

What is a Good School, and Can Parents Tell? Evidence on the Multidimensionality of School Output

Diether Beuermann, C. Kirabo Jackson, Laia Navarro-Sola, and Francisco Pardo

NBER Working Paper No. 25342

December 2018, Revised in September 2020

JEL No. I2,J01,J38

ABSTRACT

To explore whether schools' causal impacts on test-scores measure their overall impact on students, we exploit quasi-random school assignments and data from Trinidad and Tobago to estimate the causal impacts of individual schools on several outcomes. Schools' impacts on high-stakes tests are moderately related to impacts on crime but weakly related to impacts on important outcomes such as dropout, teen motherhood, and formal labor-market participation. To examine if parents value these causal impacts, we link them to parents' ranked lists of schools and employ discrete-choice models to infer preferences for school attributes. Many parents choose schools that improve high-stakes tests, that reduce criminality, and increase labor-market participation. Notably, many parents' choices indicate stronger preferences for impacts on non-academic outcomes than test-score impacts. These results reveal that evaluations based solely on test scores may be very misleading about the benefits of school choice, and education interventions more broadly.

Diether Beuermann
Inter-American Development Bank
1300 New York Ave, NW
Washington, DC 20577
dietherbe@iadb.org

C. Kirabo Jackson
Northwestern University
School of Education and Social Policy
Annenberg Hall, #204
2120 Campus Dr.
Evanston, IL 60208
and NBER
kirabo-jackson@northwestern.edu

Laia Navarro-Sola
Northwestern University
navarrosola@u.northwestern.edu

Francisco Pardo
Inter-American Development Bank
pardopajuelofrancisco@gmail.com

What is a Good School, and Can Parents Tell? Evidence on the Multidimensionality of School Output*

DIETHER W. BEUERMANN
Inter-American Development Bank

C. KIRABO JACKSON
Northwestern University

LAIA NAVARRO-SOLA
IIES, Stockholm University

FRANCISCO PARDO
University of Texas at Austin

September 1, 2020

Abstract

To explore whether schools' causal impacts on test-scores measure their overall impact on students, we exploit quasi-random school assignments and data from Trinidad and Tobago to estimate the causal impacts of individual schools on several outcomes. Schools' impacts on high-stakes tests are moderately related to impacts on crime but weakly related to impacts on important outcomes such as dropout, teen motherhood, and formal labor-market participation. To examine if parents value these causal impacts, we link them to parents' ranked lists of schools and employ discrete-choice models to infer preferences for school attributes. Many parents choose schools that improve high-stakes tests, that reduce criminality, and increase labor-market participation. Notably, many parents' choices indicate stronger preferences for impacts on non-academic outcomes than test-score impacts. These results reveal that evaluations based solely on test scores may be very misleading about the benefits of school choice, and education interventions more broadly.

JEL Codes: I20, J0

*Beuermann: Inter-American Development Bank, 1300 New York Avenue, NW, Washington DC 20577, USA (e-mail: dietherbe@iadb.org); Jackson: School of Education and Social Policy, Northwestern University, 2120 Campus Drive, Evanston IL 60208, USA (e-mail: kirabo-jackson@northwestern.edu); Navarro-Sola: Institutet för internationell ekonomi 106 91 Stockholm (e-mail: laia.navarrosola@iies.su.se); Pardo: Department of Economics, University of Texas at Austin, 2225 Speedway, Austin TX 78712, USA (e-mail: fpardo@utexas.edu). The statements made are solely the responsibility of the authors.

I Introduction

Is a school’s impact on test scores a good measure of its overall impact on students? Do parents value schools that improve high-stakes standardized tests? Do parents value school impacts on outcomes other than high-stakes tests? To shed light on these issues, we use administrative data from a variety of sources covering the full population of Trinidad and Tobago. To address the first question, we estimate individual schools’ causal impacts on high-stakes test scores, low-stakes test scores, dropout, teen motherhood, teen arrests, and labor market participation. Using the relationship between these estimates, we examine the extent to which school output is multidimensional (such that test score impacts may miss important dimensions of school quality). To address the second and third questions, we link our estimated impacts to parents’ school rankings and explore whether parents choose schools with positive causal impacts on these multiple outcomes – providing the first exploration into whether parents value school causal impacts on non-academic outcomes.

The motivations for this paper are twofold. The first motivation is to better understand the multidimensional nature of individual schools’ output. Researchers, practitioners, and policy-makers often rely on school’s performance on standardized tests as a measure of quality. However, because educational output may be multidimensional ([Hanushek 1971](#); [Heckman et al. 2006](#); [Kautz et al. 2017](#); [Jackson 2018](#); [Jackson et al. forthcoming](#)), schools that improve important longer-run outcomes (such as crime, college-going, and earnings) may have little impact on test scores. As such, policies that use test score impacts to make decisions (such as school closures, performance pay, accountability, etc.) may, at best, be sub-optimal and, at worst, have deleterious impacts on unmeasured long-run outcomes. To assess the importance of this, one must understand the joint distribution of individual school’s impacts across several different outcomes.¹ However, to date, only four studies examine the causal impact of individual schools on different outcomes ([Abdulkadiroğlu et al. 2014](#); [Dobbie and Fryer 2015](#); [Angrist et al. 2016](#); [Place and Gleason 2019](#)). To rely on *causal* impacts, these studies focus on a small number of oversubscribed schools that admit students using enrollment exams or randomized lotteries.² While focusing on oversubscribed

¹We now know that certain *groups of schools* that raise test scores may not improve other outcomes and vice versa. For example, [Deming \(2011\)](#) finds that winning a school choice lottery can reduce crime with little impact on test scores, [Deming et al. \(2014\)](#) find that school choice lotteries improve test scores and educational attainment (only for girls). [Beuermann and Jackson \(2020\)](#) find that attending a preferred school in Barbados improves long run outcomes but not test scores. Also, [Booker et al. \(2011\)](#) find that charter school attendance impacts on test score do not correlate with their impacts on college outcome. All these studies examine groups of schools rather than individual school impacts – precluding an analysis of the multidimensional nature of educational output by schools.

²[Place and Gleason \(2019\)](#) and [Angrist et al. \(2016\)](#) examine 31 and 26 oversubscribed charter schools, respectively. [Abdulkadiroğlu et al. \(2014\)](#) examine 6 elite selective enrollment schools, and [Dobbie and Fryer \(2015\)](#) examine a single charter school. [Dobbie and Fryer \(forthcoming\)](#) examine impacts of 45 charter schools in Texas. However, they do not have quasi-random variation so that the estimated school impacts may not capture causal impacts. In their words “*we assume unobserved determinants of students’ labor market outcomes are orthogonal to our school value-added measures....our estimates should be interpreted with this strong identifying assumption in mind.*”

schools overcomes selection biases, these studies examine a small number of schools that are necessarily non-representative – so that the patterns from these important studies may not generalize to all schools. Moreover, these studies have examined individual schools’ impacts on test scores and related educational outcomes (such as college going) but *have not* related schools’ test score impacts to a broad set of nonacademic outcomes. As such, no studies have identified individual school’s causal impacts across a representative group of schools and on a broad array of academic and non-academic outcomes simultaneously – which is necessary to rigorously explore the multidimensional nature of school output. To help fill this space, we rely on quasi-random variation to uncover the causal impact of attending 133 individual secondary schools in Trinidad and Tobago (84% of all public secondary schools). This is about the same number of public secondary schools as in Chicago, and more than in the entire state of Vermont. We estimate causal impacts of individual schools on a wide array of academic and non-academic short- and longer-run outcomes.

The second motivation for our work is to better understand parental preferences for schools. In theory, by aligning schools’ incentives with parents’ preferences, school choice policies may increase efficiency in educational production (Friedman 1955; Chubb and Moe 1990). However, if parents cannot discern school causal impacts, school choice policies will do little to increase education production or improve human capital. Indeed, there is a growing literature showing that parental preferences for schools are not systemically related to school impacts on test scores (MacLeod and Urquiola 2019; Beuermann and Jackson 2020). The few studies that directly examine parent preferences for school causal impacts (as opposed to peer quality or average outcomes) conclude that parents may not value school impacts *per se* (Rothstein 2006, Abdulkadiroğlu et al. 2020) – casting doubt on the likely efficacy of school choice. However, because test-score impacts may not capture *all* the educational output that parents value, this may not be the final word on this issue. Specifically, because educational output may be multidimensional, parents may value schools that improve outcomes that are not highly correlated with test score impacts. If so, school choice may improve outcomes valued by parents, but that are not well observed by the econometrician – leading to wrong inferences about parental preferences and the benefits of school choice. By linking our schools’ causal impacts on a broad set of outcomes to parent’s rankings of schools, we seek to provide the first examination of the extent to which parents are more likely to choose (or value) schools that causally improve test scores and *also* key nonacademic outcomes.

We use data on all applicants to public secondary schools in Trinidad and Tobago between 1995 and 2012. These data contain students’ identifying information, scores on the Secondary Entrance Assessment (SEA) (taken at age 11 at the end of 5th grade), and a ranked list of secondary schools the student wished to attend. We link these data (at the student level) to scores on low-stakes national exams taken three years later, high-stakes secondary school completion exams five years later, and a national tertiary certification exam taken seven years later. We also link these student

records to official police arrest records, birth registry data, and retirement contribution fund data. The resulting dataset allows us to track individual students over time and through 33 years of age across a host of different types of outcomes.

To estimate the causal effects of attending individual schools, we rely on the fact that the Ministry of Education (MOE) assigns most students to schools using a deferred acceptance algorithm (Gale and Shapley, 1962). Conditional on the information used in the assignment process, the algorithm-based assigned school is unrelated to both observed and unobserved student characteristics (Jackson, 2010). Exploiting this fact, for each of 133 secondary schools, we use the conditionally-random school assignment to that school as an instrument for *attending* that school (relative to the average school). In our setting, each school estimate reflects the average effect of attending a given school *among applicants to that school*. Because applicants may differ across schools, to obtain a set of relative school estimates that are comparable to each other requires that school effects are additive – i.e., the average effect of attending school A relative to school B is equal to the effect of attending school A relative to C, plus the effect of attending school C relative to B. This condition holds in our data. We implement several additional tests to support a causal interpretation of our estimates. As such, our school effect estimates reflect the relative improvement in outcomes the average parent can expect for their children from attending a given school (compared to the average school). Accordingly, our average school effect estimates are similar to others in the school effects literature (e.g. Deutsch et al. 2020, Abdulkadiroğlu et al. 2014; Dobbie and Fryer 2015; Angrist et al. 2016; Dobbie and Fryer (forthcoming); Place and Gleason 2019).

To estimate parent’s preferences for schools, we rely on the fact that a ranked list of four secondary schools is submitted as part of the secondary-school application process. Under Deferred Acceptance algorithms, among the set of schools ranked, it is rational to list them in order of true preference. As such, one could use a ranked-ordered logit model to learn something about parent preferences for schools. However, under Deferred Acceptance algorithms when all schools are not ranked, (a) not all school rankings can be observed, and (b) applicants may engage in strategic behaviors by accounting for the likelihood of admission when making choices (Chade and Smith, 2006). Researchers have addressed this by making some key assumptions. For example, Abdulkadiroğlu et al. (2020) account for not having all schools ranked by assuming that some set of school options were not in the choice set.³ Also, to account for strategic choices both Abdulkadiroğlu et al. (2020) and Hastings et al. (2009) appeal to patterns in the data to justify assuming that the choices made are not strategic. No solution is perfect. We propose an alternative solution that relies on somewhat weaker assumptions. Specifically, because the nature of any strategic choices is known, and we can obtain estimates of admission probabilities using historical data, we are able to

³Abdulkadiroğlu et al. (2020) assume that parents in New York City do not choose schools outside of their borough because such choices are uncommon.

condition on admission probabilities and uncover true preferences in our context. As such, in our choice model we include *all* schools (making no assumption about what schools are in the choice set) and also explicitly account for possible strategic behaviours. We implement a modified multinomial logit model (McFadden, 1973) to estimate parent preferences for all possible schools based on observed peer quality, proximity, average school characteristics, admission probabilities, and estimated school impacts. To ensure that our findings are not an artifact of the chosen methodology, we show that our main findings are similar in models that do not account for strategic behaviors explicitly (as has been done by other researchers), and models that rely only on the rank ordering of the choices made (i.e., a standard rank-ordered logit model).

Schools have meaningfully different average effects on an array of outcomes. Going from a school at the median of the impact distribution to one at the 85th percentile increases both low-stakes and high-stakes test scores by about 0.4σ , reduces school dropout by 3.6 percentage points, reduces teen births by 3.5 percentage points, reduces teen arrests by 1.6 percentage points, and increases the likelihood of being formally employed by 3.6 percentage points. We explore whether the magnitude of school impacts differ by incoming achievement levels. The magnitude of average school effects are similar to effects at the bottom and top of the achievement distribution for most outcomes. However, school effects tend to be larger for those at the bottom of the achievement distribution – particularly for dropout and teen motherhood. We also explore if the schools that improve outcomes on average differ from those that improve outcomes at the top or bottom of the distribution of incoming achievement. Average school effects are very similar to effects evaluated for those at the top and bottom of the achievement distribution – indicating minimal match effects, or effect heterogeneity by incoming achievement. This is further evidence that our average school effect estimates reflect the relative improvements the average parent can expect for their children from attending a given school (compared to attending the average school).

We next test for whether school impacts on test scores capture effects on other outcomes. After accounting for estimation errors, the correlations between school impacts on high-stakes tests and other outcomes are modest. For example, the correlation between impacts on high-stakes exams and non-dropout is 0.26, and that between impacts on high-stakes tests and being formally employed is 0.19. We show that these low correlations are not due to high-achieving students being more responsive to school impacts on academic outcomes and attending one set of schools while low-achieving students are responsive to school impacts on nonacademic outcomes and attend different schools. Rather, *even among a homogeneous set of students*, schools that improve academic skills are often not those that improve broader adult well-being (which parents may value).

Next we analyze parental choices to explore if parents value schools’ impacts. As in all studies of this type (e.g., Avery et al. 2013; Burgess et al. 2015; Abdulkadiroğlu et al. 2020; Hastings et al. 2006), we infer that parents “value” or “prefer” a school attribute if they are more likely

to choose schools and rank schools more highly with that attribute. Using this “revealed preference” approach, we replicate several results of existing studies. Parents assign higher priority to higher-performing schools with closer proximity, and with higher-achieving peers (Burgess et al. 2015; Abdulkadiroğlu et al. 2020). As in Hastings et al. (2006), choices are more sensitive to average school achievement among parents of higher-achieving students. We also present several novel results. Parents choose schools that are safe and schools with lower teen birth rates.⁴ Looking at test-score impacts, parents of higher-achieving children choose schools with larger positive causal impacts on high-stakes exams. This pattern is robust to controls for peer quality and average outcomes— suggesting that parents of high-achievers *can and do* disentangle schools that causally improve test scores from schools with strong average performance.⁵ Importantly, we establish that this pattern cannot be driven by match effects because school impacts on high-stakes exams are near identical for students throughout the incoming achievement distribution.

A second key innovation of our work is to be the first to measure parent preferences for schools’ causal impacts on nonacademic outcomes. Parents of high-achieving students (particularly males) are sensitive to school impacts on (reduced) arrests. Indeed, males’ parents’ choices at almost all achievement levels are more sensitive to a one standard deviation increase in arrest impacts than that for impacts on high-stakes exams. Interestingly, parents of both high- and low-achieving children (both male and female) are more likely to choose schools that raise formal labor-market participation. Remarkably, the pattern of estimates across groups and outcomes suggest that most parent choices are more sensitive to a 1σ increase in crime or labor market impacts than a 1σ increase in test-score impacts. Because the school impacts are largely the same for all students on these outcomes, we can rule out that our key results are driven by test score impacts having larger marginal effects for high-achieving children while labor market or crime impacts having larger effects on low-achieving counterparts.⁶ Because schools that causally improve test scores are often not those that reduce crime or improve labor market participation, these results have important implications for our understanding of parental preferences for schools.

We build on the school quality literature by presenting the first analysis of the relationships between school impacts on several academic and non-academic outcomes – providing direct evidence

⁴This is the first study to document parental preferences for school safety and teen motherhood using a revealed preference approach (as opposed to self-reported preferences such as Hart Research Associates (2017)).

⁵The only other paper to explore parental preferences for school causal impacts on test scores is Abdulkadiroğlu et al. (2020) who find that conditional on peer quality, parents do not appear to value schools test-score impacts. Our results are not entirely at odds with theirs because conditional on average outcomes and peer incoming achievement, only parents of students within the top 25 percent of the incoming achievement distribution exhibit strong preferences for schools that improve high-stakes exam scores.

⁶Note that any unaccounted-for match effects (in unobserved dimensions) would lead us to *understate* the importance of school impacts on choices because the average effects differences used would not be indicative of the real differences in value added for parents – **which would bias our method away from finding any effects.**

of the multidimensionality of school output.⁷ Our findings have important policy implications because test-score impacts are increasingly used for policy decisions. We also contribute to the work on parental preferences by providing the first study of parental choices based on school impacts on non-academic outcomes such as fertility, crime, and labor-market participation. We show that parents may have strong preferences for schools that reduce crime and increase labor market participation – impacts that are only weakly correlated with school impacts on tests. If this pattern holds in other settings, it could explain why researchers have found a weak link between parental preferences for schools and schools test score impacts.⁸ Our results suggest that existing evaluations of school choice based solely on test score impacts (without regard for schools’ nonacademic output) may be very misleading about their welfare effects.

The remainder of this paper is as follows; Section II describes the Trinidad and Tobago context and discusses the data used. Section III presents our empirical strategy for estimating school causal impacts. Section IV presents the magnitudes of the estimated school causal impacts and explores the potential multidimensionality of school output. Section V discusses our choice models, and presents our estimates of parental preferences. Section VI concludes.

II The Trinidad and Tobago Context and Data

The Trinidad and Tobago education system evolved from the English system. At the end of primary school (after grade 5, around 11 years old), parents register their children to take the Secondary Entrance Assessment (SEA) and provide a list of four ranked secondary school choices to the Ministry of Education (MOE). The SEA is comprised of five subjects that all students take: mathematics, English language, sciences, social studies and an essay. Students are allocated to secondary schools by the MOE based on the SEA scores and the school preferences submitted at SEA registration using the deferred acceptance mechanism summarized in Section III below.

Secondary school begins in form 1 (grade 6) and ends at form 5 (grade 10). We focus on public secondary schools of which there were 158 during our study period.⁹ All schools provide instruction from forms 1 through 5 and teach the national curriculum. Students take two externally graded exams at the secondary level, and one at the tertiary level. The first secondary exam is the National Certificate of Secondary Education (NCSE) taken at the end of form 3 (grade 8) by all

⁷Dobbie and Fryer (forthcoming) examine the relationship between charter school impacts on test scores, high school graduation, and earnings. However, as discussed above, they rely on selection on observable assumptions for identification so that the observed relationships may be subject to selection biases.

⁸This main point is also suggested in Beuermann and Jackson (2020) who examine the short and long run impacts of attending a preferred secondary school in Barbados.

⁹There are two types of public schools: Government schools (fully funded and operated by the government) and Government Assisted schools (managed by private bodies, usually a religious board, and all operating expenses funded by the government). There were 44 Government Assisted schools during our sample period. Private secondary schools serve a very small share of the student population (about 3.4 percent).

students in eight subjects.¹⁰ NCSE performance does not affect school progression or admission to tertiary education, and is therefore a low-stakes examination.

The second secondary exam is the Caribbean Secondary Education Certification (CSEC) taken at the end of form 5 (grade 10) which is equivalent to the British Ordinary-levels exam. CSEC exams are given in 33 subjects. To be eligible for university admission, one must pass five or more subjects including English and mathematics. Students who qualify for university admission based on CSEC performance could either apply and, if accepted, enroll in a tertiary institution or pursue the Caribbean Advanced Proficiency Examination (CAPE). In addition, entry level positions in the public sector require at least five CSEC subject passes. For these reasons, the CSEC is a high-stakes exam. The third exam, the CAPE, is the equivalent of the British Advanced-levels exam and was launched in 2005. The CAPE program lasts two years and includes three two-unit subjects and two core subjects (Caribbean and Communication studies). Passing six CAPE units is a general admission requirement for British universities. The post-secondary qualification of a CAPE Associate's Degree is awarded after passing seven CAPE units including the two core subjects. Finally, students who obtain the highest achievable grade in eight CAPE units are awarded Government sponsored full scholarships for undergraduate studies either in Trinidad and Tobago or abroad (including the US, Canada or UK). Given this, the CAPE is a high-stakes exam.

Secondary School Applications Data: The data include the official administrative SEA covering all students who applied to a public secondary school in Trinidad and Tobago between 1995 and 2012. These data include each student's name, date of birth, gender, primary school, residential census tract, religion, SEA scores, the ranked list of secondary school choices, and the administrative school placement by the MOE. The final SEA dataset contains information on 312,420 students across the 18 SEA cohorts. We link various additional datasets to the SEA data by full name (first, middle, and last), gender, and date of birth.

Examination Data: To track students' exam performance and educational attainment we collected data on the NCSE exams (taken 3 years after secondary school entry, typically at age 14), the CSEC exams (taken 5 years after secondary school entry, typically at age 16) and the CAPE exams (completed after 2 years of post-secondary school studies, typically at age 18). The NCSE was launched in 2009 and data are available for years between 2009 and 2015. These data include the scores for the eight subjects assessed. The NCSE data were linked to the 2006 through 2012 SEA cohorts.¹¹ The CSEC data are available for all years between 1993 and 2016. These data include the scores for each subject examination taken. The CSEC data were linked to the 1995 through 2011 SEA cohorts.¹² The CAPE data are available for years 2005 through 2016, and are linked to

¹⁰NCSE academic subjects include mathematics, English, Spanish, sciences, and social studies. NCSE non academic subjects include arts, physical education, and technical studies.

¹¹We matched 97.44 percent of all NCSE individual records to the SEA data.

¹²We matched 96.31 percent of all CSEC individual records to the SEA data. The non-match rate of 3.69 percent

the 1999 through 2009 SEA cohorts.¹³ These data contain scores for each exam unit taken.

Criminal Records: We obtained the official arrests records from the Trinidad and Tobago Police Service. For each arrest that occurred in Trinidad and Tobago between January 1990 and May 2017, these data include the offender's full name, date of birth, gender, and date of arrest. To explore teen crime, these data were linked to the 1995 through 2010 SEA cohorts.

Civil Registry: We obtained the official birth records from the Trinidad and Tobago Registrar General. For each live birth in Trinidad and Tobago between January 2010 and September 2016, these data include the mother's full name, date of birth, gender, and date of the live birth. To explore teen motherhood, these data were linked to the 2004 through 2010 SEA cohorts.

Labor Market Participation: We obtained the official registry of active contributors to the national retirement fund as of May 2017 from the National Insurance Board. These data include all persons who were formally employed and, therefore, contributing to the national social security system as of May 2017. For each affiliate, the data include the full current name, full original name prior to any name changes, date of birth, and gender. To explore formal employment among individuals aged 27 through 33, these data were linked to the 1995 through 2002 SEA cohorts.

Table 1 presents summary statistics for all our matched datasets. The population is roughly half female and the average admitted cohort size across all schools is about 217 students per incoming cohort per school (column 1). About 90 percent of students took the NCSE and 77.6 percent took at least one CSEC subject. The average student passed about 3.3 CSEC subjects and 36 percent passed five subjects including English language and mathematics (i.e. qualified for tertiary education). We also show the outcomes by sex and the selectivity of the assigned school (by incoming SEA scores). Incoming SEA scores are roughly 0.25 standard deviations lower for males than for females, and average scores of those assigned to the top ranked schools are 1.4 standard deviations higher than those assigned to the bottom ranked schools. Given the differences in incoming scores, there are relatively more females at selective schools. Females have lower dropout rates by age 14 than males (92.3 versus 88.7 percent took the NCSE). Likewise, girls score 0.45 standard deviations higher on the NCSE. Also, 43 percent of female students qualify for tertiary education while only 29 percent of males do. Students at the most selective schools score 0.86 standard deviations higher on the NCSE than the average student at less selective schools. They also pass about 5 CSEC subjects on average, and 59.8 percent qualify for tertiary education; while this is only accomplished by 12.7 percent of students at the least selective schools (column 5).

Looking at post-secondary education, about 20.6 percent of students took at least one CAPE unit, 15.4 percent earned an Associate's degree, and only 0.99 percent earned a CAPE scholarship. Females passed 1.8 CAPE units, and 19.3 percent earned an Associate's degree. In comparison,

closely mimics the share of students served by private schools (3.4 percent) who would not have taken the SEA.

¹³We matched 96.6 percent of all CAPE individual records to the SEA data.

males passed 1.1 units, and only 11.4 percent earned an Associate’s degree. At the most selective schools, 34.8 percent of students took at least one CAPE unit and 26.9 percent earned an Associate’s degree. Among those at less selective schools, only 4.7 percent took at least one CAPE unit and 2.5 percent earned an Associate’s degree.

Moving to nonacademic outcomes, 3.3 percent of the population had been arrested by age 18. Arrests are concentrated among males of which 5.8 percent had been arrested by age 18. Arrests rates are low (1.7 percent) among students from more selective schools, and are higher (4.8 percent) among students at the least selective schools. A similar pattern is observed for teen motherhood. While 6.5 percent of girls at the top schools had a live birth before age 19, as much as 15 percent of females at the bottom schools did. Finally, 75.7 percent of the population is formally employed (as an adult). However, formal employment is somewhat higher for males than for females, and for those assigned to more selective schools than for those assigned to less selective schools. Next, we describe how we estimate schools’ causal impacts on these key outcomes.

III Estimating School Impacts

Our first aim is to uncover the causal impact of attending each school j relative to the average school. To do this we need to exploit variation in school attendance that is unrelated to unobserved determinants of student outcomes. Following [Jackson \(2010\)](#), we exploit the fact that the Ministry of Education (MOE) uses a deferred acceptance mechanism to create an initial set of school assignments for students. This initial assignment is unrelated to unobserved student or parent actions and characteristics (by construction), so that conditional on the variables used by the algorithm to form these assignments, the assignments are unrelated to unobserved determinants of student outcomes. We rely on this exogenous variation in school assignments to uncover schools causal impacts.

School assignments are as follows: Parents submit a rank-ordered list of secondary schools they wish their children to attend *before* they sit the SEA. Once the exams are scored, the top scoring student is assigned to her top choice school, then the second highest scoring student is treated similarly, and so on until all school slots are filled. Once a given school’s slots are filled, that school is then taken out of the pool, and students who had that school as their top choice will be in the applicant pool for their second choice. This process continues until all school slots are filled or all students are assigned.¹⁴ We refer to this rule-based initial assignment as the “tentative” assignment.

A key feature of the mechanism is that each school has a test score cutoff above which applicants are tentatively assigned to the school and below which they are not. Conditional on school choices and SEA score, the tentative assignments are beyond parent, student or school admin-

¹⁴See [Appendix A](#) for a more detailed description. Because all schools have the same preferences for students, this is similar to a serial dictatorship game. If students listed a complete ranking of all schools, this would essentially be serial dictatorship. However, because the lists are not complete, is more accurately described as deferred acceptance ([Gale and Shapley, 1962](#)). Students have incentives to truthfully reveal their rankings among chosen schools.

istrator control and are therefore unrelated to unobserved determinants of student outcomes. In reality, the official MOE school placement differs from this initial assignment because principals at Government Assisted schools are allowed to admit up to twenty percent of the incoming class at their discretion.¹⁵ Even though this discretion is often not used by principals, to avoid any bias, we follow [Jackson \(2010\)](#) and do not rely on the official MOE placement, but rather use only the exogenous variation in the tentative rule-based assignment.

