theta \cdot IQ + c_{5kj}\Theta D + c_{6kj}IQ \cdot D + \tau_{B1kj} + \varepsilon_{B1kj},$$
$$k = 1,...,K,\ j = 0,...,J \tag{E.8}$$

$$\Theta_{1j}^H = d_{1j}\Theta + \sum_q d_{2qj}B_{0q} + d_{3j}D + d_{4j}\Theta \cdot IQ + d_{5j}\Theta D + d_{6j}IQ \cdot D + \tau_{H1j} + \varepsilon_{H1j},$$
$$j = 0,...,J \tag{E.9}$$

$$\ln(\lambda(t)) = e_{1j}\Theta + \sum_q e_{2qj}B_{0q} + e_{3j}D + \sum_k (e_{4kj}B_{1kj} + e_{5kj}B_{1kj}D + e_{6kj}\Theta B_{1kj} + e_{7kj}IQ \cdot B_{1kj})$$
$$+ e_{8j}\Theta \cdot IQ + e_{9j}\Theta D + e_{10j}IQ \cdot D + e_{11j}\Theta_{1j}^H + \tau_{\lambda j} + \ln(\lambda_0(t)),\ j = 0,...,J. \tag{E.10}$$

We estimate model (E.5–E.10) jointly with the measurement system, which is the same as system (IV.7–IV.8) of the main text, but health stock is estimated for different time points j: $H_{1mj}^* = \beta_{1mj}\Theta_{1j}^H + \beta_{2mj} + \eta_{H1mj}$, $m = 1,...,M$; $j = 1,...,J$. As in the main text, we set $M = 4$, $J = 2$, and $K = 8$.

The features of the more general model that are not present in the parsimonious model are the following: (a) interactions among cognitive and socioemotional skills are accounted for in all equations; (b) dependence of regression coefficients on the education level modelled by interacting education D with other regressors is present in equations (E.8–E.10); (c) all coefficients of the MPH model (E.10) are time-dependent (as shown by index j)[35];

[35] As explained in the main paper, this MPH model combines the use of discrete time j and continuous time t. Index $j = 0, 1$ denotes the discrete time of observations in years 1992 and 2004.

and (d) variables $\boldsymbol{B}_1$ and Θ_1^H of equations (E.10–E.8) are time-dependent.

We test model (E.5–E.10) and find that it can be greatly simplified. Informed by the proportional hazard (PH) test, we keep time-dependence of coefficients for socioemotional skills and education, but not for the effects of IQ and mediators. Allowing for time dependence of mediators $\boldsymbol{B}_1$ and Θ_1^H in the MPH model provides only marginal changes to the model at the high cost of many additional degrees of freedom. Also, we find no evidence of interactions among education, skills, and behaviors except for the interaction among skills in equations (E.6) and (E.7). We jointly test the equality of all interaction coefficients but α_{2q} and b_2 to zero for each gender and cannot reject such tests, with p-values of 0.697 for men and 0.480 for women.[36]

[36] In equation (E.10), for women only, we set e_{4k0} and e_{4k1} to zero for the k corresponding to the social activity mediator due to lack of variation in the sample. We also use an average of the residualized achievement measures conditional on $\boldsymbol{X}$ and IQ approximating latent Θ only for the interaction term $\Theta \cdot IQ$ in equation (E.9) to ensure convergency of the estimation procedure. We make no such approximations in our main model (IV.1–IV.6), because the interaction term $\Theta \cdot IQ$ is excluded from equation (IV.5) and time-dependence of B_{1k} is excluded too.