Because there are multiple cutoffs embedded in the assignments algorithm, this assignment mechanism generates *two* sources of exogenous variation that we can exploit ([Jackson, 2010](#)); (a) variation around the cutoffs for each school (based on applicants scoring just above and just below each cutoff) and (b) variation across cutoffs (based on *all* students including those far away from the cutoffs). We discuss each source of exogenous variation in turn.

Variation around individual cutoffs (Discontinuity Variation):

The most familiar source of exogenous variation is the variation around the cutoffs. Consider this scenario illustrated in [Figure 1](#): There are two choice groups; choice group 1 who list school 1 as their top choice and school 3 as their second; and choice group 2 who list school 2 as their top choice and school 3 as their second choice. Both groups have the same second choice school (school 3), but different top choices (school 1 and 2, respectively). Applicants to school 1 who score above 82 on the SEA are granted admission, while school 2 has a higher cut-off such that applicants to school 2 who score above 92 on the SEA are granted admission. Among those with school 1 as the top choice (i.e., those subject to the cutoff for school 1), one can estimate the impact of school 1 versus 3 using the cutoff variation just above and below 82. Similarly, among those with school 2 as the top choice (i.e., those subject to the cutoff for school 2), one can estimate the impact of school 2 versus 3 using the cutoff variation just above and below 92. This is the standard variation exploited in a regression discontinuity (RD) design. This variation is valid so long as the cutoffs are exogenous. In [Appendix A](#) we present the standard battery of empirical tests to support the exogeneity of the test score cutoffs. That is, scoring above the cutoff is not associated with a jump in density, or change in predicted outcomes, but *is* strongly associated with an increased likelihood of attending ones preferred school. This is the first (but not only) source of variation.

Variation across cutoffs (Difference in Difference Variation):

Because there are multiple cutoffs (one for each school each year), the RD variation is not all the identifying variation embedded in the assignment mechanism. That is, one can exploit the fact that individuals with the same test scores may be marginal for different schools' cutoffs due to small differences in their choices (which are observable and can be accounted for directly). To see this,

¹⁵Government Assisted schools account for about 30 percent of national enrollment. Therefore, students admitted upon discretion of principals at these schools could account at most for 6 percent of national enrollment.

consider students away from the cutoffs in [Figure 1](#). All students who score below 82 (irrespective of choices) end up in school 3. As such, any differences across choice groups (among those with low scores) cannot be due to school differences and will reflect the impact of any unobservables correlated with the different choices. However, among students who score above 92, the two choice groups face different cutoffs and end up in different schools, so that the differences reflect both the difference in school impacts and the impact of those unobservables correlated with the different choices. If the choice group effects are additively separable from that of test scores, one can use a difference-in-difference type approach to identify the effect of attending school 1 versus school 2 *even among those far away from the cutoff*.¹⁶ To show that our school impacts based on this variation are not biased by violations of the aforementioned additive separability assumption, in [Section III.3](#) we show that our estimated school effects are virtually identical in models that include flexible interactions between school choices and test scores and those that do not.

Making Comparisons Across all Schools:

The set of students described above allow one to estimate the relative effects for schools 1, 2, and 3. However, our aim is to compare the causal effect of attending each school to that of *any* other particular school – i.e., we need *all* our school effects to be directly comparable. With different groups of students who make different choices (and therefore face a different set of cutoffs) one can estimate impacts for other sets of schools (say schools 3, 4, and 5). If school effects are additive (i.e., the effect of attending school 1 relative to 5 is equal to the effect of attending school 1 relative to 3 plus the effect of attending school 3 relative to 5), then each school can be linked to all other schools through a chain of overlapping within-group comparisons and one can compare each school to every other school.¹⁷ This is the approach we employ. Because making comparisons across all schools relies on the assumption that school effects are additive, in [Section III.3](#) we show that this condition holds in our setting.

¹⁶Concretely, the difference in outcomes across choice groups among students with low scores (below 82) will be the choice-group effect, C . The difference in outcomes across choice groups among the high-scoring students (above 92) is the choice effect (C) plus the school effect (S). The difference in outcomes among high scorers ($C+S$), minus the difference in outcomes among low scorers (C), will uncover the school effect ($C+S-C=S$). By a similar logic, one can use individuals who score between 82 and 92 to identify the effect of school 1 versus school 3. In this example, one can uncover the relative effectiveness of schools 1, 2, and 3.

¹⁷Suppose group A allows a comparison of schools 1, 2, and 3, while group B allows a comparison of schools 4, 5, and 6. So long as there is some other group that has one school from each group (say schools 2, 4, and 9) then all schools in 1, 2, 3, 4, 5, 6, and 9 can be compared to each-other. This example highlights that if each school can be linked to all other schools through a chain of overlapping within-group comparisons, all schools can be compared to all other schools. This identification requirement is similar to that for estimating teacher value-added while also controlling for school effects ([Mansfield, 2015](#)).

III.1 Relying Only on the Identifying Variation

For clarity of exposition, we will refer to assigned school τ and attended school j . Based on the algorithm, a student is tentatively assigned to school τ if they (a) had school τ in their list of choices, (b) scored above the cutoff for school τ , and (c) did not score above the cutoff for a more preferred school in their choices. It follows that *conditional on school choices* students will only be assigned to different schools if some scored above the cutoff for a desired school while the others did not (i.e., variation due to their observable scores). Also, *conditional on having the same test scores* students will only be assigned to different schools if some made a different set of school choices than the others (i.e., variation due to their observable choices). Importantly, both sources of variation in school assignment described above will be unrelated to unobserved determinants of student outcomes so long as the location of school cutoffs is exogenous.

Under the testable modelling assumptions discussed above (i.e., (1) additive separability of test scores and school choices in determining outcomes, and (2) additivity of school effects), conditional on smooth flexible functions of incoming SEA scores and explicit controls for student choices, students initial tentative assignments are as good as random if locations of the test score cutoffs are exogenous to other student characteristics (as shown empirically in [Appendix A](#)). One can therefore obtain *assigned* school’s causal effects by estimating (1) by Ordinary Least Squares (OLS).

$$Y_{i\tau ct} = \Sigma(I_{i,\tau} \cdot \theta_{\tau}^{ITT}) + f(SEA_i) + \lambda_c + \mathbf{X}_{it}'\delta + C_t + \varepsilon_{i\tau ct} \quad (1)$$

In (1), $Y_{i\tau ct}$ is the outcome of interest for student i who was assigned to school τ , and belongs to choice group c and SEA cohort t . $I_{i,\tau}$ is an indicator equal to 1 if student i was *assigned* to school τ . $f(SEA_i)$ is a 5th-order polynomial of the incoming SEA score. X_{it} is a vector of individual-level baseline characteristics (measured at SEA registration) including sex, district of residence fixed effects, the distance between the primary and secondary school, and religion fixed effects. C_t denotes SEA cohort fixed effects; while $\varepsilon_{i\tau ct}$ is an individual-level disturbance.

Key variables for our analysis are the choice group fixed effects λ_c . These identify separate intercepts for groups of individuals who made the same school choices in the same order.¹⁸ Importantly, the choice-group indicators identify groups of individuals who may have the same SEA score but are subject to different school cutoffs – which allows for the difference-in-difference identification across cut-offs (using individuals away from the cutoff as outlined above).¹⁹ The

¹⁸In most years, students could list four choices. However, for SEA cohorts 2001-2006 the MOE allowed students to list up to 6 different school choices (instead of the usual 4). Therefore, we grouped students with unique combinations of the first 4 choices within one set of fixed effects, and included separate sets of fixed effects for choices 5 and 6.

¹⁹In principle one could efficiently rely only on the discontinuity variation by using the “tie breaking” approach proposed in [Abdulkadiroğlu et al. \(2019\)](#). However, this would preclude our use of the variation away from the cutoffs which is instrumental to our ability to compare school effects to the average school.

estimated $\hat{\theta}_{\tau}^{ITT}$ s from (1) identify the average treatment effect of being tentatively assigned to each school τ (relative to the same comparison school). These estimates can be standardized to reflect the impact of being assigned to school τ relative to being assigned to the average school.²⁰

III.2 Using Instruments to Obtain Causal School Attendance Impacts

The previous section shows that we can obtain the causal effect of being tentatively *assigned* to individual schools. To obtain clean Treatment-on-the Treated (TOT) effects of *attending* a particular school j relative to the average school, we use the rule-based school assignments as instruments for actual school attendance.²¹ Identification of individual school effects requires one instrument per alternative (Kirkeboen et al., 2016). We satisfy this condition in our setting by using indicators for being assigned to each school as instruments for attending each school. Ideally, all 158 schools would have strong first stages, but this is not the case. As such, to avoid being under-identified, we exclude the school assignment and attendance indicators for the 25 schools with the weakest first stages.²² We can obtain clean causal estimates for 84 percent of all public secondary schools in the nation (i.e., 133 schools). Our resulting two-stage least squares (2SLS) model is as follows:

$$Y_{i,jct} = \Sigma(I_{i,j} \cdot \theta_j^{TOTiv}) + f(SEA_i) + \lambda_c + \mathbf{X}_{it}'\delta + C_t + \varepsilon_{i,jct} \quad (2)$$

$$I_{i,j} = \Sigma(I_{i,\tau} \cdot \pi_{\tau j}) + f(SEA_i) + \lambda_c + \mathbf{X}_{it}'\delta + C_t + v_{i,jct}, \quad \text{for each } j \in J \quad (3)$$

The endogenous variables in the second stage equation (2) are the 133 individual school attendance indicators ($I_{i,j}$) and our excluded instruments are the 133 individual school assignment indicators ($I_{i,\tau}$). We code a student as attending school j if the student was enrolled in school j at

²⁰Note that because we condition on individuals rank-ordered choice lists, and proximity to the school, our approach is similar to Abdulkadiroğlu et al. (2020) who assume that “any omitted variable bias afflicting OLS value-added estimates is due either to spatial heterogeneity captured by distances to each school (D_i) or to the preferences underlying the rank-ordered lists submitted to the assignment mechanism.” (Page 1513). However, unlike Abdulkadiroğlu et al. (2020) where additionally “noncompliance with the assignment mechanism, are presumed to be unrelated to potential outcomes,” because we can observe **both the initial assignment and the school attended**, we **do not** rely on the additional identifying assumption of random compliance.

²¹Noncompliance with the initial assignment may occur for two reasons. First, to allow principals at Government Assisted schools (akin to charter schools in the U.S. or choice schools in the U.K.) some flexibility, principals at these schools (which account for 20 percent of the student population) are allowed to replace as much as the bottom 20 percent of students tentatively assigned to their schools with any student of their choosing. After principals at Assisted schools decide who they would like to admit (that are not on their tentative admit list), the MOE adjusts the initial tentative assignments before making the official MOE placements (see Appendix A for a detailed description of this process). The second source of noncompliance is that students may self-select and therefore not attend the schools to which they are placed. Specifically, students (and parents) may attempt to transfer to schools other than their initial placement or decide to attend a private school if they do not like their initial placement. While the first source of noncompliance is specific to the Trinidad and Tobago context, the second would exist in most contexts. We suspect that both sources of noncompliance would not be random and may be related to students’ potential outcomes which would render estimated impacts based on attended schools without a convincing source of exogenous variation biased.

²²This group of 25 schools constitute the omitted category in our estimation of individual school impacts.

the time of writing the CSEC exams. Therefore, attended school j and assigned school τ are the same for those who comply with the quasi-random assignment. While each attended school has its own assignment instrument, *all 133 school assignment indicators enter as regressors in each of the 133 first stages* denoted by (3). Since the $\hat{\theta}_j^{TOTIV}$ are standardized to be mean zero and unit variance, they provide an unbiased causal estimate of the effect of attending school j relative to the average school in Trinidad and Tobago for those who comply with the assignment.

We implement the approach outlined above to estimate individual schools’ causal impacts on several outcomes. These outcomes include multiple high-stakes test scores, low-stakes test scores, school dropout, arrests by age 18, teen motherhood, and formal labor-market participation. Because we have several test outcomes, we combine similar outcomes into indexes. We created a “High-Stakes Exams” index by running a factor analysis (using the principal-component factor method) on all the CSEC and CAPE outcomes and then predicting the first unrotated factor. Using this same approach, we computed a “Low-Stakes Exams” index grouping both NCSE academic and non-academic performance. Appendix Table B1 shows the individual outcomes that comprise each index and the weights used to compute each index. No dropout by age 14, no live birth by age 19, no arrests by age 18, and formal labor market participation each constitute their own single-outcome index. All indexes were standardized to have zero mean and unit variance.

III.3 Testing the Functional Form Assumptions

In addition to satisfying the standard RD exclusion restrictions (which we show hold in Appendix A), our empirical strategy relies on two important functional form assumptions. The first is the additive separability of choices and test scores (required for the difference in difference variation), and the second is that school effects are additive (required for us to be able to compare all school estimates to each-other). In this section, we describe empirical tests of these two key assumptions and show that these two conditions hold in our setting.

Additive Separability of Choices and Test Scores I: Regression Discontinuity Variation vs. All Variation: Our estimation approach (*using some of the difference-in-difference variation*) relies on the assumption that the impacts of choices and scores are additive. To show that our estimates are not driven by this assumption, we validate our school estimates (that use all the variation) using only the local RD variation through the cutoffs that do not rely on the assumption that test scores and choices are additively separable. If the results using all the variation are similar to those obtained using only the RD variation right around the cutoffs, it would indicate that (a) this additive assumption is valid and that (b) our estimated school effects are not driven by this assumption. Indeed, we show this.

Specifically, for each school in each year, we estimate the RD effect of scoring above the rule-based assignment cutoff for that school in that year. As pointed out in Kirkeboen et al. (2016), this

is the difference in school quality between attending the preferred school versus attending the set of counterfactual schools for those applying to that school in that year. We also obtain an estimate of the effect of scoring above the cutoff on the impact of the attended school (i.e., $\hat{\theta}_j^{TOTIV}$) relative to that of *the same set of counterfactual schools*.²³ If our school IV estimates reflect the causal impact of attending school j relative to the average school, then the RD effect on actual outcomes should be similar to the RD effect on our estimated IV school impacts.²⁴ To show this, we regress the RD effects on actual outcomes on the RD effects on the estimated IV school impacts and test if the estimated slope is statistically indistinguishable from 1. We first implement this test using the raw school effects. Following [Hastings et al. \(2015\)](#), to account for noisiness in the RD estimated effects, we weight each RD estimate by the inverse of its squared standard error. Additionally, to account for estimation errors in the RD effects on school impacts, we implement Empirical Bayes estimates of each cutoff effect.²⁵ The binned scatter-plot, pooled across outcomes ([Figure 2](#)) presents the relationship for the raw school impacts (left) and the Empirical Bayes estimates (right) – the conclusions are the same. Using the raw estimates the estimated slope coefficient is 1.01, and the test that this is different from 1 yields a p -value of 0.974. The slopes using the Empirical Bayes estimates (with or without weights) are close to 1, and both yield p -values well above 0.5.²⁶ This indicates that our IV estimates are valid under the weaker RD identifying assumptions.

Additive Separability of Choices and Test Scores II: Robustness to Interactions: The test above shows that the RD variation (which does not require additive separability) yields very similar results to those based on all the variation – indicating that the assumption of additive separability is satisfied in our setting. To further assess the extent to which interactions between test scores and

²³All RD estimated effects of scoring above the cutoff for each school in each year use the optimal bandwidths derived from [Imbens and Kalyanaraman \(2012\)](#).

²⁴A similar test was implemented in [Hastings et al. \(2015\)](#) and [Beuermann and Jackson \(2020\)](#). This is also similar in spirit to the random assignment validation of school value-added in [Deming et al. \(2014\)](#). See [Appendix C](#) for further discussion of this test.

²⁵Specifically, for any particular outcome, the *predicted* RD effect for school j is the weighted difference in estimated school impacts between those just above and below the cutoff for school j . We can express the estimated parameter as $\hat{\zeta}_{\hat{\theta}_j} = \sum_{k \in A} a_k(\hat{\theta}_k) - \sum_{k \in B} b_k(\hat{\theta}_k) = \sum_{k \in A} a_k(\theta_k + \varepsilon_k) - \sum_{k \in B} b_k(\theta_k + \varepsilon_k)$, where A is the set of schools the students attend above the cutoff, B the set of schools the students attend below the cutoff, a_k and b_k the proportions in which they do so and ε_k the estimation error for the school impact θ_k . If we assume that the school impacts are not independent within each cutoff, but that the estimation errors are, we can approximate the variance of this estimated parameter as $Var(\hat{\zeta}_{\hat{\theta}_j}) = \sum u_k \sigma_{\theta}^2 + \sum v_{lm} Cov_{\theta} + \sum w_k SE_k^2$, where σ_{θ}^2 is given by the magnitude of the school impacts, Cov_{θ} is approximated by the covariance between each pair of schools the students applied to and SE_k^2 is given by the square of the standard error of the school impact. If we also include the squared standard error of the RD estimate, $SE_{\hat{\zeta}}^2$, the reliability ratio of our RD estimate is given by

$$\lambda_j = \frac{\sum u_k \sigma_{\theta}^2 + \sum v_{lm} Cov_{\theta}}{(\sum u_k \sigma_{\theta}^2 + \sum v_{lm} Cov_{\theta}) + (\sum w_k SE_k^2 + SE_{\hat{\zeta}}^2)} \quad (4)$$

Our Empirical Bayes estimate of the *predicted* effect of cutoff j is therefore $[\lambda_j \times \hat{\zeta}_{\hat{\theta}_j}]$

²⁶With no adjustment for errors, the slope is 0.74 and one fails to reject that the slope differs from 1 at the 5% level.

choices is a problem in our setting, we estimate models with interactions between test scores and flexible function of choices. We then compare the 2SLS effects we obtain with these interactions to those that we obtain without them (summarized in [Appendix C](#)). In all cases, the correlations between the resulting effects is close to 1. That is, consistent with the validation test above, our results are not driven by the assumption of additive separability of choices and test scores.

Additivity of the School Effects: As mentioned above, our ability to compare all schools to all other schools relies on the assumption that the effect of schools is additive. We show that this is the case. We define the parameter $\theta_{m,k}^{ITT}$ as the effect of being assigned to school m relative to being assigned to school k . If school effects are additive, then $\theta_{m,k}^{ITT} = \theta_{m,l}^{ITT} + \theta_{l,k}^{ITT}$. If school effects are not additive, then this condition will generally not hold. As such, to test for additivity, we implement the sample analog of this test. That is, we test whether $\hat{\theta}_{m,k}^{ITT} = \hat{\theta}_{m,l}^{ITT} + \hat{\theta}_{l,k}^{ITT}$.

To do this, for each pair of schools m and k , we restrict the data to students that had both schools in their choices. We then estimate equation (5), where I_m is an indicator for being assigned to school m , and $I_{\neq m,k}$ is an indicator connoting assignment to a school other than school m or school k .²⁷

$$Y = I_m \cdot \theta_{m,k}^{ITT} + I_{\neq m,k} \cdot \theta_{n,k}^{ITT} + f(SEA) + \lambda_c + \mathbf{X}'\delta + C_t + \varepsilon \quad (5)$$

In equation (5), because the omitted school assignment is school k , $\theta_{m,k}^{ITT}$ captures the effect of being assigned to school m relative to school k . We then find all intermediate schools l such that $\hat{\theta}_{m,k}^{ITT_{sum}} = \hat{\theta}_{m,l}^{ITT} + \hat{\theta}_{l,k}^{ITT}$ can be computed. As is typical in the value-added literature, because we will use this sum of estimates as a regressor, we form Empirical Bayes estimates by multiplying each raw sum of effects $\hat{\theta}_{m,k}^{ITT_{sum}}$ by an estimate of their reliability $\hat{\lambda}_{m,k}$.²⁸ That is, our Empirical Bayes estimate of the indirect effect is $\hat{\lambda}_{mk} \hat{\theta}_{m,k}^{ITT_{sum}}$. Using (7) we test whether $\beta = 1$, to implement a formal test of additivity of school effects.

$$\hat{\theta}_{m,k}^{ITT} = \alpha + \beta [\hat{\lambda}_{mk} \hat{\theta}_{m,k}^{ITT_{sum}}] + \varepsilon \quad (7)$$

²⁷Estimated standard errors are clustered at the assigned school level.

²⁸Where $\sigma_{\theta_{sum}}^2$ is the variance of the sum of the two school effects to create the indirect estimate, and $\sigma_{e,sum}^2$ is the variance of the estimation error for the indirect effect, the reliability ratio of our indirect school effect can be written as

$$\lambda_{mk} = \frac{\sigma_{\theta_{sum}}^2}{\sigma_{\theta_{sum}}^2 + \sigma_{e,sum}^2} \quad (6)$$

Assuming that the estimation errors are uncorrelated across schools, our empirical estimate of $\sigma_{e,sum}^2$ is $(SE_{\theta_{ml}}^2 + SE_{\theta_{lk}}^2)$. If we make the additional standard assumption that the estimation errors are uncorrelated with the true effects, then the total variance of the estimated indirect school effects is $\sigma_{\hat{\theta}_{sum}}^2 = \sigma_{\theta_{sum}}^2 + \mathbf{E}(SE_{\hat{\theta}})^2$, where $\mathbf{E}(SE_{\hat{\theta}})^2$ is the sum of the squared expected standard errors $SE_{\hat{\theta}_{ml}}$ and $SE_{\hat{\theta}_{lk}}$, which are approximated by the mean standard error of the estimates (alternatively, we also approximated them with the median standard error of the estimates yielding equivalent results shown in [Appendix C](#)). By subtracting our estimates of $\mathbf{E}(SE_{\hat{\theta}})^2$ from the overall variance of the estimated indirect effects, we can estimate $\sigma_{\theta_{sum}}^2$. To create our estimated reliability ratio, we replace the parameters in (6) with the empirical estimates of these parameters ([Kane and Staiger, 2008](#)).

We implement this test on sets of school pairs that have at least 100 observations assigned to school m , school k , or school l across all years – this corresponds to about 15 observations per school per year. A scatterplot of the direct estimates against the indirect estimate (i.e., the sum) are presented in [Figure 3](#). The datapoints line up remarkably well along the 45 degree line. If we focus on more precisely estimated school effects with 150 observations, the estimated unweighted slope is 1.03 and is 0.97 when weighted by the inverse of the sum of the number of observations. In all cases, the null hypothesis that the effects are additive (i.e. $\beta = 1$) is not rejected. This is strong evidence that our school effects are additive. Note that this does not rule out the existence of *all* match effects ([Abdulkadiroğlu et al. 2020](#); [Kirkeboen et al. 2016](#)) but it does show that (a) any potential match effects do not bias our estimates in an appreciable way, and (b) our interpretation of school effects as reflecting the average relative effect that parents can expect for their kid as a result from attending school j is valid.

IV Magnitude of the School Impacts

To assess the magnitude of the school impacts on each outcome, we estimate the standard deviation of these impacts. Because the school effects are estimated with error and noise, simply reporting the variance of the estimated effect would overstate the magnitude of schools’ actual impacts. As is common practice in the school and teacher effects literatures (e.g., [Kane and Staiger 2008](#); [Chetty et al. 2014](#)), to account for this, we rely on the correlations between school effects across years to identify the variance of schools persistent effects. Following [Jackson \(2013\)](#), we do this in two steps. First we estimate the IV impacts of each school in even years ($\hat{\theta}_{j,even}^{TOT_{IV}}$) and in odd years ($\hat{\theta}_{j,odd}^{TOT_{IV}}$). Let $p \in \{\text{even, odd}\}$. These even and odd year school estimates contain a permanent school effect (θ_j^{TOT}) and a transitory effect (μ_{jp}). In a second step, under the assumption of joint normality of these components and the covariance structure in (8), we uncover Maximum Likelihood estimates of the variance of the persistent school impacts ($\sigma_{\theta_j^{TOT}}^2$) and of the transitory school impacts ($\sigma_{\mu_{jp}}^2$).

$$\begin{bmatrix} \theta_j^{TOT} \\ \mu_{jp} \end{bmatrix} \sim N \left(0, \begin{pmatrix} \sigma_{\theta_j^{TOT}}^2 I_J & 0 \\ 0 & \sigma_{\mu_{jp}}^2 I_M \end{pmatrix} \right) \quad (8)$$

[Table 2](#) reports estimates of the standard deviation of the persistent school impacts for each outcome along with their 95 percent confidence intervals.²⁹ To aid interpretation, all outcomes are coded so that higher values reflect better outcomes.

High-stakes exams: The persistent school effects for the high-stakes dimension have a standard

²⁹To be conservative, we exclude outlier schools with estimates lying 4σ away from the median. In [Appendix C](#), we show estimates of the transitory school impacts, and also estimates with and without removing outliers.

deviation of 0.39 (with a 95% confidence interval between 0.35 and 0.45). This indicates that attending a school at the 85th percentile of the impact distribution compared to attending a school at the median would increase high-stakes test performance by approximately 0.39 standard deviations. These estimated school impact sizes are larger than those found for school impacts on test scores in North Carolina (Jackson 2013; Deming 2014), and than those of attending Promise Academy in the Harlem Children’s Zone (Dobbie and Fryer, 2015); but on the same order of magnitude as that of attending Boston urban charter schools (Angrist et al., 2013).

Low-stakes exams: The magnitude of the school impacts on high-stakes and low-stakes tests are very similar. The standard deviation of the persistent school effect on the low-stakes index is 0.41 (with a 95% confidence interval between 0.36 and 0.47). That is, attending a school at the 85th percentile of the impact distribution compared to a school at the median would increase low-stakes test performance by approximately 0.41 standard deviations.

Dropout: The first non-test-score outcome we examine is dropout. Because all students take the NCSE exams around age 14, our measure of dropout is not being registered for the NCSE exams. To aid interpretation, we present the standard deviation of school impacts on the binary outcomes directly in the lower panel (as opposed to the impacts on the standardized outcome in the top panel). The estimated standard deviation of the persistent school impacts is 0.036 – indicating that attending a school at the 85th percentile of the impact distribution compared to attending a school at the median would reduce high school dropout by approximately 3.6 percentage points. Our estimated impact of attending a school with 1 σ higher impact on dropout is smaller than that of attending a charter high school (Booker et al., 2011) or winning a lottery to a choice school in North Carolina (Deming et al., 2014). As such, our estimates are conservative relative to what one might expect based on existing studies.³⁰

Crime: We present estimated school impacts on having any arrest by age 18 in the lower panel (as opposed to the crime index in the top panel). The standard deviation of the persistent school effects is 0.016 which means that being assigned to a school at 85th percentile of the impact distribution as opposed to the median would reduce the likelihood of being arrested as a teenager by 1.6 percentage points. Relative to the average arrest rate of 3.3 percent, this is almost a fifty-percent reduction in teen arrests. Coincidentally, Deming (2011) finds that winning a lottery to attend a better school reduced arrests among high-risk youth by about fifty percent. The 95 percent confidence interval does not include zero so that these school impacts are real and persistent over time.

Teen motherhood: The standard deviation of the persistent school effects on teen motherhood is 0.035. The 95 percent confidence interval is between 0.02 and 0.06. Going from a school at the median to one at the 85th percentile of the impact distribution would reduce teen live births by 3.5

³⁰The charter school and choice school literatures find impacts on high school completion between 10 and 15 percentage points. Our estimates suggest that these choice schools may be as much as 2 σ above the typical school.

percentage points. While there are many studies of the impact of teen motherhood on schooling, we believe that this is the first study to examine the extent to which individual schools have causal impacts on teen motherhood.³¹ Given that the teen live birth rate is around 10 percent on average, these represent economically important relative impacts.

Labor market participation: The final outcome we examine is participating in the formal labor market. We examine school effects on the likelihood that a student is observed with positive earnings in the formal labor market (i.e. contributing to the national social security system). To aid interpretation, we present estimated school impacts on this binary outcome in the lower panel. The standard deviation of the persistent school effects on this outcome is 0.036 (with a 95% confidence interval between 0.029 and 0.044). Going from a school at the median of the impact distribution to one at the 85th percentile would increase the likelihood of being formally employed by about 3.6 percentage points. This impact is economically meaningful.

The fact that schools have economically meaningful impacts on an array of different outcomes is not surprising. However, the policy implications of this result depends on the extent to which these school impacts are all well-measured by a school's impact on high-stakes exams. If school impacts across these outcomes are highly correlated, then school impacts on high stakes exams would identify those schools that will improve life outcomes. Using these estimates to inform policy (such as allocating funds, school closures, or rewards) would likely improve all outcomes. However, if those schools that improve high-stakes exams are a different set of schools than those that improve labor market participation or those that reduce crime, it would mean that commonly used test-based measures of school quality are incomplete. In such a scenario, using school impacts on high stakes exams to inform policy could have deleterious impacts on other outcomes and could lead to multitasking problems (Holmstrom and Milgrom, 1991). We examine the relationship between school impacts across these different outcomes below.

IV.1 Is School Quality Unidimensional?

Many recent education policy (e.g., **No Child Left Behind** in the U.S. or **League Tables** in the U.K.) are predicated on the idea that schools that raise test scores are better schools. While this may be true *on average*, if school quality is multidimensional, school impacts on test scores may not capture impacts on important dimensions of quality. To assess this, we present scatter plots of the estimated school impacts on high stakes test scores and other outcomes. In **Figure 4**, the correlations appear to be generally low – suggesting that test scores may not capture multiple dimensions of school quality. However, because the school effects are estimated with error, the correlations between the estimated school effects across outcomes will understate the correlations

³¹In related work, Jackson (2019) finds that converting existing coeducation school to single-sex reduced the teen birth rate by about 4 percentage points. Also, Beuermann and Jackson (2020) find that attending a preferred school in Barbados decreases the teen motherhood rate by about 6 percentage points.

between the true school effects.

To account for disattenuated correlations due to estimation errors in the estimated school effects, we obtain Maximum Likelihood estimates of the true correlations between each pair of outcomes (as in [Abdulkadiroğlu et al. 2020](#)). Following the notation in equation (8), consider two outcomes 1 and 2 so that θ_{1j}^{TOT} is the effect of school j on outcome 1 and θ_{2j}^{TOT} is the effect of school j on outcome 2. Similarly, μ_{1jp} and μ_{2jp} are the transitory effects of school j on outcomes 1 and 2. Under the assumption that the effects on outcomes 1 and 2 follow a joint normal distribution as in equation (9), one can estimate the correlation (net of estimation errors) between the effect on any two outcomes 1 and 2, (that is, ρ_{12}), by Maximum Likelihood.

$$\begin{bmatrix} \theta_{1j}^{TOT} \\ \theta_{2j}^{TOT} \\ \mu_{1jp} \\ \mu_{2jp} \end{bmatrix} \sim N \left(0, \begin{pmatrix} \sigma_{\theta_{1j}^{TOT}}^2 I_J & \rho_{12} I_J & 0 & 0 \\ \rho_{12} I_J & \sigma_{\theta_{2j}^{TOT}}^2 I_J & 0 & 0 \\ 0 & 0 & \sigma_{\mu_{1jp}}^2 I_M & 0 \\ 0 & 0 & 0 & \sigma_{\mu_{2jp}}^2 I_M \end{pmatrix} \right) \quad (9)$$

[Table 3](#) presents correlations between school impacts on the six outcomes. The correlations between school impacts across dimensions are modest – suggesting that different schools are good at improving different student outcomes. Focusing on high stakes test score impacts, school impacts on high-stakes tests score do not explain large shares of school effect on the other outcomes. The correlation between school impacts on the high-stakes and low-stakes exam indexes is only 0.06. While this may seem low, recall that in addition to the difference in stakes, the low stakes exams include non-academic subjects such as physical education and home economics. Indeed, the correlation between school impacts on the high-stakes exams and the low-stakes *academic* exams is 0.25, while that for the low-stakes non-academic exams is -0.05 ([Appendix Table B2](#)). The correlation with dropout is 0.26, and the correlation with being formally employed is 0.19 ([Table 3](#), column 1 - unweighted). This suggests that schools that improve high-stakes exams are associated with relatively small improvements in these other outcomes. The correlations between performance on high-stakes and the absence of teen motherhood and arrests are positive, but moderate (0.13 and 0.47 respectively). This suggests that schools that improve high-stakes exams performance also tend to improve these outcomes, on average, but that only about 22.1 percent (i.e. $0.47 \times 0.47 = 0.221$) of the variation in school impacts on reduced arrests can be explained by effects on high-stakes exams, and *vice versa*. While this may seem low, a disconnect between school impacts on high-stakes and other outcomes like low-stakes exams and crime has been documented in other settings (e.g. [Mbiti et al. 2019](#); [Deming 2011](#)).

IV.2 Accounting For Differences by Incoming Achievement

The results thus far are based on the average school effects among those enrolled at each school. One may wonder if our conclusions that (a) there are large school effects on a broad set of outcomes, and (b) that school effects on high-stakes exams are only weakly related to school impacts on other outcomes may change if we allowed for heterogeneity in school effects for students of different incoming achievement. We explore this possibility here, and show that all the conclusions are the same in models that account for differential school impacts by incoming preparedness.

Magnitude of School Effects by Incoming Achievement:

One may wonder if the magnitudes of school effects for the average student at the school, might overstate or understate the magnitude of school effects for particular students. For example, one might expect that school effects on dropout are larger for low-achieving students on the margin of dropout than for the average student. To assess this, where pct_i is the percentile of student i in the SEA distribution, we estimate each school's IV treatment effects (θ_j^{TOT}) while weighting each observation by $(1 + \frac{(X - pct_i)^2}{100})^{-1}$. This puts heavy weight on students with incoming scores close to the X^{th} percentile and lower weight on those far away from that percentile. This results in school effect estimates that are largely representative of students at the X^{th} percentile of the incoming achievement distribution. With these weighted estimates (in even and odd years) we compute the Maximum Likelihood estimate of the magnitude of school impacts on different outcomes for those at the bottom or top of the incoming achievement distribution.

The last three columns of [Table 2](#) report estimates of the standard deviation of persistent school effects for students at the 25th, 50th, and 75th percentiles of the incoming achievement distribution respectively. There are large persistent school effects for all outcomes at all percentiles. The magnitudes of the school effects are similar at different points of the incoming test score distribution for most outcomes, but tend to be larger at the bottom end of the distribution. This pattern of larger effects for students at the lower end of the incoming achievement distribution is most pronounced for dropout and teen motherhood. For students at the bottom of the test score distribution, going from a school at the 85th percentile to the median of the impact distribution would lead to an 11 percentage point decrease in dropout and a 7.5 percentage point reduction in teen births – much larger effects than at the average or the 75th percentile. This likely reflects the fact that students with low incoming test scores are most likely to be marginal for these two outcomes and therefore more responsive to school quality differences for these two outcomes. In sum, while we do find some evidence that school effects are larger for those at the lower end of the incoming achievement distribution (particularly for dropout and teen motherhood), we document considerable school impacts at all points in the distribution of incoming achievement.

Correlations of School Effects Across Outcomes by Incoming Achievement:

The fact that we use the average school impacts (among attendees) may also affect the reported correlations. Specifically, the low correlations for school impacts across outcomes *could* reflect students who are marginal for different outcomes attending different schools. For example, suppose only low-achieving students are marginal for high school dropout, while only high-achieving students are marginal for high-stakes exams. Also, suppose that school A only admits high achievers and B low achievers. Even if both schools have the same *potential* impacts on dropout and high-stakes exams, school A will appear to improve high-stakes exams while school B will only appear to influence dropout.³² To assess this possibility, we use the weighted value-added estimates to obtain the Maximum Likelihood estimates of the correlations of school impacts across outcomes for students at different points in the achievement distribution. By weighting the school value-added estimates at the same point in the achievement distribution *for all outcomes*, we ensure that the resulting correlation of school impacts across outcomes is not due to differences in the students being compared, but rather due to differences in school impacts (for the same set of students).

We present the Maximum Likelihood estimated correlations between the value added estimates of high-stakes and other outcomes when we center the weights at the 25th and the 75th percentiles of the SEA distribution in the middle and bottom panels of Table 3. Using weights centered on the 25th percentile does increase the correlations between high stakes test impacts and some of the other outcomes. The correlation with low stakes exams goes from 0.06 to 0.20, with teen motherhood increases from 0.07 to 0.23 and with formal employment increases slightly from 0.19 to 0.25. Using weights centered on the 75th percentile reduces the correlations between high stakes test impacts and no dropout from 0.26 to 0.16. Taken together, the results do not support the notion that these low correlations are mainly due to different schools serving students who are marginal for different outcomes. That is, even using this re-weighting, none of the correlations with high stakes exam impacts is greater than 0.5 – suggesting that the low correlations reflect different schools having impacts on different outcomes.

Overall, the patterns in Table 2 and Table 3 indicate that (a) schools have economically meaningful impacts on a range of outcomes and (b) impacts on these different dimensions are not very strongly related. This suggests that school impacts on no single outcome can serve as a “summary measure” for the quality of that school. As such, the extent to which parents choose different

³²Imagine two schools. Both schools have effect 1 on high stakes and 1 on dropout. However, only students below 60 have a dropout effect and only student above 40 have a high stakes effect. School 1 ability is uniformly distributed between 0 and 60. In this scenario, the effect on high-stakes for the average admit to school 1 is [1 for 33% of students above 40] = 0.33. Also, the effect on dropout for the average admit to school 1 is [1 for all students] = 1. In contrast, School 2 ability is uniformly distributed between 40 and 100. As such, the effect on high-stakes for the average admit to school 1 is [1 for all 100% of students] = 1. Also, the effect on dropout for the average admit to school 1 is [1 for 33% of students below 60] = 0.33. In this highly stylized example, the correlations between the effects would be -1 even though they would have the same effect had they admitted the same students.

schools for their children may have to do with the extent to which they value school impacts on different dimensions. We showed that school impacts on non-academic dimensions are economically meaningful and large. As such, the fact that parents may not choose schools that improve test scores (e.g. [Abdulkadiroğlu et al. \(2020\)](#) and [MacLeod and Urquiola \(2015\)](#)) may reflect parents choosing schools that improve other outcomes (that are weakly related to test score impacts). We explore these possibilities in Section [V](#).

V Estimating Preferences for Schools

In this section we will use the average school impact estimates and the set of secondary school choices to examine (a) the extent to which parents choose schools based on their causal impacts, and (b) explore the extent to which they choose schools with casual impacts on outcomes other than high-stakes tests. We use a modified exploded multinomial logistic model (also known as rank-ordered multinomial logit) to estimate the relationship between different school attributes (including their estimated causal impacts) and choices. As in all studies of this type (e.g. [Avery et al. 2013](#); [Burgess et al. 2015](#); [Abdulkadiroğlu et al. 2020](#); [Hastings et al. 2006](#)), we infer that parents “*value*” or “*prefer*” a school attribute if they are more likely to choose schools and rank schools more highly with that attribute. Even though our main findings are robust across several models and to inclusion of a rich set of controls (as we will show), as in other studies of parent choices (e.g. [Burgess et al. 2015](#); [Abdulkadiroğlu et al. 2020](#); [Hastings et al. 2006](#)) we cannot entirely rule out that there are unobserved determinants of parent choices that affect our results.

V.1 A Model of School Choices

For ease of exposition, we assume one parent per child. We derive the choice probability from the utility-maximizing behavior of parent $i \in N$, on behalf of student $i \in N$. Parents choose a finite number (R) of schools among all schools in the nation. Each school is indexed by $j \in J$. The utility parent i derives from student i attending each school alternative j has the following general form:

$$U_{ij} = U(X_i, Z_{ij}, \epsilon_{ij}) = \delta(X_i, Z_{ij}) + \epsilon_{ij} \quad (10)$$

where $U(\cdot)$ is the function mapping school attributes and student characteristics to utility values U_{ij} , X_i are observed student characteristics, Z_{ij} are observed school-specific attributes that may vary at the student i level (such as proximity to primary school), and ϵ_{ij} is a random error.

The school choice set is the same for all parents (i.e., $J(i) = J \forall i$), and each parent submits a single ranked-ordered list. Let $U_{ij}^{r_{is}}$ indicate the utility parent i gets from school j that they ranked in position s ($r_i = s$), so that $U_{ij}^{r_{i1}}$ is their utility for the school ranked first, $U_{ij}^{r_{i2}}$ is their utility for the school ranked second, and so on. Let $U_{ij}^{r_{i0}}$ indicate the utility parent i gets from attending school j that they did not rank. Under the algorithm used to assign students to schools, among the ranked

schools, parents have incentives to truthfully reveal their preference rankings (Haeringer and Klijn 2009; Pathak and Sönmez 2013). If parents make rational choices then:

$U_{ij}^{ria} > U_{ik}^{rib} \forall k \neq j \in J, a < b \text{ and } b \neq \emptyset$: Parent i prefers their a -ranked school over any other school k ranked below.

One can rely only on comparisons within the set of submitted choices to infer preferences about school (e.g. Avery et al. (2013) and Beuermann and Jackson (2020)). In Appendix E, we show that our main result are similar in standard rank-ordered logit models based only on the choices submitted. However, comparisons made only among chosen schools can *potentially* be misleading about particular attributes if the set of choices is not random. To see this, imagine that all parents chose four schools that are very close to home. If one were to look only within the set of schools listed, one might infer that proximity is unrelated to choices when the opposite is true. To avoid this problem, one must compare choices made (or at least one of the choices made) against *all* possible choices (e.g. Hastings et al. (2009), or make assumptions about the set of options that could have been chosen (e.g. Abdulkadiroğlu et al. (2020)). We follow the less restrictive approach that requires fewer assumptions, and compare the top choice to *all* the un-chosen schools.

When parents are unconstrained in the number of schools they can list, then the top choice is the most preferred school of all possible schools, that is $U_{ij}^{ria} > U_{ik}^{rib} \forall k \neq j \in J$ (Roth and Oliveira Sotomayor, 1992).³³ However, when the number of allowed school choices is limited, parents may act strategically so that the top listed choice is not necessarily the school they prefer. Specifically, Chade and Smith (2006) demonstrate that when the number of choices is limited, it is rational for parents to maximize the expected value of the set of choices, where the expected value of applying to a set of schools is a function of both the ex-post utility of attending the listed schools and the likelihoods of being admitted to those schools.³⁴ They point out that when listing a finite set of schools, it is rational to trade-off the ex-post utility associated with attending a school against the probability of being admitted. As such, the top choice school may not be the school with the highest ex-post utility, but rather will be the school with the highest ex-post utility *given the probability of admission*. As is proven formally in Chade and Smith (2006), if a parent's ex-post most preferred school (i.e. the school with the highest U_{ij}) is not the top choice, it must be because the probability of admission to that ex-post preferred school is too low. A useful empirical prediction from Chade and Smith (2006) is that *even with strategic choices* so long as the parents are rational, conditional on the probability of admission, the top choice school must have higher ex-post utility than any unranked school. Where p_{ij} is the probability that student i is admitted to school j , this yields

³³Many papers assume that this condition holds without testing it explicitly even when choices are constrained.

³⁴When there are finite choices, as in our setting, with a limit of up to four schools to list, rational agents will choose the portfolio of four schools that as a whole provide the greatest expected utility. Once this set of schools is decided, they will order them by ex-post utility (as above).

$U_{ij}^{r_{i1}} | p_{ij} > U_{ik}^{r_{i0}} | p_{ik} \forall k \neq j \in J$: Conditional on the admission probabilities, parent i prefers their first-ranked school over any school not in the submitted set of choices.

The expression above suggests that if one had measures of the admission probabilities for each student at each school, one could condition on these probabilities and infer ex-post preferences across all schools based on the choices. The two conditions above suggest that, where R_i is the maximum number of alternatives ranked by parent i , assuming rational choices, the probability that a parent i submits a particular ranking on over all schools is given by equation (11) below.

$$\begin{aligned} \text{Prob}[r_{i1}, r_{i2}, \dots, R_i] = & Pr[(U_{ij}^{r_{i1}} | p_{ij} > U_{ik}^{r_{i0}} | p_{ik} \forall k \neq j \in J) \\ & \cap (U_{ij}^{r_{i1}} > U_{ik}^{r_{im}}, 1 < m, \forall m \in \{2, \dots, R_i\}) \cap \dots \cap (U_{ij}^{r_{iR_i-1}} > U_{ik}^{r_{iR_i}})] \end{aligned} \quad (11)$$

V.2 Modified Exploded Multinomial Logistic Model

Equation (11) defines the likelihood of observing a set of choices as a function of parent utilities for schools and random errors. We make some assumptions on the form of U_{ij} and the distribution of ε_{ij} to use the observed choices to infer parental preferences for school attributes. Following [Hastings et al. \(2009\)](#) and [Abdulkadiroğlu et al. \(2020\)](#), we assume that the choices are rational and we parametrize δ_{ij} as a linear-in-parameters function of the school characteristics

$$U_{ij} = \beta' Z_{ij} + \varepsilon_{ij} \quad (12)$$

where β is a vector of deterministic components of school preferences. We further assume that ε_{ij} is distributed i.i.d extreme value, that is $F(\varepsilon_{ij}) = e^{-e^{(-\varepsilon_{ij})}}$. Under this standard distributional assumption (see [Train \(2009\)](#) and [McFadden \(1973\)](#)), the probability that a parent i submits a particular ranking on over all schools (i.e. Equation (11)) is simply a product of standard logit formulas. That is, where p_{ij} is the probability of admission for student i to school j , and parameter vector $\beta = [\beta_1, \pi]$, the probability that parent i chooses the ranking $\{r_{i1}, r_{i2}, \dots, R_i\}$ is:

$$\begin{aligned} \text{Prob}[r_{i1}, r_{i2}, \dots, R_i] = & \frac{\exp(\beta_1' Z_{ij}^{r_{i1}} + \pi p_{ij})}{\exp(\beta_1' Z_{ij}^{r_{i1}} + \pi p_{ij}) + \sum_{k=1}^{J-R_i} \exp(\beta_1' Z_{ik}^{r_{i0}} + \pi p_{ik})} \\ & \cdot \frac{\exp(\beta_1' Z_{ij}^{r_{i1}})}{\sum_{k=1}^{R_i} \exp(\beta_1' Z_{ik}^{r_{ik}})} \cdots \frac{\exp(\beta_1' Z_{ij}^{r_{iR_i-1}})}{\exp(\beta_1' Z_{ij}^{r_{iR_i-1}}) + \exp(\beta_1' Z_{ik}^{r_{iR_i}})} \end{aligned}$$

Accordingly, the log likelihood of observing all the choices lists can be written as:

$$\log L(\beta) = \sum_{i=1}^N \log l_i(\beta) = \sum_{i=1}^N \log (\text{Prob}[r_{i1}, r_{i2}, \dots, R_i]). \quad (13)$$

One can obtain estimated preferences for school attributes β_k by estimating this model by maximum likelihood (i.e. finding the β vector that maximizes this expression).

Our model is conceptually similar to others in the literature but there are two key differences. First, we include an additional choice to the standard exploded logit model: In our first pseudo-observation, the individual chooses her first-ranked school over the set of *all* unranked schools in Trinidad and Tobago. As discussed above, including this additional first pseudo-observation allows us to anchor each individual’s choices to a common set of schools for all parents – making the choices and preferences comparable across individuals. The second key difference is that, as informed by the theory, when comparing the top choice to all unranked choices, we include the admission probability as a covariate in the model, but we do not include it when comparing schools within the chosen list.³⁵ These two modifications to the conventional multinomial logit model allows us to anchor each individual’s choice set while also explicitly accounting for strategic behaviors.³⁶ Our main conclusions are the same in more restrictive models that do not include the additional pseudo-observation or account for admission probabilities (See [Appendix E](#)).

V.3 Estimating Admission Probabilities

When we compare the top choice school to all un-chosen schools, we account for the probability that student i would have been assigned to each school j had they applied. In many research settings, this probability is difficult to uncover. Fortunately, because we have many years of admissions data and students are assigned to schools based on a known algorithm, we can approximate this probability with the historical likelihood that student i would have scored above the cutoff for school j given their own incoming SEA score. We report these assignment probabilities for four different schools by the percentile of incoming SEA scores in [Figure 5](#). School 9 is a very selective secondary school. Across all years, no student below the 82nd percentile scored above the cutoff for that school and all students above the 92nd percentile did. We show similar figures for less selective schools in other panels. Depending on their incoming SEA score, students may be marginal admits for some schools (predicted probabilities greater than zero and less than 1), be virtually guaranteed assignment at some schools, and have virtually no chance at others.

Because there is uncertainty for any student regarding their exact score, we compute the likelihood that student i is within “striking distance” of a given cutoff as follows: For each school in each year and each SEA score, we code the “rough” likelihood as 0 if the score was more than 5

³⁵Our main results are robust to including the admission probabilities in all comparisons. Also, our main results are robust to excluding the admission probabilities entirely.

³⁶Our model differs from [Hastings et al. \(2005\)](#) and [Hastings et al. \(2006\)](#) in that it uses a version of the exploded logit model with fixed coefficients, instead of estimating random coefficients by using mixed logit utility models. [Abdulkadiroğlu et al. \(2020\)](#) use the rank-ordered multinomial logit model to estimate a single measure of each school’s popularity separately for different covariate cells, whereas we use the modified version of the same model to estimate average population preferences for different school attributes.

percentile points below the cutoff; 0.5 if it was within 5 percentiles of the cutoff; and 1 if it was above the cutoff by more than 5 percentiles. We then compute the probability of student i with a particular SEA score being assigned to school j (p_{ij}) as the average of these “rough” likelihoods for that SEA score across all years (excluding the year when the student actually applied). We also show these probabilities for the selected schools in [Figure 5](#). So long as students are somewhat aware of these relationships based on historical precedent, our estimated probabilities proxy for the real admission probabilities used when making choices.

V.4 Choice Parameter Estimates

We examine whether parents express preferences for schools based on their impacts on academic and non-academic dimensions, above and beyond easily observed school attributes. We use the average school impacts in our choice model as predictors of choices. This is motivated by the fact that, for all outcomes except dropout, the average estimated school impacts are very highly correlated with the school impacts estimated for those at the bottom and the top of the incoming achievement distribution ([Table 4](#)). As such, the average school estimates represent meaningful differences in school impacts for all incoming achievement levels, and reflect real differences in school impacts that parents may value. **It is worth noting** that insofar as the average school impacts are inaccurate measures of school impacts for particular kinds of students (i.e., there are large match effects in unobserved dimensions), it would bias our results toward zero – making it *less* likely to find any association between choices and estimated school impacts. As such, *even if there were considerable match effects in dimensions other than incoming achievement*, any systematic relationships we find between choices and school impacts would reflect real relationships.

Our full estimation sample includes 312,420 households making school choice decisions. We estimate choice models separately for each (SEA score ventile) \times (gender) cell to allow preferences to vary based on the student’s gender and incoming achievement. Estimated standard errors are adjusted for clustering at the school-district level. Because the point estimates of the modified exploded multinomial logistic model are not easily interpretable, we report on the relative magnitudes and statistical significance of the estimated coefficients. Except for the natural log of distance to school, all attributes have been standardized to be mean zero and unit variance. We present results from two main models; (a) a Impacts Only Model (which includes schools’ causal impact estimates for all outcomes, peer quality, and log distance), and (b) a Full Model that includes schools’ causal impact estimates for all outcomes, the school-level averages for all the outcomes, peer quality, and log distance. All specifications include control variables for whether the secondary school is on the same island, whether it is all-girls, whether it is all-boys, and the estimated “rough” likelihoods of school admission (but only when comparing the top choice to all unranked schools).

Before discussing the relationship between school impacts and school choices, we investigate

the importance of proximity, peer quality, and admission probabilities. The coefficient estimates on distance for each cell are presented in the top panel of [Figure 6](#). This figure reveals three key patterns. First, all students rank closer schools more highly - for all cells the estimate on distance is negative (with p -value < 0.01). Second, parents of lower-achieving students are more responsive to distance than those of higher-achieving students. Note that the patterns for the impacts only models (left) and the full models (right) are largely the same – indicating that distance to school is largely unrelated to other school attributes. Third, while parent choices for girls are somewhat less sensitive to distance than that for choices for boys, the choices of parents of boys and girls are similar within each achievement group (as evidenced by the overlapping confidence intervals).

A second key attribute when choosing a potential school is the average peer academic quality ([Hastings et al. 2005](#); [Hastings et al. 2006](#); [Hastings and Weinstein 2008](#)). The middle panel of [Figure 6](#) shows the coefficients on the potential peers’ academic quality (the average SEA score of the incoming cohort). In the impacts only model (left), one rejects that choices are unrelated to peer achievement for almost every cell at the 5 percent level. This figure also reveals that parents of girls are less responsive to peer quality than parents of boys. This gender difference is driven by the fact that low-achieving boys and girls have similar responsiveness to peer quality, but higher-ability boys are more responsive to peer quality than low-achieving boys (while there is a small achievement gradient for girls). Estimated coefficients from the full model (right) fall by about 40 and 30 percent for females and males, respectively – suggesting that parents’ apparent preferences for better peers’ outcomes may partially reflect preferences for better average outcomes.

The importance of proximity and peer quality in shaping the schooling decision is consistent with other studies (e.g., [Hastings et al. 2005](#); [Abdulkadiroğlu et al. 2020](#)). A comparison of the coefficients on peer quality and proximity implies that, on average, increasing peer quality by about 0.71 standard deviations (roughly the difference between a student’s top choice school and the third choice school) is associated with the same difference in choices as doubling the distance between the primary and secondary school. That is, the choice parameters imply that parents may be willing to travel about 2.4 times farther to attend a secondary school with one standard deviation higher incoming peer scores. The average distance is about 6 kilometers, so that this amounts to travelling an *additional* 8.4 kilometers (or 5.5 miles) to attend a school with one standard deviation higher peer achievement. While this is true on average, there is heterogeneity. Point estimates suggest that parents of males and females in the bottom decile of the test score distribution may be willing to increase travel distance by about 50 percent to attend a school with 1 standard deviation higher peer quality. In contrast, point estimates imply that females in the top decile may be willing to double their distance to attend a school with one standard deviation higher-achieving peers, and males in the top decile would be willing to increase their distance more than threefold.

As discussed in [Section V.3](#), a key conditioning variable in our analysis is the admission prob-

ability. The bottom panel of [Figure 6](#) shows the coefficients on the admission probability under the two models. More desirable schools (for both observed and unobserved reasons) are those with lower admission probabilities, by construction. As such, a negative coefficient on admission probability would indicate that there are unaccounted-for school attributes that are negatively correlated with the admission probability. However, if the school attributes included accurately reflect those dimensions of school quality that parents value, the coefficient on admission probability should be positive. In the impacts only model the point estimates for all groups are positive – that is, conditional on peer quality, proximity and impacts on key outcomes, parents are more likely to choose schools to which their child is more likely to be admitted. In the full model, as expected, the coefficient on the admission probability becomes more positive (particularly for low-achieving students) – indicating that schools that are more desirable tend to have better average outcomes. The fact that the coefficient on the admission probability is positive for all groups provides empirical validation of the theoretical predictions from [Chade and Smith \(2006\)](#), and suggests that our included variables capture much of the determinants of parents’ school choices.

V.4.1 Academic Outcomes

Parents in Trinidad and Tobago make similar choices to parents in other settings and appear to make rational choices. We now turn to the importance of schools’ causal impacts. [Figures 7 through 12](#) plot the estimates separately for each outcome. On the left panels, we plot coefficients on schools causal impacts from the impacts only model. In the middle panels, we plot coefficients on schools causal impacts from the full model (that also includes school level averages). In the right panels, we plot the coefficients on school level average outcomes from the full model.

We first discuss preferences for school impacts on high-stakes exams. [Figure 7](#) reveals three patterns: (1) School impacts on high stakes exams influence the choices of many parents, but there is considerable heterogeneity in this pattern, (2) The choices of parents of high-achieving students are more strongly related to school impacts on high-stakes exams than those of parents of lower-achieving students, and (3) at all levels of incoming achievement, the choices of girls’ parents are more strongly related to school impacts on high-stakes exams than that of parents of boys. These patterns are illustrated by the positive and significant relationship between the individual’s score percentile and the coefficient magnitude (which is more pronounced for girls) in the left panel of [Figure 7](#). While one can reject that the choices of parents of higher-achieving students are unrelated to school impacts on high-stakes exams ($p\text{-value} < 0.001$), the figure also reveals that parents of children with low incoming scores *may not* prefer schools that raise high-stake tests. Among those in the bottom third of the incoming test-score distribution, parents are no more likely to list a top choice school with higher high-stakes test-score impact – in fact the point estimates are slightly negative. To put these point estimates in perspective, we compare the coefficients on high-stakes impacts to that of the natural log of distance. For those in the top decile of incoming

achievement, the estimates imply that parents of girls (conditional on average peer scores) may be willing to travel more than three times farther, or about 18 kilometers farther, to attend a secondary school at the 85th percentile of the high-stakes impact distribution than one at the median. Parents of boys in the top decile of incoming scores, would be willing to travel about twice as far to attend a secondary school at the 85th percentile of the high-stakes test scores effectiveness distribution than one at the median. While not as responsive as parents of high-achieving girls, the choices of parents of high-achieving boys are very strongly related to school impacts on high-stakes exams.

As we control for school averages (i.e average exam scores and that for other outcomes), the importance of high-stakes exam impacts remains largely unchanged for girls and is somewhat weaker for boys (Figure 7, middle). That is, even conditioning on peer quality and average high stakes exams, the coefficient on high-stakes impacts remains positive and highly statistically significantly different from zero for parents of high-achieving children— suggesting that these parents, on average, can distinguish causal effects from selection. This result stands in contrast to [Abdulkadiroğlu et al. \(2020\)](#) who find that after conditioning on peer quality, parents do not appear to value school impacts on school exit examination performance. We discuss possible reasons for these differences in Section V.6. Looking at the school average high-stakes scores, there is little evidence that this influences choices in the full model. It is important to keep in mind that these estimates are all conditional on average incoming scores – indeed, in models without peer quality the coefficients on average high-stakes scores are large and positive. As such, our results are not inconsistent with [Hastings et al. \(2006\)](#) who find that parents value schools with better average outcomes, or with [MacLeod and Urquiola \(2019\)](#) who argue that parents may value schools with better average outcomes. However, our results do suggest that (at least in our setting) parents being more likely to choose high-achieving schools may largely reflect a preference for higher-achieving peers.

Our next academic outcome is low-stakes exams. While there is strong evidence that certain parents may prefer schools that raise high-stakes exam performance, there is little evidence that parents value schools that improve low-stakes exam performance.³⁷ Specifically, the results (Figure 8, left and middle) show insignificant estimates across the entire incoming achievement distribution for both boys and girls. Moreover, these point estimates are much smaller in magnitude than those for high-stakes exam impacts – further evidence that parents school choices are sensitive to schools’ high-stakes impacts but not to their low-stakes impacts. Looking at average outcomes, parent choices are largely unrelated to better average performance on the low-stakes exam (conditional on the averages of the other outcomes) (Figure 8, right). In Trinidad and Tobago, average school outcomes on high-stakes exams are made public, while average school outcomes on low-stakes

³⁷ An insignificant or small estimated coefficient could indicate that either parents don’t value that particular school attribute or, alternatively, that parents care about it but they don’t have enough information about it. We favor the interpretation that an insignificant school feature does not play an important role in the schooling decision, remaining agnostic about which reason is more likely to occur in each particular case.

exams are not. As such, the results are consistent with parents discerning school impacts on high-stakes exams but not on low-stakes exams. They are also consistent with such parents not caring about school impacts on low-stakes tests precisely because they are low stakes.

Our final academic outcome is dropout. As with low-stakes exams, the patterns (Figure 9, left and middle panels) suggest that parents are no more likely to choose schools that causally reduce dropout. For none of the gender-by-achievement cells can one reject that the coefficient on dropout impacts is zero. However, looking at dropout *rates*, the patterns suggest that low-achieving male and female students are more likely to choose schools with low average dropout rates. See the positive and significant estimated coefficients at the bottom of the incoming achievement distribution in the right panel. For females, the point estimates are positive for most of the achievement distribution and one can reject that the estimates are zero for females in the bottom third of the incoming achievement distribution. While one cannot reject that parents of males' school choices are unrelated to dropout rates (conditional on the other average outcomes), the lowest-achieving males may prefer schools with lower dropout rates. As with the low-stakes exams, school-level dropout rates are not publicly reported. Moreover, unlike outcomes such as teen motherhood or arrests which would be visible to classmates, school dropouts are absent and by their very nature difficult to observe. As such, school dropout rates and school impacts on dropout may be particularly difficult for parents to observe and therefore respond to.

V.4.2 Non-Academic Outcomes

Next we examine how parents take into account school impacts on non-academic outcomes when making school choices. Recall that all variables are coded so that positive values indicate better outcomes. We start with teen motherhood. The patterns in Figure 10 reveal little association between school choices and impacts on teen motherhood. Indeed, in the impacts only model the 95 percent confidence interval for all of the estimates include zero. In the full model that also includes the teen motherhood rate (and the averages for all other outcomes), some of the point estimates for females are negative and significantly different from zero at the 5 percent level. However, these point estimates are modest compared to those for high-stakes value added, and most of the point estimates cannot be distinguished from zero.

While the evidence on parental preferences for teen motherhood impacts is mixed, parents *are* much more likely to choose schools with low average teen motherhood rates. Between two schools with the same average academic performance and impacts on academic performance, parents prefer schools with lower average teen motherhood rates (Figure 10, right). For most cells, the estimated coefficients on average teen motherhood outcomes are statistically significant, and their magnitudes are much larger than the average high-stakes exam performance coefficient, indicating that parents may place greater value on the school's average prevalence of high-risk adolescent behavior than on

the average academic performance of the school. The notion that parents may value school safety (or low risk schools) is not new, but this is the one of the first studies to document this rigorously in a discrete-choice framework. The point estimates imply that the average parent would be willing to increase their distance by about 53 percent to send their child to a school that was at the 85th percentile of the average (non) teen motherhood distribution versus one at the median. These choice patterns are even stronger for parents of high-achieving males. The estimates imply that parents of males in the top decile of the achievement distribution may be willing to more than double their distance to attend a school at the 85th percentile of the (non) teen motherhood distribution versus one at the median. In sum, while there is little evidence that parents choose schools that have larger impacts on teen motherhood (perhaps because they cannot observe or infer it), they do choose schools that have low motherhood rates.

Another important non-academic measure is teen arrests. Unlike teen motherhood, the results in [Figure 11](#) reveal that parent choices are related to school impacts on teen arrests. This is driven largely by parents of males and parents of high-achieving students (left panel). Males of almost all incoming achievement levels choose schools that reduce crime, and this behaviour is stronger for the higher-achieving males. One can reject that this relationship is zero for males above the 40th percentile of the incoming achievement distribution. In contrast, while the parents of higher-achieving females are more likely to choose schools that reduce crime, one cannot reject null impacts for girls in most achievement groups. Given that most arrests are of males, the fact that parents of males are more responsive to school impacts on arrests than girls makes sense. Overall, parents of the highest-achieving students appear to have the strongest preference for schools that causally reduce crime. Point estimates suggest that parents of both boys and girls in the top decile may be willing to increase their distance by about 130 percent (roughly 10 kilometers) to send their child to a school that was at the 85th percentile of the (non) teen arrest impact distribution versus one at the median. In fact, a comparison of the estimates for high-stakes exam impacts reveal that for the parents of males below the 85th percentile, impacts on arrests matter more for choices than impacts on high-stakes exams. This basic pattern is similar, but slightly attenuated, after controlling for average arrest rates at the school ([Figure 11](#), middle). Similar to parental choices regarding low-risk schools, parents of all achievement levels choose schools with low teen arrest rates ([Figure 11](#), right). This preference is somewhat larger for parents of females. These results, in conjunction with those for teen motherhood, suggest that parents prefer schools with a lower prevalence of risky behaviors (conditional on peer achievement and average test score outcomes). The results also indicate that parents may value schools that reduce these behaviours (i.e., arrests) in their own children, even conditional on school impacts on academics and peer quality.

The last outcome we examine is formal employment at ages 27 and older. All parents choose schools with positive causal impacts on employment ([Figure 12](#), left). The point estimates indicate

that, on average, parents would be willing to increase their distance by about 30 percent to send their child to a school that was at the 85th percentile of the employment impacts distribution versus one at the median. These patterns are noteworthy for two reasons. First, this is the first direct demonstration that parents choose (i.e., prefer) schools that have causal impacts on formal employment (above and beyond peer quality and impacts on academic outcomes). Second, even though choices for children in the bottom half of the incoming achievement distribution are largely unrelated to school impacts on high-stakes test scores, they *are* strongly related to school impacts on formal adult employment – suggesting that for more than half of the population, school impacts on employment may matter more than impacts on high-stakes exams. The estimates on employment impacts are largely similar in models that include school average labor market participation rates (Figure 12, middle). As with the other dimensions, the right panel of Figure 12 reveals that parents also choose schools with higher average employment. Estimates suggest that, on average, parents would be willing to increase their distance by about 50 percent to send their child to a school at the 85th percentile of the formal employment distribution versus one at the median. Interestingly, the implied preference for schools with higher average employment rates is larger for parents of children at the lower end of the incoming achievement distribution.

In sum, we show that parents choose, and therefore likely value, schools that have higher causal impacts on certain academic and non-academic outcomes. We show that this is not simply due to parents choosing schools with better average outcomes or better peers. Also, consistent with school quality being multidimensional, parents choose schools that have causal impacts on outcomes other than high-stakes tests such as crime and formal labor-market participation. Importantly, the correlations between school impacts on high stakes exams and impacts on arrests and formal labor market participation are relatively low. This suggests that strong parental preferences for school impacts on non-academic outcomes (that are largely unrelated to test score impacts) are a plausible explanation for the weak link between parental preferences and school impacts on test scores.³⁸

Even though we do not find that parents choose schools with impacts on all outcomes, we do find evidence that parents choose schools with lower dropout rates, lower teen arrest rates, lower teen motherhood rates and higher future labor market participation. It is important to note that valuing schools' average outcomes (but not causal impacts) is not necessarily irrational, because

³⁸ All of our models use the 2SLS estimated school impacts as explanatory variables across all years (i.e., across as much as 17 years). Because the choice year is included when forming this estimate, one may worry about mechanical correlation between the estimated impacts and the desirability of the school. Because our school effects are based on several years of data, and we use quasi-random variation for identification, we believe that this is unlikely. However, to assuage this concern, we estimate our choice models using leave-year-out 2SLS estimates. Because the 2SLS estimates are based on 133 excluded instruments, leave-year-out estimates can vary a lot for the same school from year to year. As such, while the leave-year-out estimates remove potential mechanical biases, they also introduce non-trivial estimation errors. As one might expect, this introduces additional noise to our estimates, but as one can see in Appendix D, our results using both leave-year-out estimated impacts and leave-year-out school average outcomes are qualitatively similar, and our central conclusions unchanged.

schools with higher average outcomes may confer benefits to students that are not measured in our data.³⁹ Overall, parents appear to be relatively sophisticated in their understanding of school quality.

V.5 Are Differential Preferences by Incoming Scores Due to Match Effects?

The choice models results indicate that parents of high- and low-scoring children make different choices. This could be due to parents (a) having different preferences regarding school quality, or (b) school quality having different effects on children by incoming achievement – a form of match effect. To assess this, we explore the extent to which the school effects estimates for those at the 25th, median, and 75th percentiles of the incoming achievement distribution, differ from that of the average student. Specifically, we estimate school effects for even and odd years based on regressions that weight toward the 25th, median and 75th percentiles of the incoming achievement (as outlined in Section IV.1). We then implement the same maximum likelihood procedure outlined in Section IV.1 to obtain correlations that are not attenuated by estimation errors. If the differences in parent choices are due to match effects, then the correlation between the school effects for different students (classified according to their incoming achievement) will be very different. We examine this for the three outcomes that appear to influence parent choices (high stake exams, teen arrests, and formal labor market participation). In Table 4, the correlations between the average effects and those at the 25th, 50th and 75th percentiles are all above 0.98 for high stakes exams, and all above 0.91 for arrests and formal labor market participation. Insofar as the correlations are not perfect, we cannot rule out the existence of match effects. However, given the very high correlations (all above 0.9) between effects for high- and low scoring-students, we are confident that the differences in choices we document do not reflect match effects. That is we can rule out that our key results are driven by test score impacts having larger marginal effects for high-achievement children while labor market or crime impacts having larger effects on low-achieving children. Rather, our results likely reflect differences in preferences (or differences in information).

V.6 Discussion of Parental Preference Results

One of our key findings is that parents may value school impacts on multiple outcomes above and beyond peer quality and average outcomes. The only other paper to formally test this notion is Abdulkadiroğlu et al. (2020) who find that parents do prefer schools that improve academic outcomes, but not after controlling for peer quality. Our results are a nice counterpoint to their work because we demonstrate that context matters. Also, by moving beyond academic outcomes and examining parental preferences for non-academic outcomes such as crime, teen fertility, and labor market participation, we shed light on the extent to which parents value school impacts beyond

³⁹Indeed, MacLeod and Urquiola (2019) present a model that rationalizes preferences for schools' absolute achievement when this attribute serves as a signal that improves labor market matching.

academics – this is very important given that many school choice evaluations use test scores alone.

Another potential explanation for differences between our findings and [Abdulkadiroğlu et al. \(2020\)](#) is market size. Several studies show that when individuals are faced with too many options they often opt for simplicity (e.g., [Iyengar and Kamenica \(2010\)](#)), are more likely to rely on heuristics (e.g., [Besedeş et al. \(2012\)](#)) and less likely to make the optimal choice (e.g., [Schram and Sonnemans \(2011\)](#)). [Abdulkadiroğlu et al. \(2020\)](#) examine parent choices in the largest school district in the United States (which offers over 700 programs at over 400 schools). Their setting is a context in which sub-optimal behaviors are most likely to occur. In contrast, in our setting, individuals choose from a set of 158 options. While this is by no means a small market, it is much smaller than New York City (as are most markets), and therefore individuals’ choices are less likely to be subject to errors induced by “overchoice.”

Our finding that only parents of high-achieving students are able to discern school impacts on high-stakes examinations relates to the overall lack of robust achievement effects, on average, of attending schools that parents prefer ([Beuermann and Jackson, 2020](#)). However, in the Trinidad and Tobago context, school impacts may be easier to infer for relatively sophisticated parents. Average incoming scores are well known and publicly reported. Additionally, school averages for the high-stakes exams are also reported at the school level. As such, it is plausible for a relatively sophisticated parent to observe schools with similar average outcomes and infer which one likely has larger impacts (based on average incoming test scores). In settings where average incoming scores are not reported or well known, this calculation may be much more difficult to conduct – offering another plausible explanation for our finding that some parents (i.e., those of higher achieving students) can discern school impacts (conditional on average outcomes) while some other studies do not find so. However, the fact that parents of high-achieving students value schools that reduce arrests, and all parents value schools that raise employment (even conditional on school averages) suggests that in some instances parents can discern school impacts even when information is imperfect (perhaps through some combination of knowing the incoming student characteristics and reputation effects regarding average outcomes).

VI Conclusions and Policy Implications

Individual schools have meaningful causal effects on an array of outcomes; these include test scores in low-stakes exams (both academic and non-academic subjects), dropout by age 14, teen motherhood, performance on high-stakes school leaving exams, being arrested, and formal labor market participation. However, consistent with school quality being multidimensional, the correlations between school impacts on high-stakes tests and other outcomes is surprisingly low. From a policy perspective, our results suggest that school impacts on test scores may not be the best measure of a school’s impacts on longer-run outcomes. Accordingly, policymakers should be cautious

(and thoughtful) regarding using test score impacts in accountability systems and incentive pay schemes and may wish to adopt a more holistic view of school quality.

When we link these causal estimates to choice data we find that parents choose schools that have larger positive impacts on high-stakes tests. However, they also choose schools that decrease crime and increase labor market participation. Importantly, parents seem to value schools that causally improve outcomes above and beyond average school outcomes and peer quality. These results suggest that parents may be using reasonable measures of school quality when making investment decisions for their children – a key requirement for the potential benefits of school choice (Friedman, 1955). The fact that parents do not *only* choose schools that improve academics but also those that improve non-academic and longer-run outcomes suggests that the benefits to school choice may extend to a wide range of outcomes (not just test scores). This result provides a plausible explanation for the fact that parental preferences for schools are not strongly related to school’s test score impacts (MacLeod and Urquiola, 2019). It also suggests that policy evaluations based solely on test scores may be very misleading about the effects of school choice on welfare.

While we find that parents choose, and therefore likely value, positive school impacts, we do find heterogeneity. Parents value impacts on high-stakes exams, reduced teen crime, and formal adult employment. However, parents of high-achieving children are most responsive to schools’ causal impacts. This pattern has important distributional implications because it suggests that high-achieving children may benefit more from school choice – exacerbating any pre-existing inequalities among children (in both academic and nonacademic domains). It also suggests that the market forces that may drive efficient competition among schools may be much weaker (or even non-existent) for schools serving largely low-achieving student populations. If these differences across parents reflect differences in information, there may be value to the provision of information to parents regarding the causal impacts of schools (as opposed to school averages) on a wide array of academic outcomes (such as high-stakes test scores and school dropout) and nonacademic outcomes (such as teen motherhood and crime).⁴⁰ The provision of such information may improve the decisions of all parents (not just those of low-achieving children) and could increase the *potential* allocative efficiencies and competitive benefits of school choice.

⁴⁰A recent experimental study in Chile (Allende et al. 2019) find that a personalized information intervention led parents to chose schools with higher causal impacts which in turn improved their children’s academic achievement.

Tables and Figures

Table 1: Summary Statistics

	All Schools	Male	Female	Above median	Below median
	(1)	(2)	(3)	(4)	(5)
<i>Panel A: SEA data (cohorts: 1995 - 2012)</i>					
Female (%)	50.38 (50.00)			53.52 (49.88)	47.24 (49.92)
Admitted cohort size	217.72 (165.11)	220.55 (163.97)	215.06 (166.12)	220.55 (163.97)	215.06 (166.12)
Standardized SEA score	0.00 (1.00)	-0.13 (1.04)	0.12 (0.94)	0.70 (0.62)	-0.70 (0.79)
Individuals	312,420	155,017	157,403	156,168	156,252
<i>Panel B: NCSE data (linked to SEA cohorts: 2006 - 2012)</i>					
Took NCSE (%)	90.48 (29.34)	88.72 (31.64)	92.26 (26.72)	95.27 (21.22)	83.24 (37.35)
Standardized NCSE score	0.00 (1.00)	-0.23 (1.00)	0.22 (0.95)	0.31 (0.89)	-0.55 (0.95)
Individuals	108,097	54,209	53,888	65,084	43,013
<i>Panel C: CSEC data (linked to SEA cohorts: 1995 - 2011)</i>					
Took at least 1 subject (%)	77.58 (41.70)	72.24 (44.78)	82.85 (37.70)	89.27 (30.95)	66.13 (47.33)
Number of subjects passed	3.28 (3.11)	2.66 (2.98)	3.89 (3.11)	4.93 (2.94)	1.66 (2.31)
Qualified for tertiary (%) *	36.00 (48.00)	28.90 (45.33)	42.97 (49.50)	59.75 (49.04)	12.70 (33.30)
Individuals	296,838	147,212	149,626	146,950	149,888
<i>Panel D: CAPE data (linked to SEA cohorts: 1999 - 2009)</i>					
Took at least 1 unit (%)	20.58 (40.43)	16.01 (36.67)	25.10 (43.36)	34.84 (47.65)	4.65 (21.05)
Number of units passed	1.45 (2.98)	1.11 (2.66)	1.79 (3.22)	2.50 (3.58)	0.28 (1.38)
Earned Associate Degree (%)	15.35 (36.05)	11.38 (31.75)	19.28 (39.45)	26.85 (44.32)	2.50 (15.60)
Earned scholarship (%)	0.99 (9.92)	0.69 (8.30)	1.29 (11.29)	1.88 (13.57)	0.01 (0.86)
Individuals	198,762	98,810	99,952	104,897	93,865
<i>Panel E: Criminal records (linked to SEA cohorts: 1995 - 2010) - in percent</i>					
Arrested by 18	3.28 (17.82)	5.81 (23.39)	0.80 (8.89)	1.73 (13.04)	4.77 (21.31)
Individuals	281,578	139,581	141,997	137,693	143,885
<i>Panel F: Birth records (linked to SEA cohorts: 2004 - 2010) - in percent</i>					
Live birth by 19			9.88 (29.84)	6.49 (24.64)	14.97 (35.68)
Individuals			42,167	25,337	16,830
<i>Panel G: Labor market data (linked to SEA cohorts: 1995 - 2002) - in percent</i>					
Formally employed	75.67 (42.91)	79.58 (40.31)	71.87 (44.96)	78.35 (41.19)	73.79 (43.98)
Individuals	149,238	73,504	75,734	61,377	87,861

Notes: Standard deviations reported in parentheses below the means. *Qualification for tertiary education requires passing five CSEC examinations including English language and mathematics. Columns (4) and (5) report statistics differentiated by the rank of the assigned school based on the SEA score mean of students assigned to each school.

Table 2: Standard Deviation of Persistent School Impacts

				School Level ($\sigma_{\theta_j^{TOT}}$)		
Outcome	Size of Impact	[95% CI]	Size of impacts for Students at the			
			25th %ile of the achievement distribution	median of the achievement distribution	75th %ile of the achievement distribution	
<i>Standardized outcomes</i>						
High-Stakes Index	0.393	0.347	0.446	0.423	0.454	0.384
Low-Stakes Index	0.412	0.359	0.474	0.555	0.485	0.382
No Dropout by 14	0.114	0.097	0.135	0.361	0.131	0.091
Not arrested by 18	0.089	0.076	0.103	0.124	0.111	0.078
No live birth by 19	0.115	0.067	0.197	0.248	0.129	0.137
Formally employed 27+	0.082	0.067	0.100	0.095	0.094	0.104
<i>Binary outcomes</i>						
No Dropout by 14	0.036	0.030	0.042	0.113	0.041	0.028
Not arrested by 18	0.016	0.014	0.019	0.023	0.020	0.014
No live birth by 19	0.035	0.020	0.060	0.075	0.039	0.042
Formally employed 27+	0.036	0.029	0.044	0.041	0.041	0.045

Notes: The table reports estimates of the standard deviation of the persistent attended school impacts along with their respective 95 percent confidence intervals obtained from the maximum likelihood approach. We removed students who attended schools with outlier estimated impacts (i.e. beyond 4σ of the median school). The last 3 columns show results when using school impacts estimated with weights centered around different percentiles of the achievement distribution. We do this using $weight_i = (1 + \frac{(X - pct_i)^2}{100})^{-1}$, where $X = 25, 50, 75$ and pct_i is the student's percentile in the achievement distribution.

Table 3: Maximum Likelihood Correlations Between School Impacts

	High-Stakes Index	Low-Stakes Index	No Dropout by 14	Not arrested by 18	No live birth by 19	Formally employed 27+
<i>Unweighted</i>						
High-Stakes Index	1.00					
Low-Stakes Index	0.06	1.00				
No Dropout by 14	0.26	0.23	1.00			
Not arrested by 18	0.47	-0.03	0.43	1.00		
No live birth by 19	0.13	-0.03	-0.20	-0.51	1.00	
Formally employed 27+	0.19	0.00	-0.01	0.21	-0.16	1.00
<i>Weighted around 25th percentile of the SEA score</i>						
High-Stakes Index	1.00					
Low-Stakes Index	0.20	1.00				
No Dropout by 14	-0.07	0.38	1.00			
Not arrested by 18	0.34	0.22	0.12	1.00		
No live birth by 19	0.23	0.26	-0.20	-0.10	1.00	
Formally employed 27+	0.25	0.15	-0.10	0.13	0.03	1.00
<i>Weighted around 75th percentile of SEA score</i>						
High-Stakes Index	1.00					
Low-Stakes Index	0.13	1.00				
No Dropout by 14	0.16	-0.28	1.00			
Not arrested by 18	0.50	-0.15	0.28	1.00		
No live birth by 19	0.06	-0.08	0.26	-0.39	1.00	
Formally employed 27+	0.20	0.17	0.06	0.25	-0.37	1.00

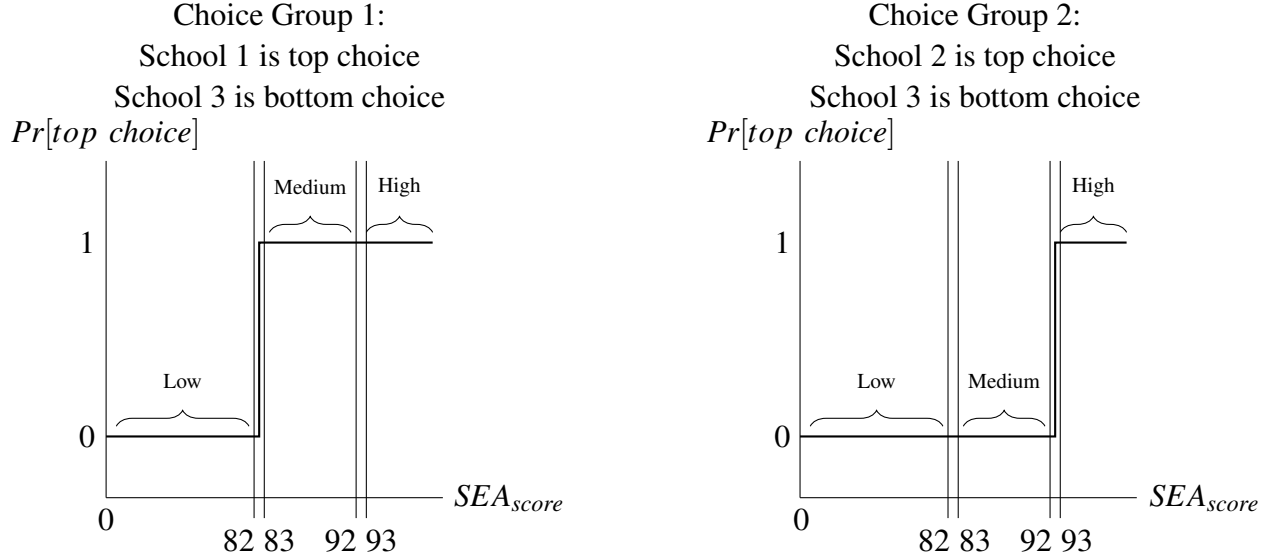
Notes: The table reports correlations of persistent school effects across different outcomes. Correlations were computed using the maximum likelihood approach described in the text. Correlations in the top panel were computed with estimated school impacts for the average student and those in the middle and bottom panels were computed using school impacts with weights centered around different percentiles of the achievement distribution. We do this using $weight_i = (1 + \frac{(X - pct_i)^2}{100})^{-1}$, where $X = 25, 75$ and pct_i is the student's percentile in the achievement distribution. We removed students who attended schools with outlier estimated impacts (i.e. beyond 4σ of the median school).

Table 4: Maximum Likelihood Correlations Using Different Weights

	Correlation between average school impacts and school impacts weighted around the		
	25th %ile of the achievement distribution	median of the achievement distribution	75th %ile of the achievement distribution
High-Stakes Index	1.00	0.99	0.98
Low-Stakes Index	0.95	0.96	0.98
No Dropout by 14	0.66	0.80	0.53
Not arrested by 18	0.91	0.94	0.96
No live birth by 19	0.73	0.90	0.91
Formally employed 27+	0.97	0.91	0.97

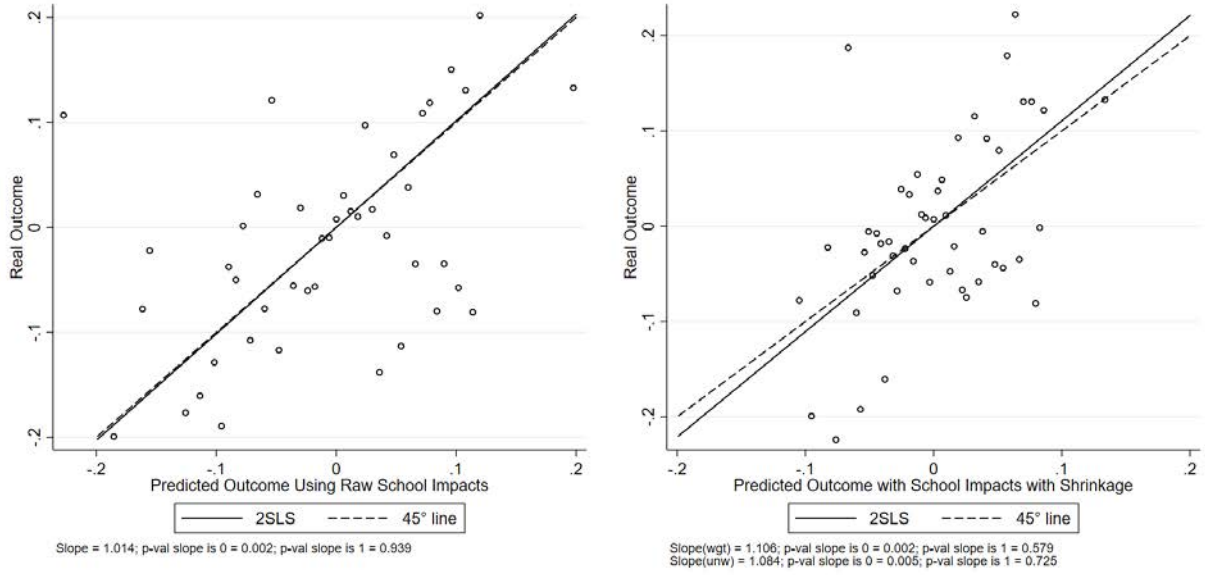
Notes: The table reports correlations of persistent school effects for each outcome across the different weighted estimates. Correlations were computed using the maximum likelihood approach described in the text. We show correlations between estimated school effects for the average student and those weighted around the 25th percentile, the median, and the 75th percentile of the students' SEA score distribution. Weighted school impacts are estimated using $weight_i = (1 + \frac{(X - pct_i)^2}{100})^{-1}$, where $X = 25, 50, 75$ and pct_i is the student's percentile in the achievement distribution. We removed students who attended schools with outlier estimated impacts (i.e. beyond 4σ of the median school).

Figure 1: Exemplar of Variation



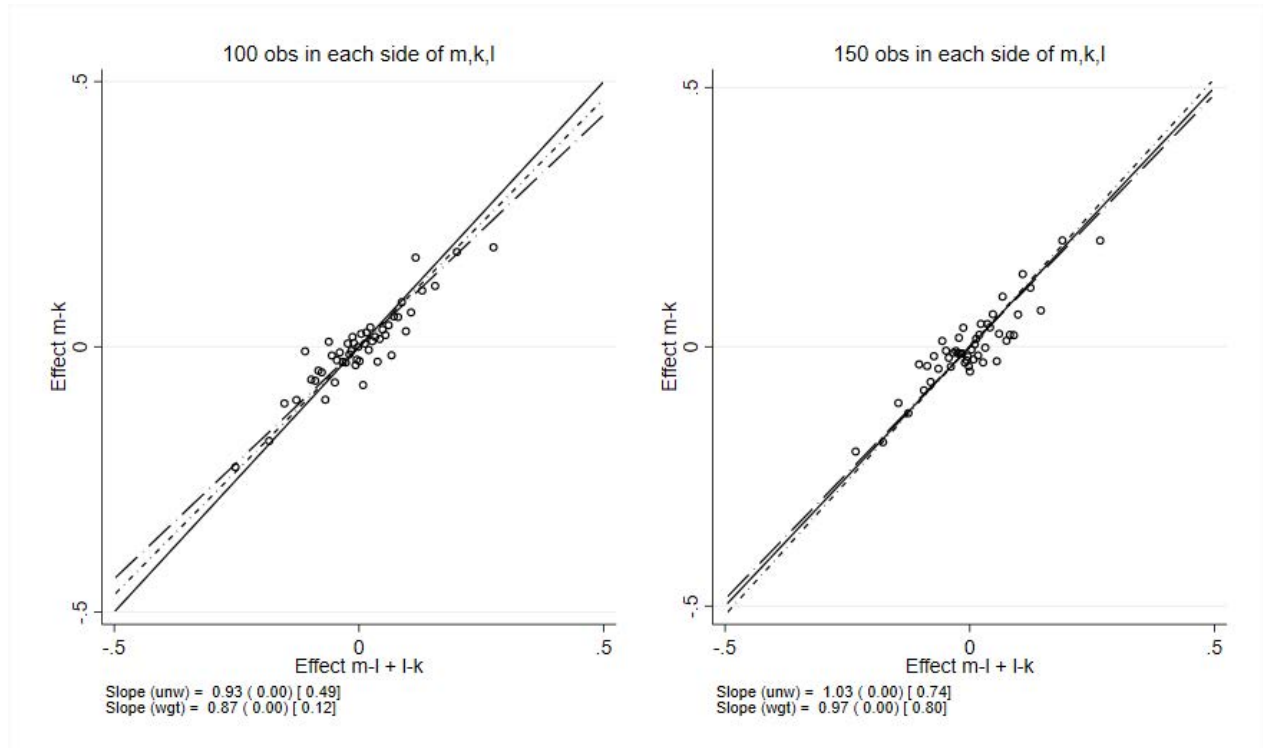
Notes: The Y-axis represents the probability of student i being assigned to her top choice. The X-axis represents the student's SEA score. The left panel shows the cutoff for school 1 and the right panel shows the cutoff for school 2. In both panels those who score below the cutoff for the preferred school are assigned to school 3. This figure illustrates the two different sources of variation. **The RD variation** identifies the effect of being assigned to school 1 by comparing the Low scoring group (right below the cutoff for school 1) to the Medium scoring group (right above the cutoff for school 1) in the left panel. Similarly, The RD variation identifies the effect of being assigned to school 2 by comparing the Medium scoring group (right below the cutoff for school 2) to the High scoring group (right above the cutoff for school 2) in the right panel. **The Difference in Difference variation** comes from making comparisons *across* cutoffs even among those who are not right above or below a cutoff. For example, the difference in outcomes between the Low scoring group at Choice Group 1 and the Low scoring group at Choice Group 2 will reflect differences in choices (as both groups were assigned to the same school 3). By contrast, the difference in outcomes between the High scoring group at Choice Group 1 and the High scoring group at Choice Group 2 will reflect both the effect of being assigned to school 1 (relative to school 2) and the differences in choices. Therefore, if the effect of choices and test scores are additively separable, then [the difference in outcomes between High scorers] minus [the difference in outcomes between Low scorers], will identify the effect of attending school 1 relative to school 2.

Figure 2: Predicted Cutoff Effects versus Actual Cutoff Effects



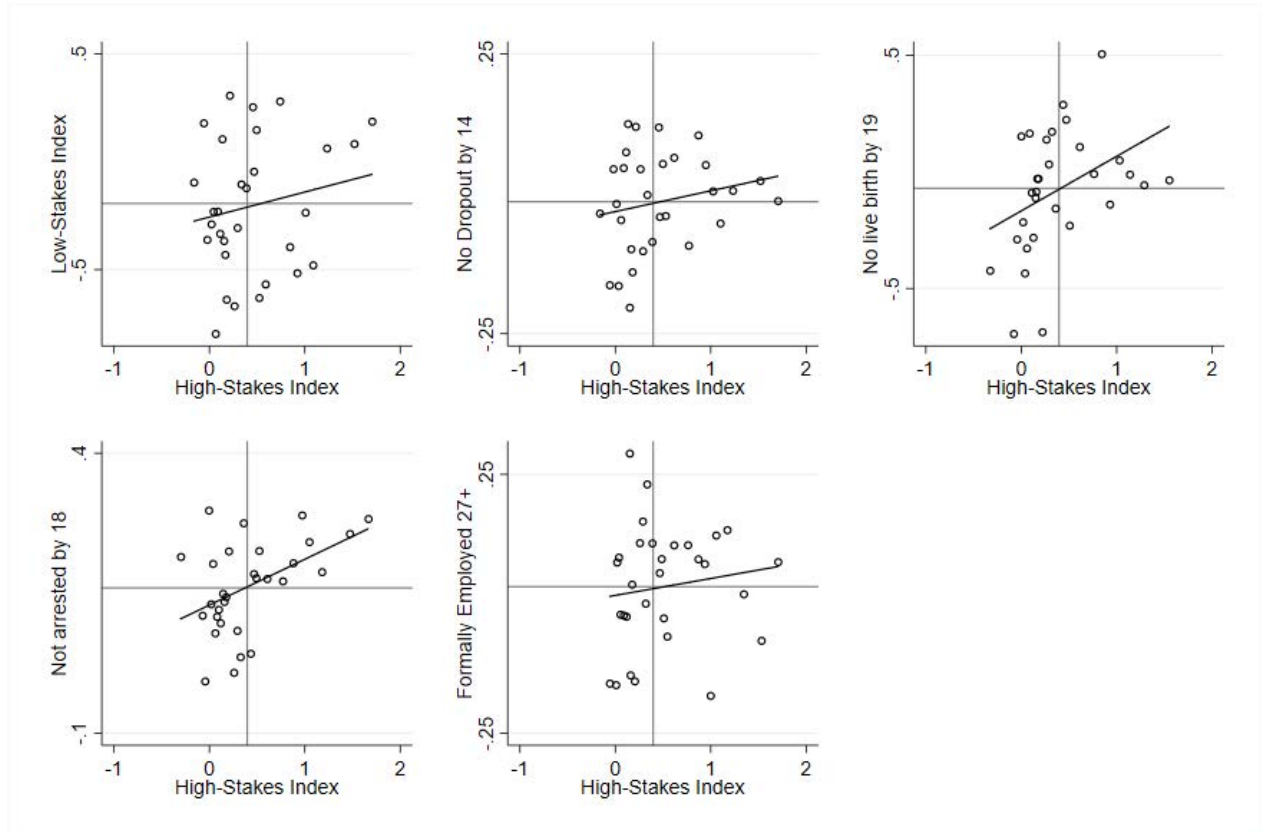
Notes: In both panels, the X-axis represents the estimated coefficients on an indicator for scoring above the rule-based cutoff resulting from an RD model that controls for a fifth degree polynomial of the SEA score, gender, district of residence at SEA registration, and religion; estimated for each school j and for each outcome where the estimated TOT school impacts ($\hat{\theta}_j^{TOTiv}$) enter as dependent variables. The Y-axis represents the estimated coefficients on an indicator for scoring above the rule-based cutoff resulting from an RD model that controls for a fifth degree polynomial of the SEA score, gender, district of residence at SEA registration, and religion; estimated for each school j and for each outcome where the individual level outcomes enter as dependent variables. The solid line represents a linear fit; while the dashed line is the 45° line. Estimated slope and p-values resulting from testing for whether the slope differs from both 0 and 1 are shown below the graph. Schools have been grouped in bins across the X-axis. Outliers above 4 standard deviations away from the median were removed. In the left panel, we show unweighted results and results were all schools estimated effects are weighted by the inverse of the squared standard error of each estimated coefficient for the predicted outcome. In the right panel, all schools estimated effects are adjusted by the reliability ratio in equation (20) and weighted by the inverse of the squared standard error of each estimated coefficient for the real outcome.

Figure 3: Additivity of the School Effects.



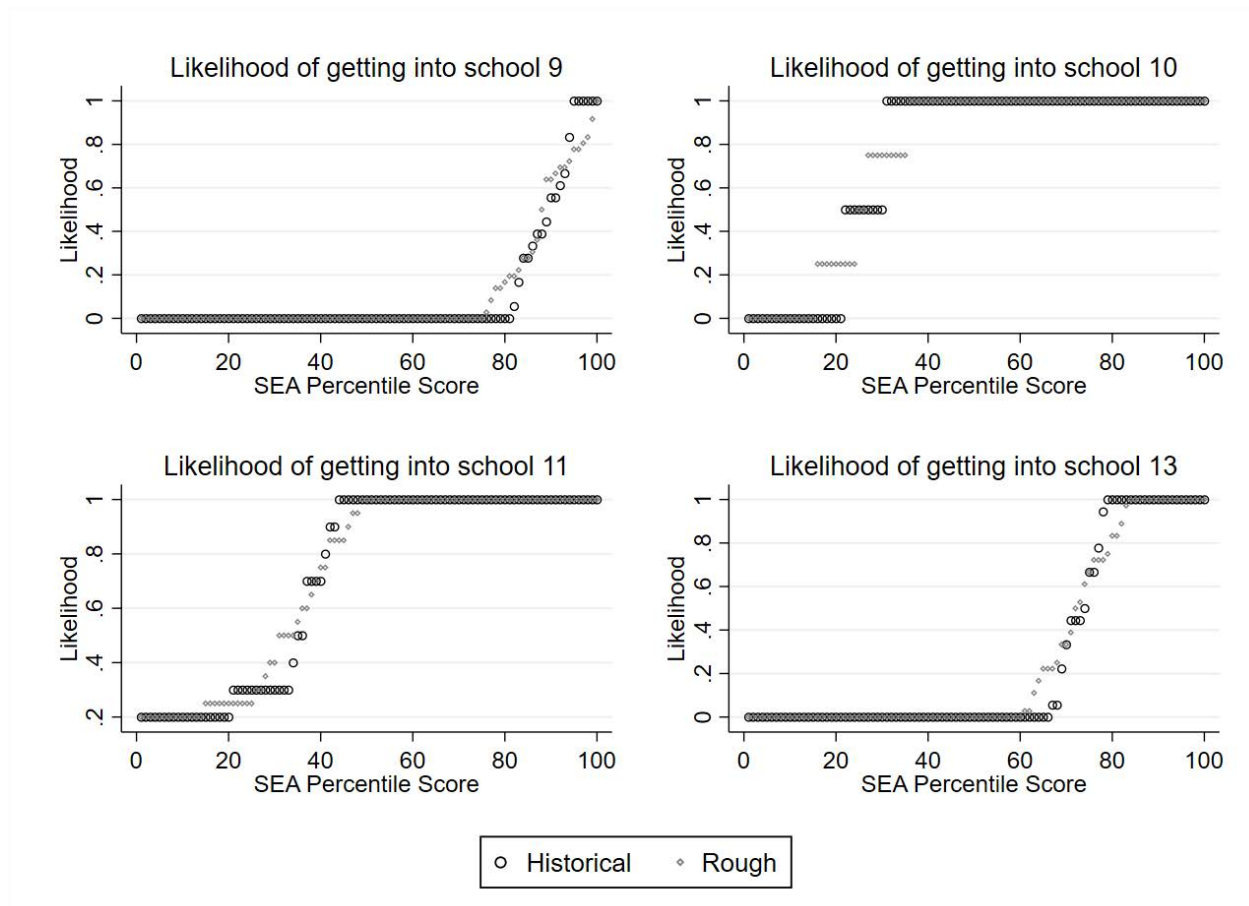
Notes: Both panels display estimated relative school effects pooled across all outcomes. The p -values associated with the hypothesis that the slope is 0 are in (parenthesis), and p -values associated with the hypothesis that the slope is 1 are in [brackets]. The short dashed line shows the unweighted fit, the long dashed line the weighted fit and the straight line the 45° relationship. We use the mean of the estimated standard errors of the school effects when computing the reliability ratio. Similar results are obtained when employing the median and are shown in [Appendix C](#).

Figure 4: Relationships Between High-Stakes and Other Impacts



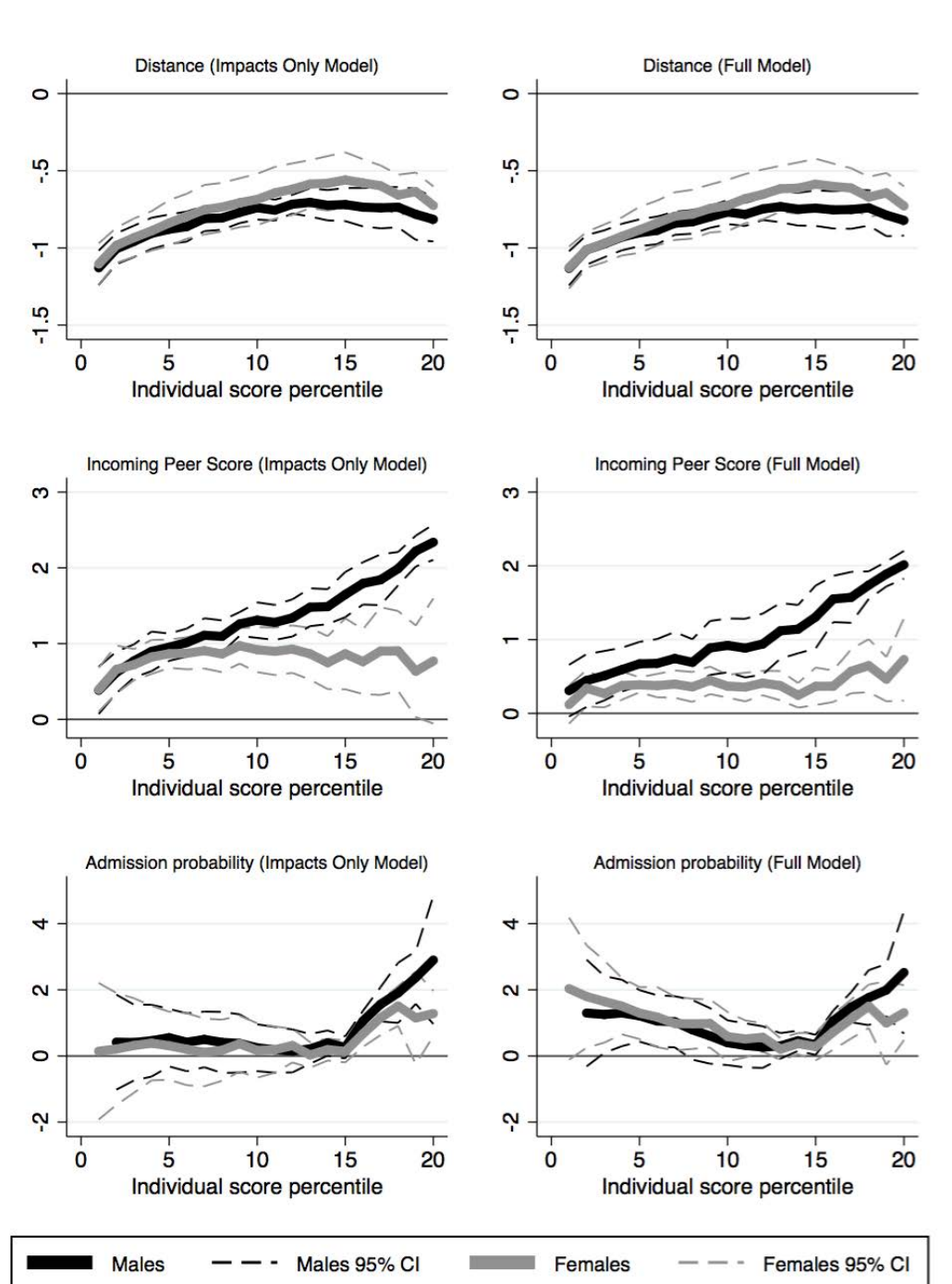
Notes: The X-axis represents estimated school impacts ($\hat{\theta}_j^{TOTIV}$) on the high-stakes index. The Y-axis represents the estimated school impacts on the other outcomes. A linear fit is added to each plot. Schools are grouped in 30 bins.

Figure 5: Estimated Admission Probabilities for Selected Schools



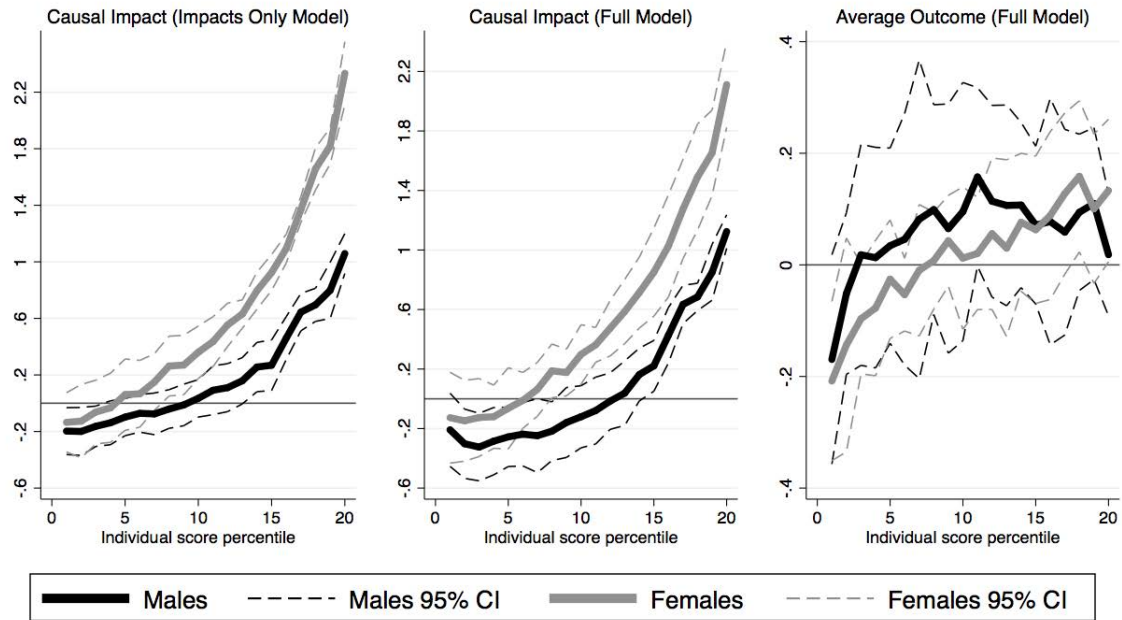
Notes: The X-axis represents the SEA score percentile. The Y-axis represents the likelihood of being assigned to a given school. School 9 is a very selective secondary school, where the school algorithm-based assignment cutoff has always been above the 82nd percentile. School 10, on the other hand, is a less selective school where students above the 30th percentile have always scored above the cutoff. Historical probabilities are depicted with black circles. “Rough” probabilities (calculated as described in the text) are depicted with gray diamonds.

Figure 6: Distance, Peer Quality, and Admission Probability



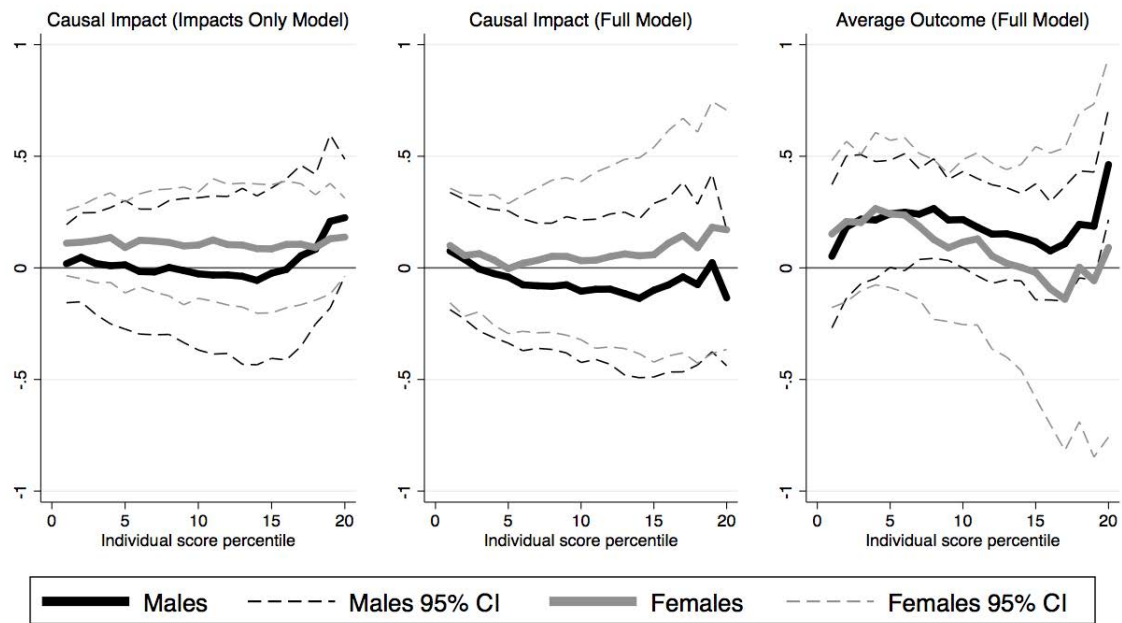
Notes: The connected lines represent the estimated coefficients, computed separately for each (SEA score ventile) × (gender) cell for two different models: Impacts Only Model (left panel), and Full Model (right panel). The dashed lines represent the associated 95% confidence intervals.

Figure 7: High-Stakes Index



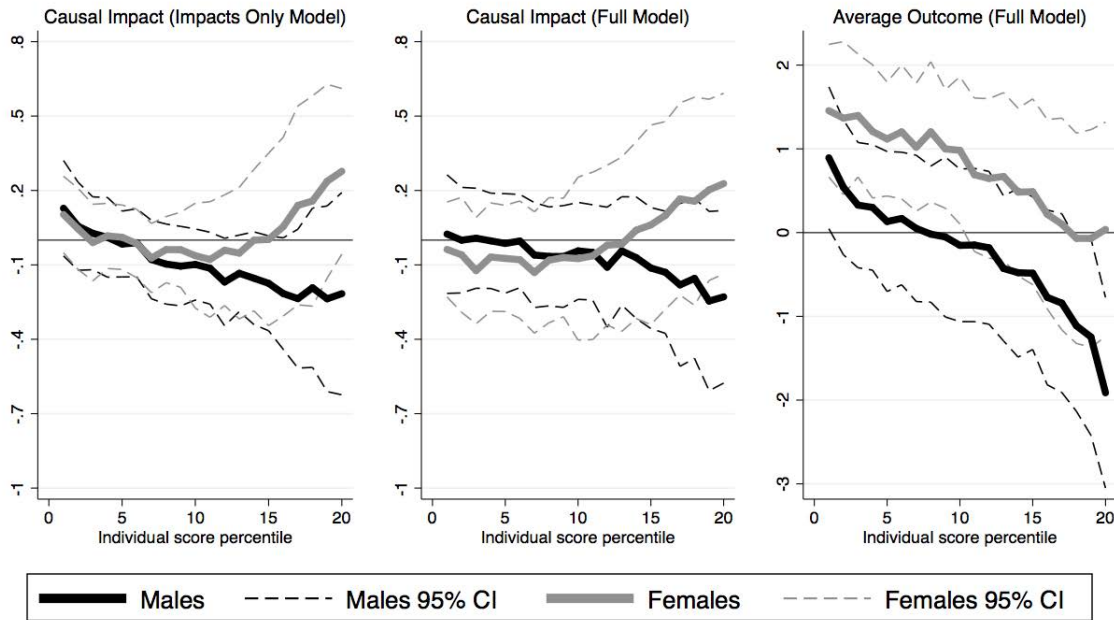
Notes: The connected lines represent the estimated coefficients, computed separately for each (SEA score ventile) \times (gender) cell for two different models: Impacts Only Model (left panel), and Full Model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure 8: Low-Stakes Index



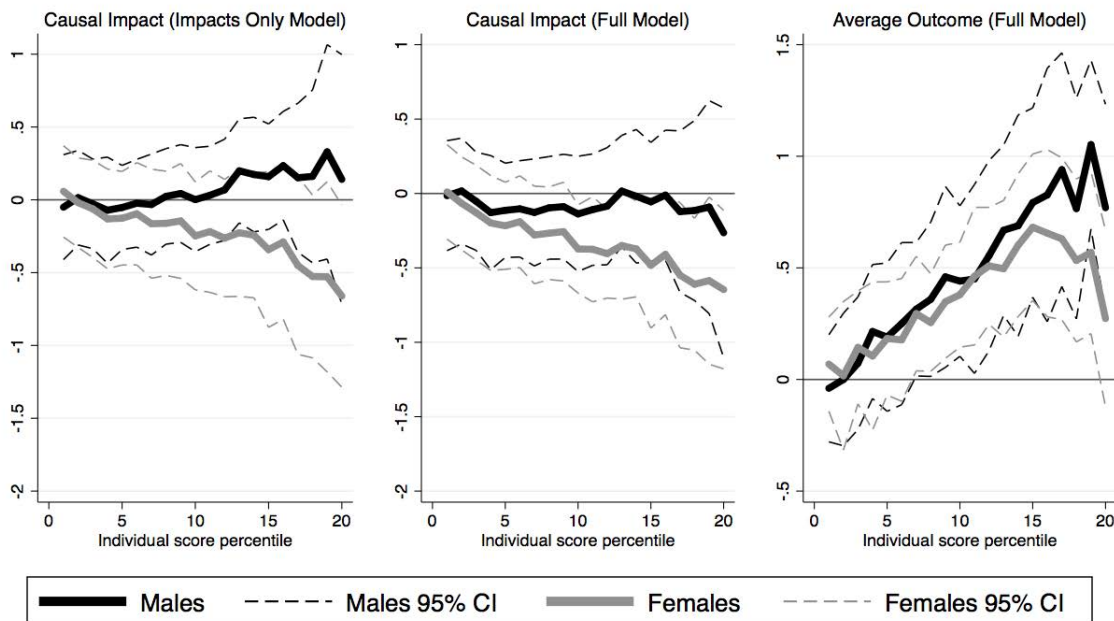
Notes: The connected lines represent the estimated coefficients, computed separately for each (SEA score ventile) \times (gender) cell for two different models: Impacts Only Model (left panel), and Full Model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure 9: No Dropout by Age 14



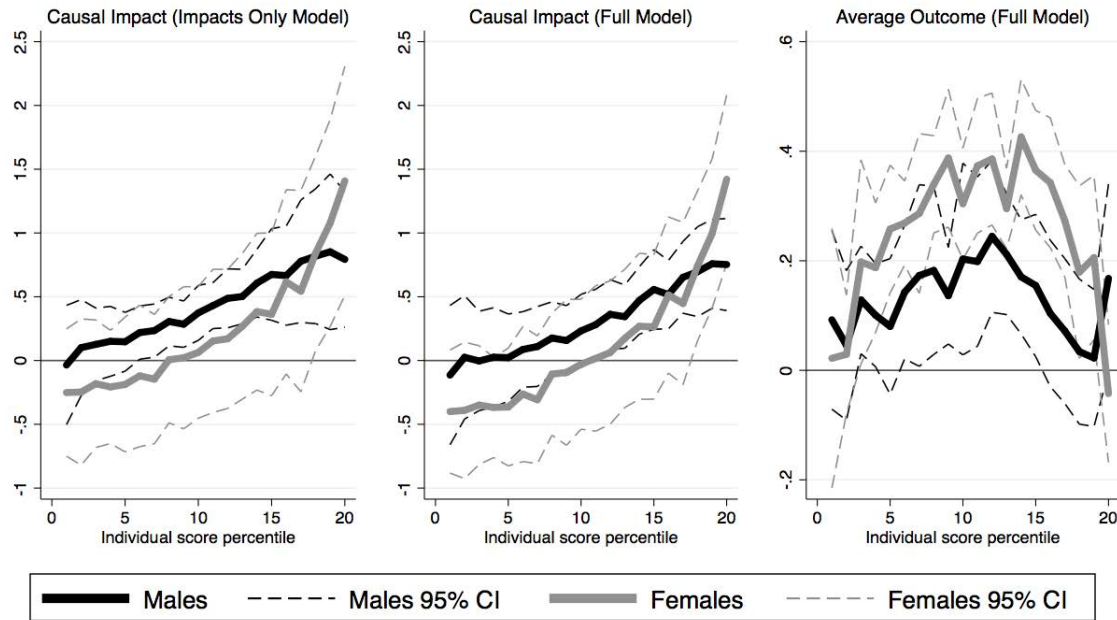
Notes: The connected lines represent the estimated coefficients, computed separately for each (SEA score ventile)×(gender) cell for two different models: Impacts Only Model (left panel), and Full Model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure 10: No Live Birth by Age 19



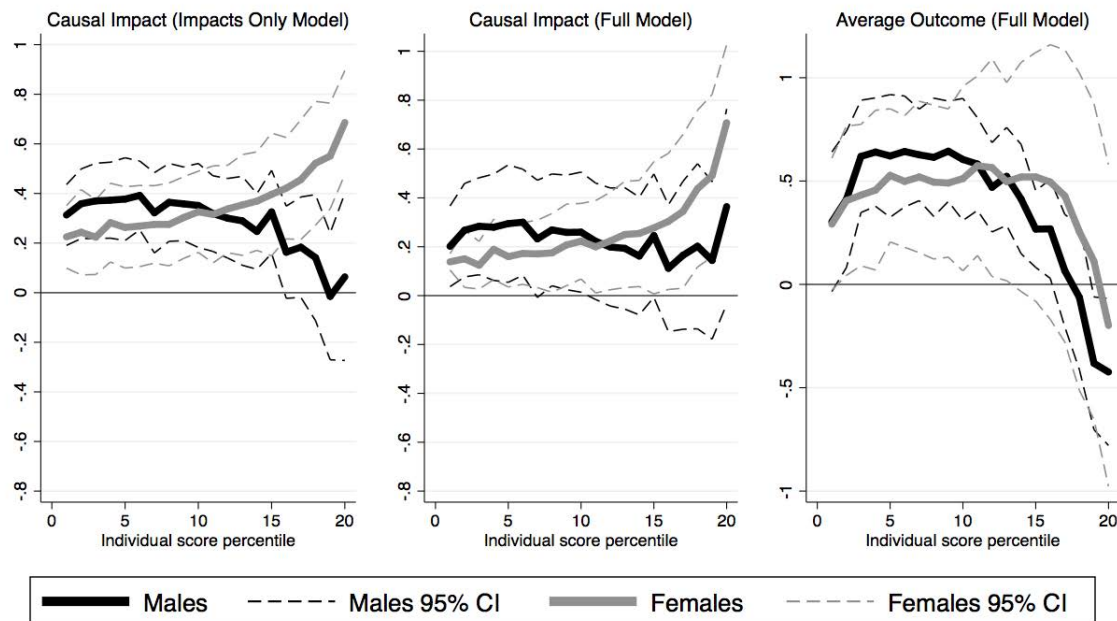
Notes: The connected lines represent the estimated coefficients, computed separately for each (SEA score ventile)×(gender) cell for two different models: Impacts Only Model (left panel), and Full Model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure 11: Not Arrested by Age 18



Notes: The connected lines represent the estimated coefficients, computed separately for each (SEA score ventile) \times (gender) cell for two different models: Impacts Only Model (left panel), and Full Model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure 12: Formally Employed at 27+ Years Old



Notes: The connected lines represent the estimated coefficients, computed separately for each (SEA score ventile) \times (gender) cell for two different models: Impacts Only Model (left panel), and Full Model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Acknowledgments. We are deeply grateful to Sabine Rieble-Aubourg and Dana King from the Inter-American Development Bank for their invaluable support in establishing the necessary contacts to assemble the data used in the study. We are indebted to Chief Education Officer Harrilal Seecharan of the Trinidad and Tobago Ministry of Education (TTMOE) for his continuous support. We also thank Ria Boafu, Lisa Henry-David, Shalini Maharaj, Brenda Moore, and Peter Smith of the TTMOE for facilitating access to the educational data needed for the study, their assistance, and their generosity. We thank Registrar General of Trinidad and Tobago Karen Bridgewater for kindly granting us access to the national birth records; Amos Sylvester from the Crime and Problem Analysis Branch of the Trinidad and Tobago Police Service for facilitating access to arrests records; and Executive Director Niala Persad of the National Insurance Board of Trinidad and Tobago, as well as Andy Edwards, Arlene Grant, Feyaad Khan and Bernard Smith for their support and generosity while working in their facilities to match employment records while maintaining individual confidentiality. Tatiana Zárate and Diego Zúñiga provided excellent research assistance. This paper benefited from comments by seminar participants at the NBER, MIT, UChicago, LACEA, Zurich, Lund, and Wharton.

References

- Atila Abdulkadiroğlu, Joshua Angrist, and Parag Pathak. The Elite Illusion: Achievement Effects at Boston and New York Exam Schools. *Econometrica*, 82(1):137–196, 2014.
- Atila Abdulkadiroğlu, Joshua D Angrist, Yusuke Narita, Parag A Pathak, Don Andrews, Tim Armstrong, Eduardo Azevedo, Yeon-Koo Che, Glenn Ellison, Brigham Frandsen, John Friedman, Justine Hastings, Guido Imbens, Jacob Leshno, Whitney Newey, Ariel Pakes, and Pedro Sant’anna. Breaking Ties: Regression Discontinuity Design Meets Market Design. Discussion Paper 2170, Cowles Foundation for Research in Economics. Yale University., 2019.
- Atila Abdulkadiroğlu, Parag Pathak, Jonathan Schellenberg, and Christopher Walters. Do Parents Value School Effectiveness? *American Economic Review*, 110(5):1502–1539, 2020.
- Claudia Allende, Francisco Gallego, and Christopher Neilson. Approximating the Equilibrium Effects of Informed School Choice. *Princeton University Mimeo*, 2019.
- Joshua D Angrist, Parag A Pathak, and Christopher R Walters. Explaining Charter School Effectiveness. *American Economic Journal: Applied Economics*, 5(4):1–27, 2013.
- Joshua D. Angrist, Sarah R. Cohodes, Susan M. Dynarski, Parag A. Pathak, and Christopher R. Walters. Stand and Deliver: Effects of Boston’s Charter High Schools on College Preparation, Entry, and Choice. *Journal of Labor Economics*, 34(2):275–318, 2016.
- Christopher N. Avery, Mark E. Glickman, Caroline M. Hoxby, and Andrew Metrick. A Revealed Preference Ranking of U.S. Colleges and Universities. *The Quarterly Journal of Economics*, 128(1):425–467, 2013.
- Tibor Besedeş, Cary Deck, Sudipta Sarangi, and Mikhael Shor. Age effects and heuristics in decision making. *Review of Economics and Statistics*, 94(2):580–595, 2012.
- Diether W. Beuermann and C. Kirabo Jackson. The Short and Long-Run Effects of Attending The Schools that Parents Prefer. *Journal of Human Resources*, pages 1019–1053R1, apr 2020.
- Kevin Booker, Tim R. Sass, Brian Gill, and Ron Zimmer. The Effects of Charter High Schools on Educational Attainment. *Journal of Labor Economics*, 29(2):377–415, 2011.

- Simon Burgess, Ellen Greaves, Anna Vignoles, and Deborah Wilson. What Parents Want: School Preferences and School Choice. *The Economic Journal*, 125(587):1262–1289, 2015.
- Hector Chade and Lones Smith. Simultaneous Search. *Econometrica*, 74(5):1293–1307, 2006.
- Raj Chetty, John N. Friedman, and Jonah E. Rockoff. Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review*, 104(9):2593–2632, 2014.
- John E Chubb and Terry M. Moe. *Politics, markets, and America’s schools*. 1990.
- David J. Deming. Better Schools, Less Crime? *The Quarterly Journal of Economics*, 126(4): 2063–2115, 2011.
- David J. Deming. Using School Choice Lotteries to Test Measures of School Effectiveness. *American Economic Review*, 104(5):406–411, 2014.
- David J. Deming, Justine S. Hastings, Thomas J. Kane, and Douglas O. Staiger. School Choice, School Quality, and Postsecondary Attainment. *The American Economic Review*, 104:991–1013, 2014.
- Jonah Deutsch, Brian Gill, and Matthew Johnson. The Promotion Power Impacts of Louisiana High Schools (Executive Summary). *Mathematica Policy Research Reports*, 2020. URL <https://www.mathematica.org/our-publications-and-findings/publications/es-the-promotion-power-impacts-of-louisiana-high-schools>.
- Will Dobbie and Roland Fryer. Charter Schools and Labor Market Outcomes. *Journal of Labor Economics*, forthcoming.
- Will Dobbie and Roland G. Fryer. The Medium-Term Impacts of High-Achieving Charter Schools. *Journal of Political Economy*, 123(5):985–1037, 2015.
- Milton Friedman. *The Role of Government in Education*. University of Chicago Press, 1955.
- D. Gale and LS Shapley. College Admissions and the Stability of Marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.
- Guillaume Haeringer and Flip Klijn. Constrained school choice. *Journal of Economic Theory*, 144 (5):1921–1947, 2009.
- Eric A. Hanushek. Teacher characteristics and gains in student achievement: Estimation using micro data. *American Economic Review*, 61(2):280–288, 1971.
- Hart Research Associates. Public School Parents On The Value Of Public Education, 2017.
- Justine Hastings, Thomas Kane, and Douglas Staiger. Parental Preferences and School Competition: Evidence from a Public School Choice Program. NBER Working Paper 11805, Cambridge, MA, 2005.
- Justine Hastings, Thomas Kane, and Douglas Staiger. Preferences and Heterogeneous Treatment Effects in a Public School Choice Lottery. NBER Working Paper 12145, Cambridge, MA, 2006.
- Justine Hastings, Christopher Neilson, and Seth Zimmerman. The Effects of Earnings Disclosure on College Enrollment Decisions. NBER Working Paper 21300, Cambridge, MA, 2015.
- Justine S. Hastings and Jeffrey M. Weinstein. Information, School Choice, and Academic Achievement: Evidence from Two Experiments. *Quarterly Journal of Economics*, 123(4):1373–1414, 2008.

- Justine S Hastings, Thomas J Kane, and Douglas O Staiger. Heterogeneous Preferences and the Efficacy of Public School Choice. *Working Paper*, 2009.
- James J. Heckman, Jora Stixrud, and Sergio Urzua. The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, 24(3): 411–482, 2006.
- Bengt Holmstrom and Paul Milgrom. Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, & Organization*, 7:24–52, 1991.
- Guido Imbens and Karthik Kalyanaraman. Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies*, 79(3):933–959, 2012.
- Sheena S. Iyengar and Emir Kamenica. Choice proliferation, simplicity seeking, and asset allocation. *Journal of Public Economics*, 94(7-8):530–539, 2010.
- C. Kirabo Jackson. Do Students Benefit from Attending Better Schools? Evidence from Rule-based Student Assignments in Trinidad and Tobago. *The Economic Journal*, 120(549):1399–1429, 2010.
- C. Kirabo Jackson. Match Quality, Worker Productivity, and Worker Mobility: Direct Evidence from Teachers. *Review of Economics and Statistics*, 95(4):1096–1116, 2013.
- C. Kirabo Jackson. What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes. *Journal of Political Economy*, 126(5):2072–2107, 2018.
- C. Kirabo Jackson. Can Introducing Single-Sex Education into Low-Performing Schools Improve Academics, Arrests, and Teen Motherhood? *Journal of Human Resources*, 2019.
- C. Kirabo Jackson, Shanette Porter, John Easton, Alyssa Blanchard, and Sebastián Kiguel. School Effects on Socio-emotional Development, School-Based Arrests, and Educational Attainment. *American Economic Review: Insights*, feb forthcoming.
- Thomas Kane and Douglas Staiger. Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation. 2008.
- Tim Kautz, James J. Heckman, Ron Diris, Bas ter Weel, and Lex Borghans. Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success. NBER Working Paper 20749, 2017.
- Lars J. Kirkeboen, Edwin Leuven, and Magne Mogstad. Field of Study, Earnings, and Self-Selection. *The Quarterly Journal of Economics*, 131(3):1057–1111, 2016.
- W. Bentley MacLeod and Miguel Urquiola. Reputation and School Competition. *American Economic Review*, 105(11):3471–3488, 2015.
- W. Bentley MacLeod and Miguel Urquiola. Is Education Consumption or Investment? Implications for the Effect of School Competition. NBER Working Paper 25117, Cambridge, MA, 2019.
- Richard K. Mansfield. Teacher Quality and Student Inequality. *Journal of Labor Economics*, 33(3):751–788, 2015.
- Isaac Mbiti, Karthik Muralidharan, Mauricio Romero, Youdi Schipper, Constantine Manda, and Rakesh Rajani. Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania. *The Quarterly Journal of Economics*, 134(3):1627—1673, 2019.
- Justin McCrary. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2):698–714, 2008.

- Daniel McFadden. Conditional Logit Analysis of Qualitative Choice. In Ed. Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, 1973.
- Parag A Pathak and Tayfun Sönmez. School Admissions Reform in Chicago and England: Comparing Mechanisms by their Vulnerability to Manipulation. *American Economic Review*, 103(1): 80–106, 2013.
- Kate Place and Philip Gleason. Do Charter Middle Schools Improve Students’ College Outcomes? (Study Highlights). Technical report, Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, 2019.
- Cristian Pop-Eleches and Miguel Urquiola. Going to a Better School: Effects and Behavioral Responses. *American Economic Review*, 103(4):1289–1324, 2013.
- Alvin E Roth and Marilda A Oliveira Sotomayor. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis (Econometric Society Monographs)*. 1992.
- Jesse M Rothstein. Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions. *American Economic Review*, 96(4):1333–1350, 2006.
- Donald B Rubin. Formal Modes of Statistical Inference for Causal Effects. *Journal of Statistical Planning and Inference*, 25:279–292, 1990.
- Arthur Schram and Joep Sonnemans. How individuals choose health insurance: An experimental analysis. *European Economic Review*, 55(6):799–819, 2011.
- Kenneth E. Train. *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge, 2009.

Appendix: NOT FOR PUBLICATION

Appendix A: School Placement Rules and Validity of the Regression Discontinuity Identification Strategy

The School Assignment Algorithm

School slots are assigned in rounds such that the most highly subscribed/ranked school fills its spots in the first round, then the next highly subscribed school fills its slots in the second round, and so on until all school slots are filled. This is done as follows: (1) the number of school slots at each school n_j is predetermined based on capacity constraints. (2) Each student is tentatively placed in the applicant pool for her first choice school and is ranked by SEA score. (3) The school at which the n_j^{th} ranked student has the highest SEA score is determined to be the most highly subscribed/ranked school and the top n_{j1} students in the applicant pool for top-ranked school j_1 are assigned to school j_1 . The SEA score of the n_{j1}^{th} student is the cutoff score for school j_1 . (4) The top-ranked school slots and the assigned students are removed from the process, and the second choice becomes the new "first choice" for students who had the top-ranked school as their first choice but did not gain admission. (5) This process is repeated in round two to assign students to the second highest ranked school j_2 and determine the cutoff score for the second-ranked school, and this is repeated in subsequent rounds until all slots are filled. This assignment mechanism is a deferred acceptance algorithm (Gale and Shapley 1962) in which students have incentives to truthfully reveal their rankings among chosen schools.

However, there is an important exception to the school assignment algorithm-based rule. Specifically, Government assisted schools (which are privately managed public schools – akin to Charter schools in the US) can admit up to 20 percent of their incoming class at the principal's discretion. As such, the rule is used to admit at least 80 percent of the students at these schools, while the remaining students can be hand-picked by the principal before the next-highest ranked school fills any of its slots. For example, suppose the highest ranked school has 100 slots and is a Government assisted school. The top 80 applicants to that school will be admitted, while the principal can hand-pick up to 20 other students at her discretion. The remaining 20 students would be chosen based on for example family alumni connections, being relatives of teachers, religious affiliation, and so on. These hand-picked students may list the school as their top choice, but this need not be the case. Students receive one admission decision and are never made aware of other schools they would have been admitted to had they not been hand-picked. Only after all the spots (including both admitted students based on the algorithm and on the hand-picking) at the highest ranked school have been filled will the process be repeated for the remaining schools. As such, school admissions are based partly on the described deterministic function of student test scores and student choices and partly on the endogenous selection of students by school principals at Government assisted schools.

In addition, there are other circumstances by which the attended school would differ from the algorithm-based assigned school. First, students who do not score high enough to be assigned to a school on their choice list receive an administrative placement from the Ministry of Education (made to the administrative school zoned to the students' residential location). Finally, due to unforeseen circumstances some schools may have less capacity than expected or may close (this may happen due to flooding etc.). In such rare cases, the Ministry will place students to schools based on open slots in nearby schools, open slots in other schools in the choice list, and proximity.

Simulating the School Assignments Using the Algorithm-Based Rule

Because the assignment algorithm is known and we have the same data used by the Ministry of Education to tentatively assign students, we can identify the algorithm-based assignment cutoffs and, therefore, the algorithm-based school assignments (i.e. those that would have been the actual school allocations if Government assisted schools could not select any of their own students). This algorithm-based or tentative assignment removes the part of the actual admission process that may be driven by endogenous selection and leaves only the variation in the assignments that are known deterministic functions of students' test scores and school choices.

Following Jackson (2010) and Pop-Eleches and Urquiola (2013), we stack the data across all application pools for each year to each school (that is, we stack data for all the cutoffs into a single cutoff) into one single database. As such, we stack all application cutoffs and re-center the SEA scores for applicants to each school in each year around the algorithm-based assignment cutoff for that school-year.⁴¹ Scoring above zero means scoring above the cutoff for a preferred school. Figure A1 shows the relationship between actually *attending* to one's preferred school as a function of one's incoming test score relative to the assignment cutoff for that school.⁴² Consistent with our assignment cutoffs capturing real exogenous variation in actual school attendance, there is a sudden increase in the likelihood of attending a preferred school as one's score goes from below to above the assignment cutoff. Appendix Table A2 reports this first-stage estimated coefficient evidencing its high significance. This shows that there are meaningful differences in preferred school attendance associated with scoring above versus below an assignment cutoff that are not due to selection or hand-picking. Next, we provide direct supporting evidence on the exogeneity of the algorithm-based assignment cutoffs.

Testing the Exogeneity of the Assignment Cutoffs

The RD variation used in this paper is driven by the assignment cutoffs. As such, here we present evidence that this identification strategy is likely valid. One key diagnostic is to test for smoothness of density across the simulated cutoffs (McCrary 2008). As such, we formally test for any differential density across simulated cutoffs within each of our SEA cohorts by regressing the density of observations at each relative SEA score on an indicator for scoring above the cutoff along with smooth functions of the relative score.⁴³ As one can see in Appendix Table A1, these tests evidence no statistically significant relationship between scoring above the cutoff and the density. Therefore, there is little evidence of gaming around the cutoffs regarding the density of observations at each test score.

The validity of the identification strategy also requires that there be no sorting of students around the cutoff (i.e. that latent outcomes are smooth through the cutoff). Given that students are unaware

⁴¹Specifically, for each school we find all students who list that school as their top choice, re-center those students' SEA scores around the simulated cutoff for that school, and create a sample of applicants for each school. To mimic the sequential nature of the algorithm, we remove students assigned to their top choice schools, replace students' first choice with their second choice, and repeat this process with their second, third, fourth, fifth, and sixth choices. The applicant samples for all schools are then stacked so that every student has one observation for each school for which she/he was an applicant. We use four or six choices, as relevant per cohort limit. Only for SEA cohorts 2001-2006 students were allowed to list up to 6 school choices. Therefore, most of SEA cohorts in our data (1995-2000 and 2007-2012) could list up to 4 school choices.

⁴²We consider that one student attended school j if the student was enrolled in school j at the time of writing the CSEC examinations.

⁴³We implement these tests using the *rddensity* command in Stata.

of the location of the cutoffs and are forced to make school choices before they take the SEA examinations, it is very unlikely that there is any sorting around the test score cutoffs. However, to provide further evidence that the variation employed (due to the cutoffs) is valid, we compute predicted outcomes (using the available baseline information) and test for whether scoring above the assignment cutoff is associated with any significant change in predicted outcomes.

Specifically, we first regress our outcomes on the number of SEA attempts (repeater status in 5th grade), the student's sex, the student's religion, selectivity of the student's primary school (measured by the average SEA scores of each primary school-year), selectivity of the student's secondary school choices (measured by the average SEA scores of the incoming class to each school choice-year), month of birth (to measure quarter of birth effects), age at SEA, and SEA cohorts fixed effects. These variables are relatively good predictors of the examination indexes such that, as shown in column 1 of [Appendix Table A2](#), they yield adjusted R-squares ranging from 0.27 to 0.31. However, the predictive power for the nonacademic binary outcomes is low.

We then take the fitted values from these prediction regressions as our predicted outcomes. If there was some gaming of the cutoff, one would likely see that scoring above the cutoff (conditional on smooth functions of the relative SEA score) should be associated with better "predicted" scores. However, with no gaming there should be no relationship between scoring above the cutoff and one's predicted outcomes. To test for this, we estimate the following model using our stacked database:

$$Y_{ijt}^p = \pi \cdot Above_{i\tau t} + f(SEA_{it}) + C_{\tau t} + \epsilon_{ijt} \quad (14)$$

where Y_{ijt}^p is the predicted outcome for individual i who attended school j at time t . $Above_{i\tau t}$ is an indicator for scoring above the algorithm-based assignment cutoff for school τ . Among those who comply with the cutoff, $j=\tau$. $f(SEA_{it})$ is a 5th order polynomial of the incoming SEA score net of the cutoff score for preferred school τ fully interacted with the $Above_{i\tau t}$ indicator. $C_{\tau t}$ is an cutoff fixed effect for applicants to school τ in year t . The inclusion of cutoff fixed effects ensures that all comparisons are among students who applied to the same school in the same year. Because the same individual can enter the data for multiple cutoffs, the estimated standard errors are clustered at the individual level.

Consistent with no gaming, column 2 of [Appendix Table A2](#), shows that there is no relationship between scoring above the cutoff and one's predicted outcomes. The estimated coefficients, $\hat{\pi}$, are small in magnitude and statistically indistinguishable from zero – indicating no gaming across the assignment cutoffs. Furthermore, we also report the estimated RD effects on the actual outcomes in columns (4) and (6) showing that reduced-form effects of scoring above the school assignment cutoff are associated with significant improvements in students' examination indexes and that these estimates are not sensitive to the inclusion of baseline sociodemographic controls in the model.

As an additional check on this model, we estimated model (14) for different bandwidths around the cutoff. [Figure A2](#) presents these results visually. As one can see for any choice of bandwidth, there are no effects of scoring above the cutoff on predicted outcomes. Taken together, the patterns suggest that the variation due to the algorithm-based assignment cutoffs is likely exogenous and, therefore, valid to identify causal school impacts.

Table A1. Testing for differential density around the school assignment cutoff

SEA Cohort	<i>p-value</i>	SEA Cohort	<i>p-value</i>
1995	0.1422	2004	0.8890
1996	0.9412	2005	0.2668
1997	0.5555	2006	0.6074
1998	0.8301	2007	0.5605
1999	0.6588	2008	0.1919
2000	0.7422	2009	0.7875
2001	0.7008	2010	0.7668
2002	0.9717	2011	0.2378
2003	0.4672	2012	0.5204

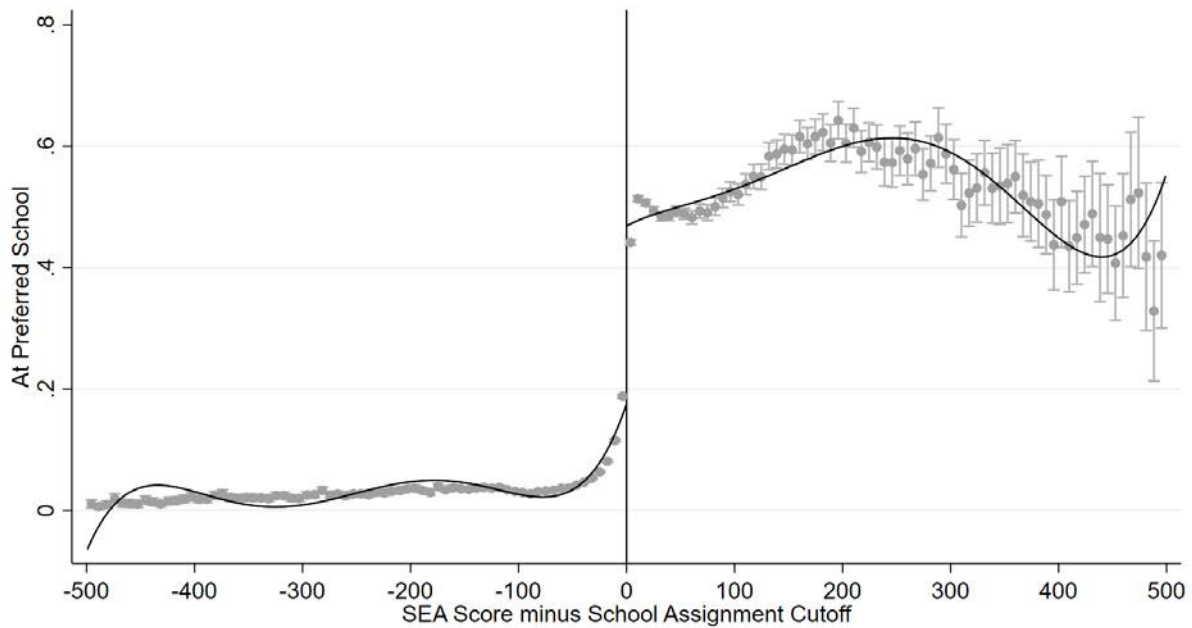
Notes: This table reports *p-values* of differential density tests across school assignment cutoffs for each SEA cohort included in the study.

Table A2. First Stage and Reduced-Form Effects

	Predicted Outcomes			Actual Outcomes			
	Prediction R2	Effect	<i>p-value</i>	Effect	<i>p-value</i>	Effect	<i>p-value</i>
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
First Stage:							
Attended preferred school				0.364	<0.001	0.365	<0.001
Reduced-Form Effects:							
High-Stakes Index	0.31	-0.001	0.821	0.061	<0.001	0.061	<0.001
Low-Stakes Index	0.27	-0.001	0.729	0.043	<0.001	0.045	<0.001
No Dropout by 14	0.09	-0.002	0.386	-0.007	0.429	-0.007	0.433
No live birth by 19	0.02	0.000	0.843	0.000	0.999	-0.002	0.878
Not arrested by 18	0.03	0.000	0.618	-0.004	0.521	-0.004	0.573
Formally employed 27+	0.04	-0.001	0.525	0.003	0.794	0.003	0.773
Cutoff fixed effects		Yes		Yes		Yes	
Sociodemographics		No		No		Yes	

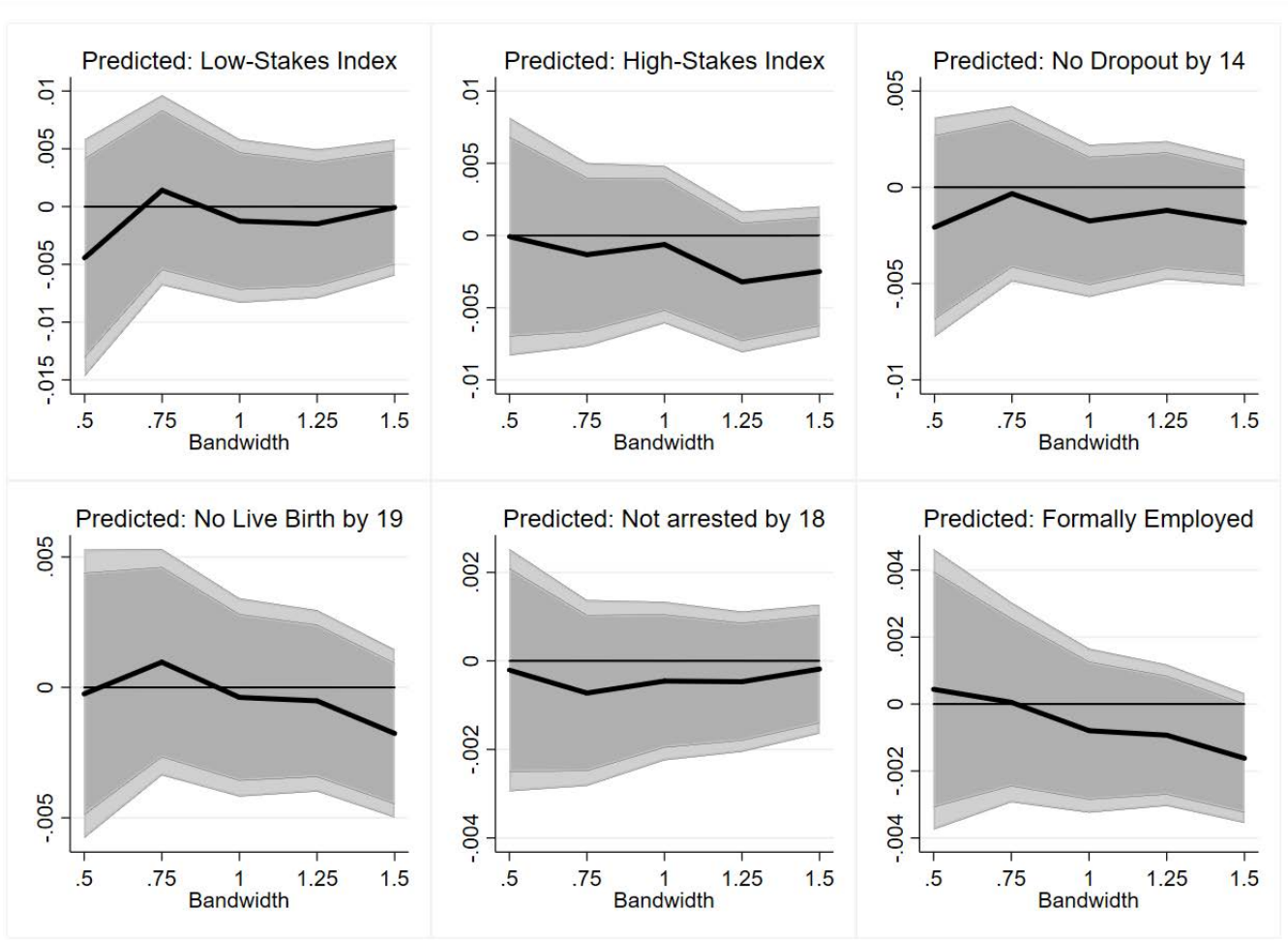
Notes: This table reports estimated coefficients on 'Above' resulting from equation (14). Models were estimated using all available observations within a bandwidth of ± 1 standard deviations from the school assignment cutoff. Sociodemographics include sex, primary school district fixed effects, and religion fixed effects. Estimated standard errors are clustered at the individual level in all regressions. *P-values* for the null of $\pi=0$ shown next to the estimated coefficients.

Figure A1. Discontinuity in Preferred School Attendance Through Assignment Cutoffs



Notes: The Y-axis represents the likelihood of preferred school attendance (i.e. the school where the student was enrolled at the time of taking the CSEC examinations). The X-axis is the SEA score relative to the deferred acceptance rule-based assignment cutoff. The circles are means corresponding to 7-point bins of the relative score. The solid lines are the fitted school attendance rates generated by fitting a fifth degree polynomial of the relative score fully interacted with an indicator for scoring above the school assignment cutoff. The gray vertical bars depict the 90 percent confidence intervals for each bin average.

Figure A2. Reduced-form effects on predicted outcomes by bandwidth



Notes: This figure reports estimated coefficients on 'Above' resulting from equation (14). The estimated coefficients are reported for each bandwidth between $\pm 0.5sd$ and $\pm 1.5sd$ from the school assignment cutoff. The 90 (95) percent confidence intervals of the estimated coefficients are presented in dark (light) gray.

Appendix B: Appendix Tables

Table B1: Weights Used to Compute Indexes

High-Stakes Index	Weight
Number of CSEC subjects passed	0.202
CSEC tertiary qualification	0.192
CSEC tertiary qualification attempt	0.140
CAPE scholarship	0.068
CAPE scholarship attempt	0.213
Number of CAPE units passed	0.219
CAPE Associate's degree	0.213
Low-Stakes Index	Weight
NCSE Total Academic	0.546
NCSE Total Non academic	0.546

Notes: Indexes are computed from a separate factor analysis (using the principal-component factor method) applied to the individual outcomes that integrate each index. The weights for individual outcomes within the indexes are determined by predicting the first underlying principal-component applied separately to each group of outcomes that integrate each index. The computed indexes are standardized to have zero mean and unit variance. CSEC tertiary qualification is obtained when passing 5 subjects including English language and mathematics. "CSEC tertiary qualification attempt" denotes that the student took 5 subjects including English language and mathematics. CAPE scholarship is awarded when passing eight CAPE units (including Caribbean and Communication studies) with the maximum possible grade. "CAPE scholarship attempt" denotes that the student took eight CAPE units (including Caribbean and Communication studies). CAPE associate's degree is awarded when passing seven CAPE units (including Caribbean and Communication studies). NCSE academic subjects include mathematics, English, Spanish, sciences, and social studies. NCSE non academic subjects include arts, physical education, and technical studies.

Table B2: Maximum Likelihood Correlations Between School Impacts

	High-Stakes Index	Low-Stakes Index	No Dropout by 14	Not arrested by 18	No live birth by 19	Formally employed 27+	Low-Stakes Index (A)	Low-Stakes Index (NA)
Unweighted								
High-Stakes Index	1.00							
Low-Stakes Index	0.06	1.00						
No Dropout by 14	0.26	0.23	1.00					
Not arrested by 18	0.47	-0.03	0.43	1.00				
No live birth by 19	0.13	-0.03	-0.20	-0.51	1.00			
Formally employed 27+	0.19	0.00	-0.01	0.21	-0.16	1.00		
Low-Stakes Index (A)	0.25	0.83	0.33	0.14	-0.08	-0.04	1.00	
Low-Stakes Index (NA)	-0.05	0.96	0.15	-0.12	0.00	0.02	0.62	1.00
Weighted around 25 percentile of the SEA score								
High-Stakes Index	1.00							
Low-Stakes Index	0.20	1.00						
No Dropout by 14	-0.07	0.38	1.00					
Not arrested by 18	0.34	0.22	0.12	1.00				
No live birth by 19	0.23	0.26	-0.20	-0.10	1.00			
Formally employed 27+	0.25	0.15	-0.10	0.13	0.03	1.00		
Low-Stakes Index (A)	0.26	0.89	0.52	0.24	0.25	0.10	1.00	
Low-Stakes Index (NA)	0.15	0.96	0.24	0.18	0.19	0.17	0.72	1.00
Weighted around 75 percentile of SEA score								
High-Stakes Index	1.00							
Low-Stakes Index	0.13	1.00						
No Dropout by 14	0.16	-0.28	1.00					
Not arrested by 18	0.50	-0.15	0.28	1.00				
No live birth by 19	0.06	-0.08	0.26	-0.39	1.00			
Formally employed 27+	0.20	0.17	0.06	0.25	-0.37	1.00		
Low-Stakes Index (A)	0.37	0.71	0.00	0.07	0.19	0.08	1.00	
Low-Stakes Index (NA)	0.00	0.96	-0.34	-0.22	-0.17	0.17	0.45	1.00

Notes: The table reports correlations of persistent school effects across different outcomes. Correlations were computed using the maximum likelihood approach described in the text. Correlations in the top panel were computed with estimated school effects for the average student and those in the middle and bottom panels were computed with school effects weighted around the 25th and 75th percentile of the students' SEA score distribution respectively. We removed students who attended schools with outlier estimated impacts (i.e. beyond 4σ of the median school).

Appendix C: Validation and Magnitude of School Impacts

Additive Separability of Choices and Test Scores I: Regression Discontinuity Variation vs. All Variation

Existing papers that have explored parental preferences for school causal impacts have either relied on school average outcomes (which may not reflect their impacts *per se*) or estimated school impacts that may be biased due to selection.⁴⁴ If one's measures of school effectiveness do not accurately reflect schools' causal impacts, it may distort one's conclusions regarding parental preferences for school effectiveness. For this reason, validating the estimated school impacts as reflecting *causal* impacts is important. A key strength of our context and data is that we are able to validate our estimated school impacts using exogenous variation only. We test the validity of our value-added estimates by exploring if they are consistent with what one would obtain using quasi-random variation only.

Under the algorithm used to create the tentative school assignments (discussed in [Appendix A](#)), each school has a minimum score above which applicants are tentatively admitted and below which they are not. As such, the marginal effect of being tentatively assigned to each school (relative to the next lowest ranked school) can be estimated with a regression discontinuity design. That is, among students who are applicants to a given school τ , the causal effect of being tentatively assigned to school τ is simply the effect of scoring above the admission cutoff for school τ (conditional on smooth functions of ones incoming SEA score). In our setup, students are considered to be applicants to a school if that school is in their ranked list and they do not score above the cutoff for a more preferred school. Note, therefore, that students can be applicants to more than one school.⁴⁵

To obtain the reduced-form Regression Discontinuity (RD) effect of being tentatively assigned to any school τ , we estimate RD models for each outcome among all applicants to school τ .⁴⁶ Under the RD identifying assumptions, the reduced-form effect of being tentatively assigned to school τ on outcome Y , is captured by estimating the equation below.

$$Y_{ij} = Above_{i\tau} \cdot \gamma_{\tau} + f(SEA_i) + \mathbf{X}_i' \delta + \varepsilon_{ij} \quad (15)$$

Where Y_{ij} is the outcome of student i who attended school j , and $Above_{i\tau}$ is an indicator for scoring above the algorithm-based assignment cutoff for school τ . Among those who comply with the cutoff, $j=\tau$. The parameter γ_{τ} captures the difference in outcomes (all else equal) between those exogenously assigned to a preferred school τ (due to scoring above the cutoff) versus scoring below the cut off and attending the student's counterfactual school q (that is, the school that the students would have attended had they not scored above the cutoff for school τ). As such, in the

⁴⁴In related work [Abdulkadiroğlu et al. \(2020\)](#) examine parents responsiveness to school impacts that rely on selection on observables assumptions (similar to our estimates). However, they are unable to validate these school impact estimates using exogenous variation in school attendance.

⁴⁵For example, a student that was assigned to her first choice will only appear once as a (successful) applicant to her top choice school. However, a student who is assigned to her second choice school will appear twice: as an (unsuccessful) applicant to her top choice school and as a (successful) applicant to her second choice school.

⁴⁶That is we estimate separate reduced-form models where, in each one, we consider all persons who applied to a particular school τ in each year.

neighborhood of the cutoff,

$$E[\hat{\gamma}_\tau | X_i, SEA_i] = E(Y_{ij} | Above = 1) - E(Y_{iq} | Above = 0) \quad (16)$$

To simplify equation (16), we consider this expression for compliers and for non-compliers. Under the assumption of unconfoundedness (Rubin 1990), it follows that $E[Y_{ij} - Y_{iq} | X_i, SEA_i] = \theta_j^{TOT} - \theta_q^{TOT}$. That is, if there is no selection on observables, the average difference in outcomes between observationally equivalent individuals who attended school j and school q reflects the difference in causal impacts between school j and school q . Among the compliers, school j is school τ if they score above the cutoff.⁴⁷ As such, for compliers, $E[\hat{\gamma}_\tau | X_i, SEA_i] = \theta_\tau^{TOT} - E[\theta_q^{TOT}]$, where $E[\theta_q^{TOT}]$ is the average impact of the counterfactual schools for the applicants to school τ . Among non-compliers, the cutoff does not change the school attended so that $E[\hat{\gamma}_\tau | X_i, SEA_i] = 0$. It follows that for the average applicant to school τ , equation (16) can be written as equation (17) below.

$$E[\hat{\gamma}_\tau | X_i, SEA_i] = \bar{p}_\tau \times (\theta_\tau^{TOT} - E[\theta_q^{TOT}]) \quad (17)$$

In words, in expectation, the estimated effect of scoring above the cutoff for school τ is the difference between the impact of attending preferred school τ and that of attending the average counterfactual school q , all times the compliance rate (\bar{p}_τ). This is simply the weighted cutoff effect for the compliers and the non-compliers.

Consider now, estimating this same model, but replacing each student's actual outcome with the predicted TOT impact of the school they attended, $\hat{\theta}_j^{TOTiv}$, as below.

$$\hat{\theta}_j^{TOTiv} = Above_{it} \cdot \zeta_\tau + f(SEA_i) + \mathbf{X}_i' \delta + \varepsilon_{ij} \quad (18)$$

The parameter ζ_τ is the difference in predicted TOT school impacts (all else equal) between those scoring above the cutoff for preferred school τ versus not. In the neighborhood of the cutoff, the RD effect on the predicted TOT impacts of an individual's attended school is $E[\zeta_\tau | X_i, SEA_i] = E(\hat{\theta}_j^{TOTiv} | Above = 1) - E(\hat{\theta}_q^{TOTiv} | Above = 0)$. Using the same logic as above for compliers and non-compliers, it follows that

$$E[\hat{\zeta}_\tau | X_i, SEA_i] = \bar{p}_\tau \times (E[\hat{\theta}_\tau^{TOTiv}] - E[\hat{\theta}_q^{TOTiv}]) \quad (19)$$

In words, in expectation, the estimated difference in predicted school TOT impacts of scoring above the cutoff for school τ is the difference between the estimated TOT impact of attending preferred school τ and that of the average counterfactual school q , all times the compliance rate (\bar{p}_τ).

Inspection of (17) and (19) reveals that if our treatment on the treated estimated impacts for attended school j and the average counterfactual school q are unbiased, then by the law of iterated expectations, $E[\hat{\zeta}_\tau | X_i, SEA_i] = E[\hat{\gamma}_\tau | X_i, SEA_i]$. In words, if the estimated TOT school impacts ($\hat{\theta}_\tau^{TOTiv}$) are consistent estimates of the causal effect of attending school τ , then for each applicant school, the RD estimates using the actual outcomes as left hand side variables and the RD estimates using the TOT school impacts of the attended school j as left hand side variables, should be equal in expectation.

⁴⁷In most instances school q will be the next ranked school in the choice list, but could be any fallback school (such as a private school or Government Assisted school that can admit students irrespective of their SEA scores).

This motivates a validation test of our TOT school estimates. In related work, [Hastings et al. \(2015\)](#) implement a very similar test to validate the reliability of using predicted versus actual earnings when disseminating information on the expected returns to attend alternative colleges and majors. To implement this test, first, we estimate $\hat{\gamma}_\tau$ and $\hat{\zeta}_\tau$ for each preferred school τ (using the optimal bandwidth from [Imbens and Kalyanaraman \(2012\)](#)). Following [Hastings et al. \(2015\)](#), to account for estimation errors in the RD effects on school impacts, we implement Empirical Bayes estimates of each cutoff effect.⁴⁸ We then regress the former estimated coefficients on the latter (to account for noisiness in the RD estimated effects, we weight each estimate by the inverse of its squared standard error when doing this regression). Finally, we test for whether the estimated slope is statistically indistinguishable from 1. The results from this approach are reported in Figure 2 in the main text.

Exploring the Potentially Biased OLS Estimates: A similar procedure can be implemented with the potentially biased OLS estimates derived from equation (1) in the main text but where the school assignment indicators are replaced with school attendance indicators. In this case, we use the OLS estimates on the school attendance indicators, $\hat{\theta}_j^{TOTOLS}$, instead of the IV TOT school impacts as below.

$$\hat{\theta}_j^{TOTOLS} = Above_{i\tau} \cdot \eta_\tau + f(SEA_i) + \mathbf{X}_i' \delta + \varepsilon_{ij} \quad (21)$$

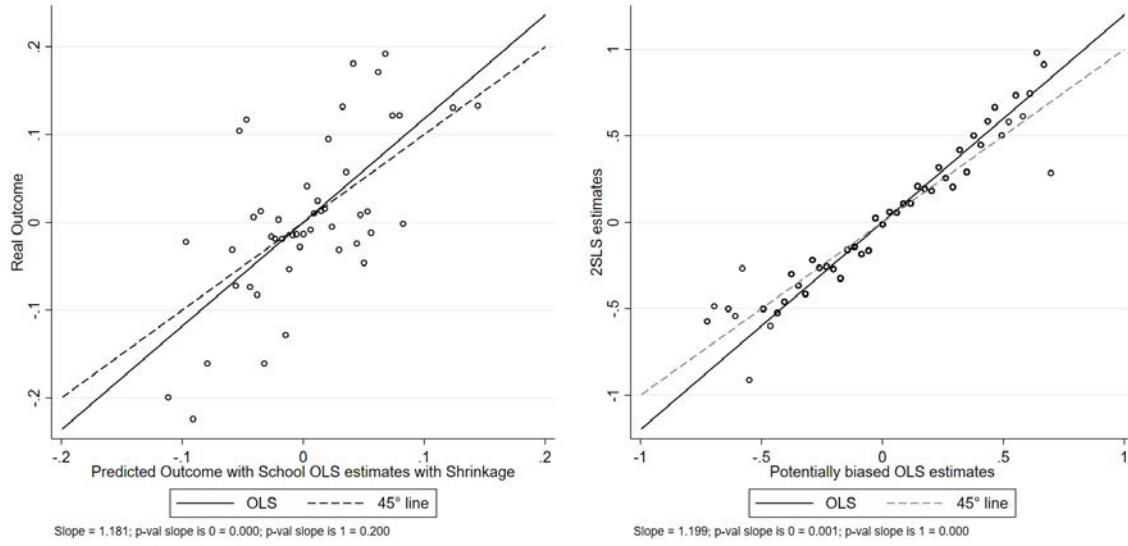
The results from this approach are shown in the left panel of Figure C1. There we regress $\hat{\gamma}_\tau$ on $\hat{\eta}_\tau$. The estimated slope is 1.18, although it cannot be rejected that it is equal to 1. However, because this is a noisy test, we also test whether the OLS estimates of school effects ($\hat{\theta}_j^{TOTOLS}$) are equivalent to the IV TOT ones ($\hat{\theta}_j^{TOTIV}$). This model (right panel) yields an estimated slope of 1.199 and one can reject that this slope is 1. Given that any estimation error in the OLS estimates would bias this slope toward zero, we can be confident that the slope is not 1—that is, we can reject that the OLS and IV TOT estimates yield the same effects in expectation. Given that the IV TOT estimates are almost identical to those implied by the RD model, we take this as evidence of bias in the OLS estimates.

⁴⁸Specifically, for any particular outcome, the *predicted* RD effect for school j is the weighted difference in estimated school impacts between those just above and below the cutoff for school j . We can express the estimated parameter as $\hat{\zeta}_{\hat{\theta}_j} = \sum_{k \in A} a_k(\hat{\theta}_k) - \sum_{k \in B} b_k(\hat{\theta}_k) = \sum_{k \in A} a_k(\theta_k + \varepsilon_k) - \sum_{k \in B} b_k(\theta_k + \varepsilon_k)$, where A is the set of schools the students attend above the cutoff, B the set of schools the students attend below the cutoff, a_k and b_k the proportions in which they do so and ε_k the estimation error for the school impact θ_k . If we assume that the school impacts are not independent within each cutoff, but that the estimation errors are, we can approximate the variance of this estimated parameter as $Var(\hat{\zeta}_{\hat{\theta}_j}) = \sum u_k \sigma_\theta^2 + \sum v_{lm} Cov_\theta + \sum w_k SE_k^2$, where σ_θ^2 is given by the magnitude of the school impacts, Cov_θ is approximated by the covariance between each pair of schools the students applied to and SE_k^2 is given by the square of the standard error of the school impact. If we also include the squared standard error of the RD estimate, $SE_{\hat{\zeta}}$, the reliability ratio of our RD estimate is given by

$$\lambda_j = \frac{\sum u_k \sigma_\theta^2 + \sum v_{lm} Cov_\theta}{(\sum u_k \sigma_\theta^2 + \sum v_{lm} Cov_\theta) + (\sum w_k SE_k^2 + SE_{\hat{\zeta}}^2)} \quad (20)$$

Our Empirical Bayes estimate of the *predicted* effect of cutoff j is therefore $[\lambda_j \times \hat{\zeta}_{\hat{\theta}_j}]$

Figure C1. Predicted Cutoff Effects versus Actual Cutoff Effects using Biased OLS Value Added



Notes: On the left panel, the X-axis represents the estimated coefficients on the 'Above' indicator resulting from model (21); estimated for each school τ and for each outcome (estimated OLS school effects, $\hat{\theta}_j^{TOTOLS}$, enter as dependent variables). The Y-axis represents the estimated coefficients on the 'Above' indicator resulting from model (15); estimated for each school τ and for each outcome (individual level outcomes enter as dependent variables). The connected line represents a lineal fit; while the dashed line is the 45° line. Estimated slope and p-values resulting from testing for whether the slope differs from both 0 and 1 are shown below the graph. Schools have been grouped in bins across the X-axis. Outliers above 4 standard deviations away from the median were removed. All schools estimated effects are adjusted by the reliability ratio given in equation (20) of the main text and weighted by the inverse of the squared standard error of each estimated coefficient for the real outcome. On the right panel, the Y-axis represents the estimated IV TOT school impacts derived from model (2) and (3), i.e. $\hat{\theta}_j^{TOTIV}$. The X-axis represents the estimated OLS school effects described in this section (i.e. $\hat{\theta}_j^{TOTOLS}$). In both panels, the solid line represents a lineal fit; while the dashed line is the 45° line. Estimated slope and p-values resulting from testing for whether the slope differs from both 0 and 1 are shown below the graph.

Additive Separability of Choices and Test Scores II: Robustness to Interactions

Our identification relies that there are no considerable interaction effects between test scores and choices. In order to test for this, we estimate our base model defined in (2) and (3) with four different interactions between choices and SEA scores. The *first interaction* considers the selectivity of choices. For this, we include an interaction between the polynomial of the student's own SEA score and the selectivity of each of the first 4 choices, this being approximated by the average peer SEA score of those assigned to that choice.

$$Y_{ijct} = \Sigma(I_{i,j} \cdot \theta_j^{TOTIV}) + f(SEA_i) + \Sigma(g_k(SEA_i) \cdot \overline{SEA}_{choice_k}) + \lambda_c + \mathbf{X}'_{it}\delta + C_t + \varepsilon_{ijct} \quad (22)$$

$$I_{i,j} = \Sigma(I_{i,\tau} \cdot \pi_{\tau j}) + f(SEA_i) + \Sigma(g_k(SEA_i) \cdot \overline{SEA}_{choice_k}) + \lambda_c + \mathbf{X}'_{it}\delta + C_t + v_{ijct}, \quad \text{for each } j \in J \quad (23)$$

Where $g_k(SEA_i)$ is a fifth order polynomial and $\overline{SEA}_{choice_k}$ is the mean total SEA scores for incoming assigned students for each choice k .

The *second interaction* is between SEA scores and student's choice sets fixed effects. Instead of using one fixed effect per choice set, we use two, one for those below the median SEA score in that choice set and one for those above.

$$Y_{ijct} = \Sigma(I_{i,j} \cdot \theta_j^{TOTIV}) + f(SEA_i) + \lambda_{ci} + \mathbf{X}'_{it}\delta + C_t + \varepsilon_{ijct} \quad (24)$$

$$I_{i,j} = \Sigma(I_{i,\tau} \cdot \pi_{\tau j}) + f(SEA_i) + \lambda_{ci} + \mathbf{X}'_{it}\delta + C_t + v_{ijct}, \quad \text{for each } j \in J \quad (25)$$

Now, instead of λ_c we will have λ_{ci} which will take one value if student i with choicesset c is above the median SEA score within that choice set and another if it is below.

The *third interaction* is between the student's SEA score and its n -th choice of school:

$$Y_{ijct} = \Sigma(I_{i,j} \cdot \theta_j^{TOTIV}) + f(SEA_i) + \Sigma(SEA_i \cdot I_{i,k}) + \lambda_c + \mathbf{X}'_{it}\delta + C_t + \varepsilon_{ijct} \quad (26)$$

$$I_{i,j} = \Sigma(I_{i,\tau} \cdot \pi_{\tau j}) + f(SEA_i) + \Sigma(SEA_i \cdot I_{i,k}) + \lambda_c + \mathbf{X}'_{it}\delta + C_t + v_{ijct}, \quad \text{for each } j \in J \quad (27)$$

Where $I_{i,k}$ is an indicator equal to 1 if the n -th choice of student i is school k . This is done for the first 4 choices.

Finally, the *fourth interaction* is between the student's SEA score polynomial and its 1st choice of school:

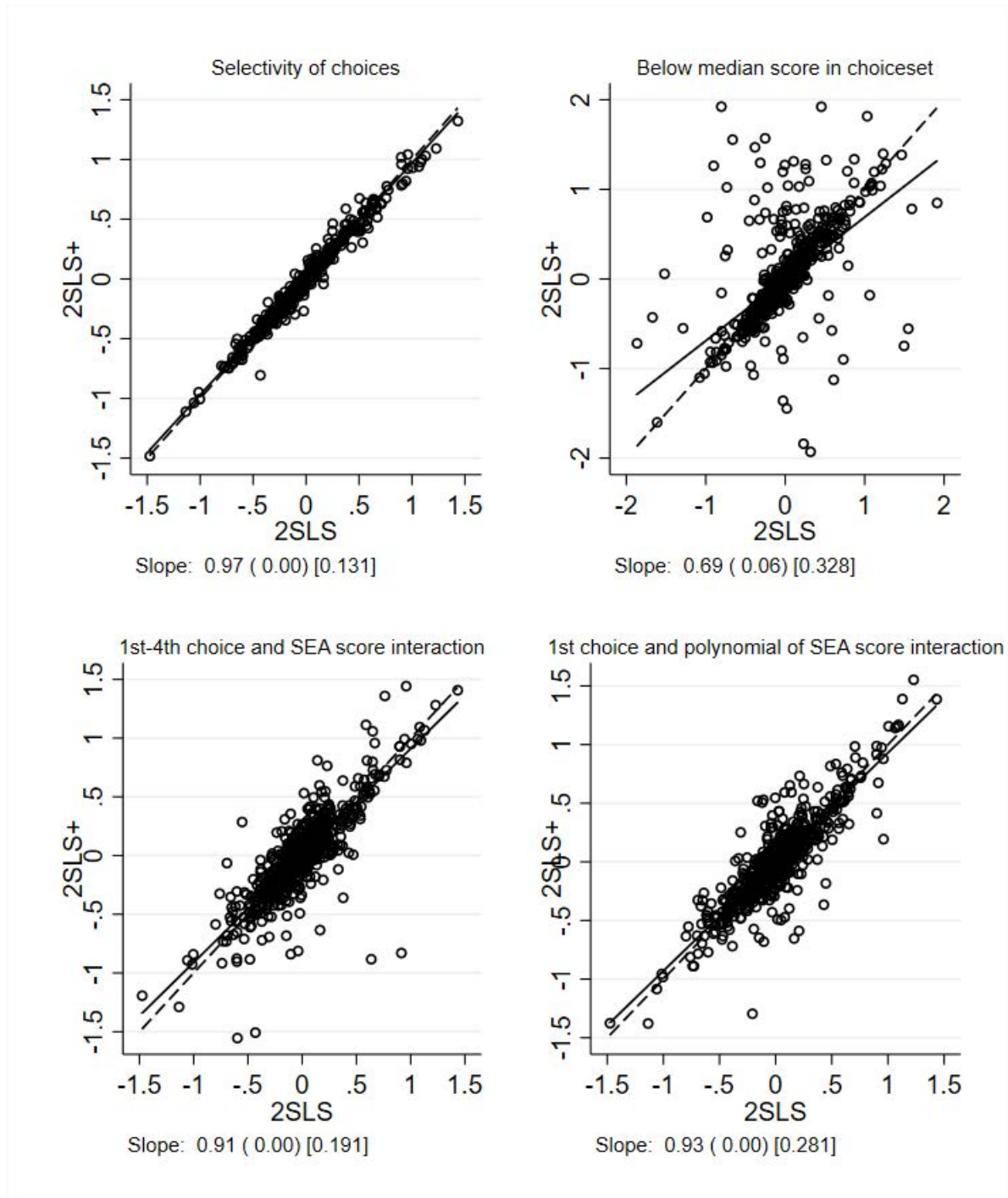
$$Y_{ijct} = \Sigma(I_{i,j} \cdot \theta_j^{TOTIV}) + f(SEA_i) + \Sigma(g_k(SEA_i) \cdot I_{i,k}) + \lambda_c + \mathbf{X}'_{it}\delta + C_t + \varepsilon_{ijct} \quad (28)$$

$$I_{i,j} = \Sigma(I_{i,\tau} \cdot \pi_{\tau j}) + f(SEA_i) + \Sigma(g_k(SEA_i) \cdot I_{i,k}) + \lambda_c + \mathbf{X}'_{it}\delta + C_t + v_{ijct}, \quad \text{for each } j \in J \quad (29)$$

Where $g_k(SEA_i)$ is a fifth order polynomial and $I_{i,k}$ is an indicator equal to 1 if the first choice of student i is school k .

Results for all four models are shown in [Figure C2](#)

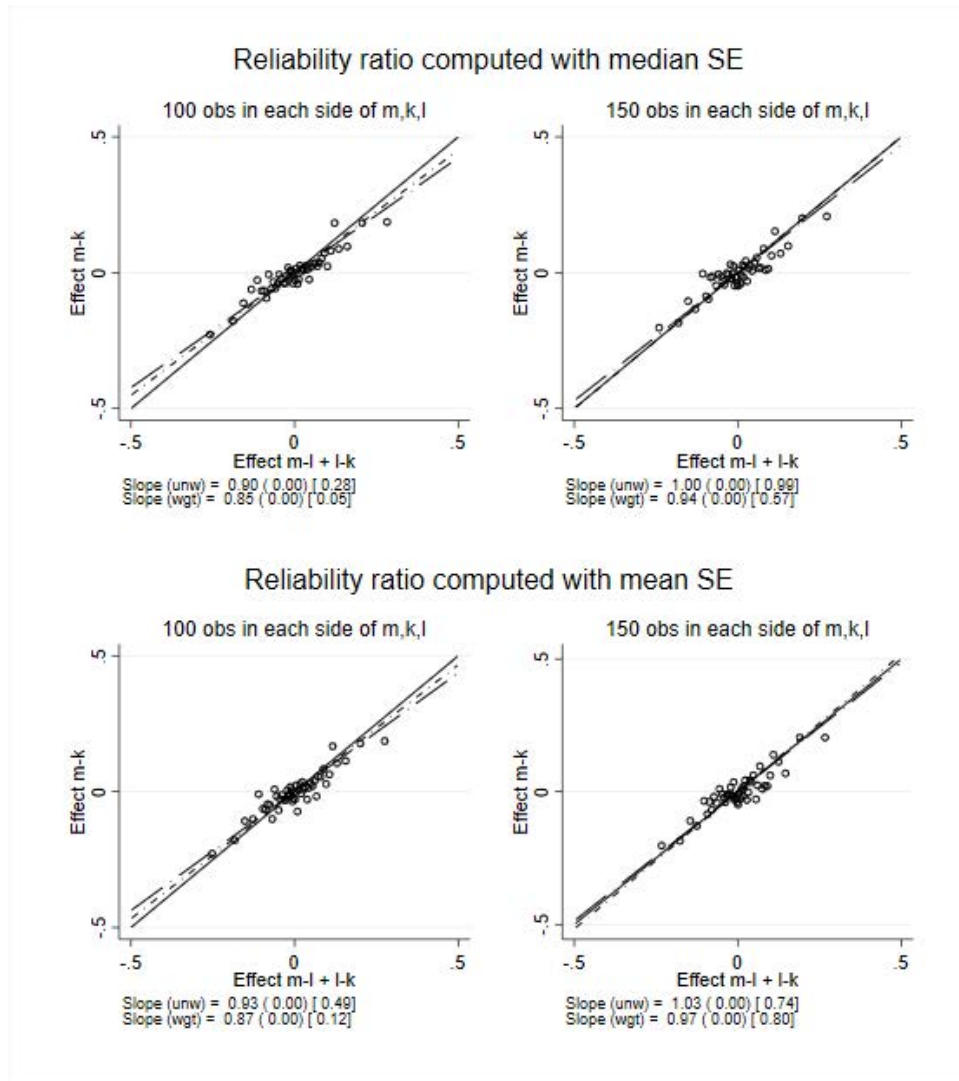
Figure C2: Additivity Test (2SLS): Scores and choice separability



Notes: We show the relationship between the estimated school impacts with the interacted model (y-axis) and the model without interactions (x-axis). Each panel shows results for one of the interacted models. In all panels, the solid line represents a linear fit; while the dashed line is the 45° line. The p -values associated with the null hypothesis of slope being 0 in parentheses, p -values of slope being 1 in brackets

Additivity of the School Effects

Figure C3: Additivity of the School Effects.



Notes: All panels display estimated effects on all outcomes combined. P-values of slope being 0 in (parenthesis), p-values of slope being 1 in [brackets]. The short dashed line shows the unweighted fit, the long dashed line the weighted fit and the straight line the 45° relation. The upper panel employs the median of the estimated standard errors of the school effects when computing the reliability ratio. The lower panel employs the mean of the estimated standard errors of the school effects when computing the reliability ratio.

Magnitude of school impacts

This section reports Table C2 which includes two models to estimate the size of the school effects. One uses only school level effects (left panel) and the other also includes school-year level effects (right panel).⁴⁹ The latter model allows us to estimate permanent ($\sigma_{\theta_j^{TOT}}$) and transitory school effects ($\sigma_{\mu_{jp}}$). Two alternative samples are used: a first sample that includes all students, and a second one that excludes those who attended schools with outlier school effects.

Table C2. Standard Deviation of School Impacts

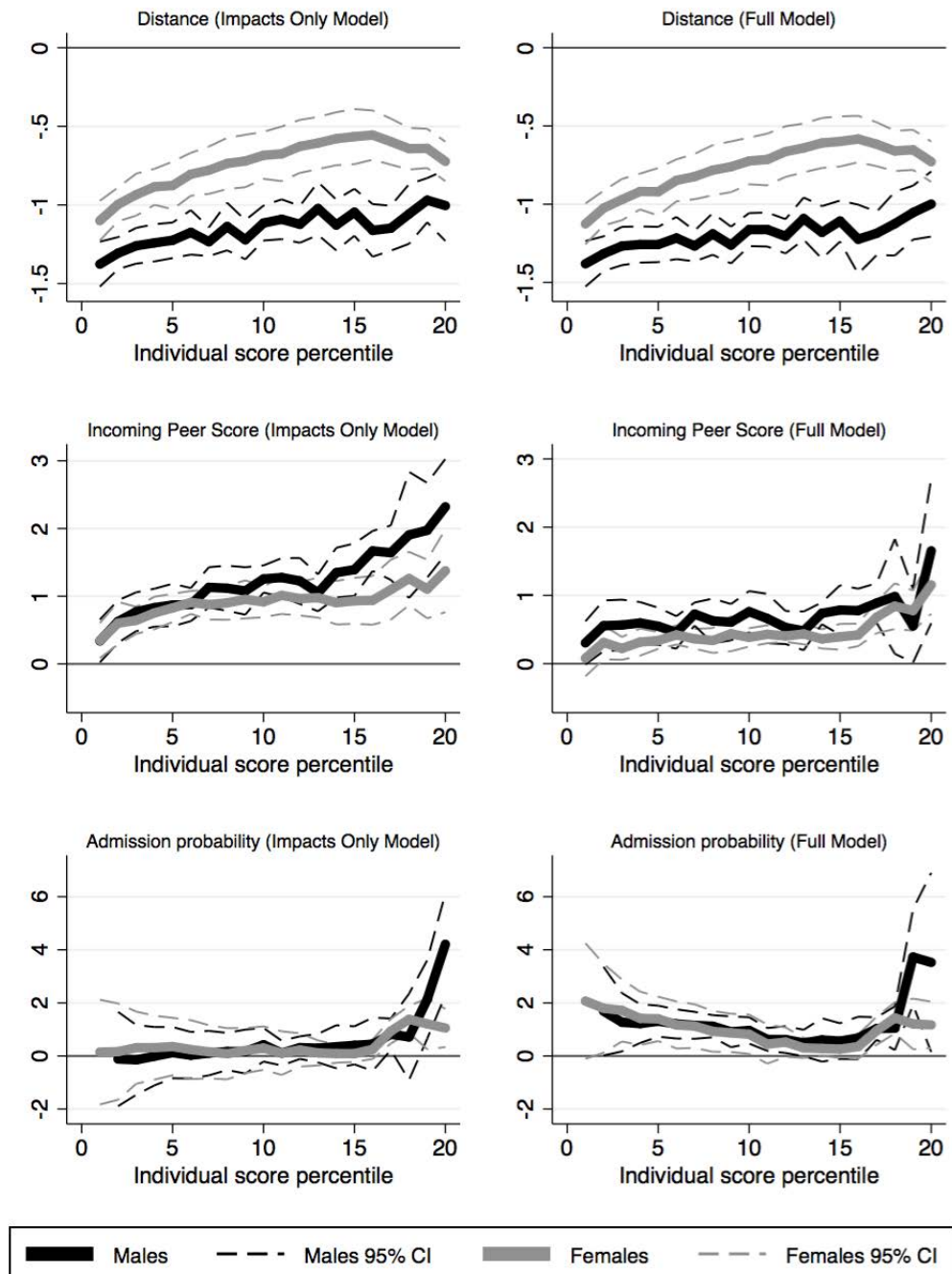
	School Level ($\sigma_{\theta_j^{TOT}}$)			School Level ($\sigma_{\theta_j^{TOT}}$)			School-Year Level ($\sigma_{\mu_{jp}}$)		
Outcome	Size of Impact	[95% CI]		Size of Impact	[95% CI]		Size of Impact	[95% CI]	
High-Stakes Index									
All Schools	0.398	0.353	0.449	0.393	0.347	0.446	0.108	0.095	0.124
Dropping Outliers	0.398	0.353	0.449	0.393	0.347	0.446	0.108	0.095	0.124
Low-Stakes Index									
All Schools	0.433	0.381	0.493	0.418	0.364	0.480	0.164	0.142	0.190
Dropping Outliers	0.427	0.375	0.486	0.412	0.359	0.474	0.149	0.129	0.171
No Dropout by 14									
All Schools	2.443	2.148	2.780	0.714	0.218	2.337	2.438	2.138	2.779
Dropping Outliers	0.124	0.108	0.143	0.114	0.097	0.135	0.065	0.054	0.078
Not arrested by 18									
All Schools	0.101	0.089	0.116	0.096	0.083	0.112	0.046	0.039	0.056
Dropping Outliers	0.093	0.082	0.106	0.089	0.076	0.103	0.042	0.035	0.051
No live birth by 19									
All Schools	1.154	1.003	1.327	1.140	0.989	1.315	0.217	0.182	0.260
Dropping Outliers	0.196	0.167	0.231	0.115	0.067	0.197	0.209	0.175	0.250
Formally employed									
All Schools	0.183	0.150	0.222	0.177	0.145	0.218	0.046	0.034	0.064
Dropping Outliers	0.089	0.075	0.105	0.082	0.067	0.100	0.046	0.033	0.063

Notes: This table reports estimates of the implied standard deviation of the permanent and transitory attended school impacts along with their respective 95 percent confidence intervals obtained from the maximum likelihood approach. The second line for each outcome shows estimates dropping students who attended schools with low compliance or with outlier value added (i.e. school impacts more than 4σ away from the median school). In the case of Low-Stakes Index, 2 outlier schools; for High-Stakes index there were no outlier schools; for no dropout by 14, 7 outlier schools; for no live birth by 19, 12 outlier schools; for not arrested by 18, 6 outlier schools; for formally employed, 6 outlier schools.

⁴⁹For each school, we estimate effects for two sets of years, even and odd years. These are the effects used in the model for the school-year level.

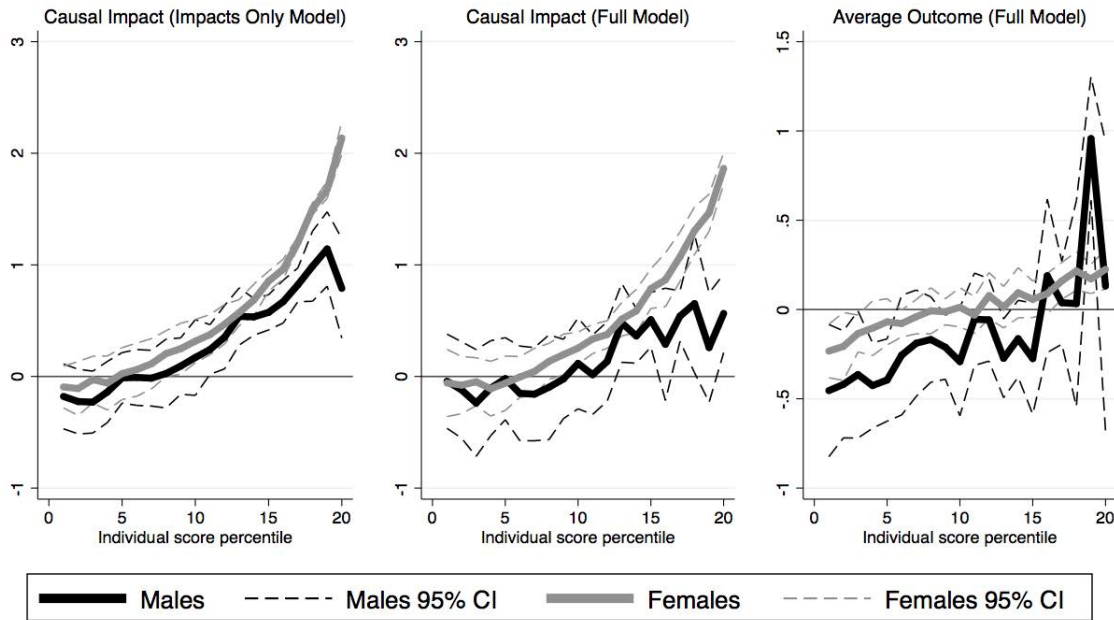
Appendix D: Choice Model using Out of Sample School Impacts

Figure D1: Distance, Peer Quality, and Admission Probability



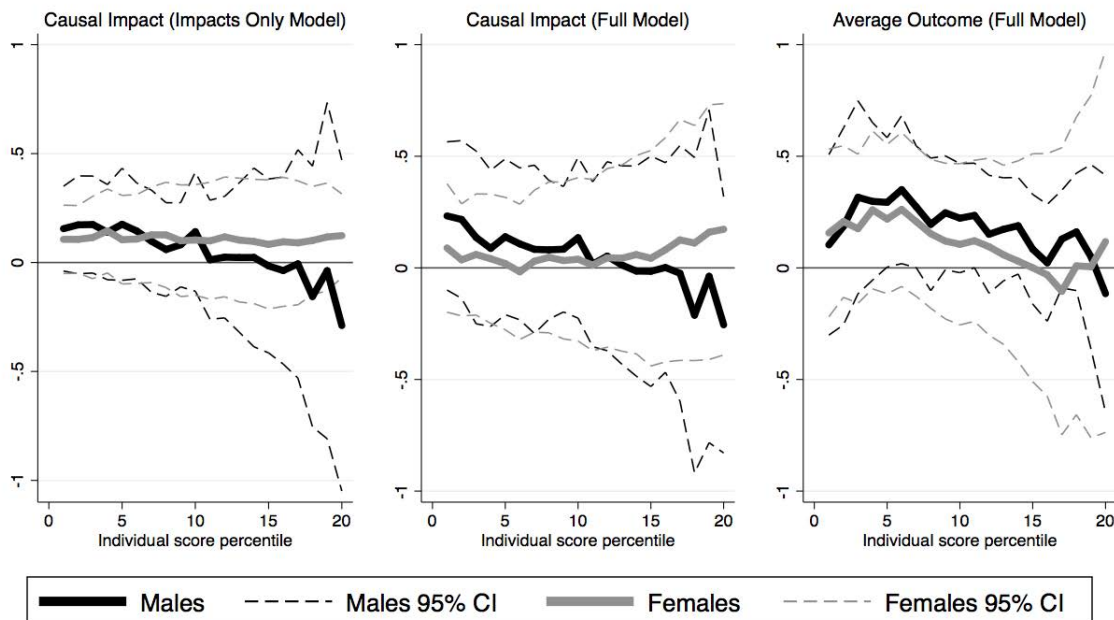
Notes: The connected lines represent the estimated coefficients, computed separately for each (SEA score ventile)×(gender) cell for two different models: Impacts Only Model (left panel), and Full Model (right panel). The dashed lines represent the associated 95% confidence intervals.

Figure D2: High-Stakes Index



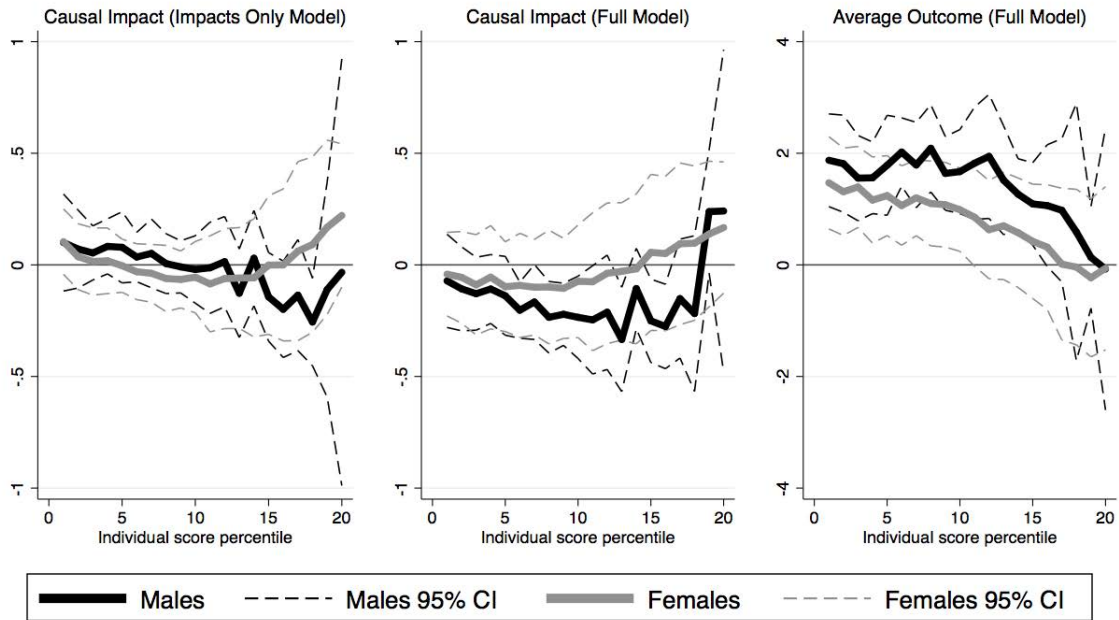
Notes: The connected lines represent the estimated coefficients, computed separately for each (SEA score ventile) \times (gender) cell for two different models: Impacts Only Model (left panel), and Full Model (middle and right panels). The dashed lines represent the associated 95% confidence interval.

Figure D3: Low-Stakes Index



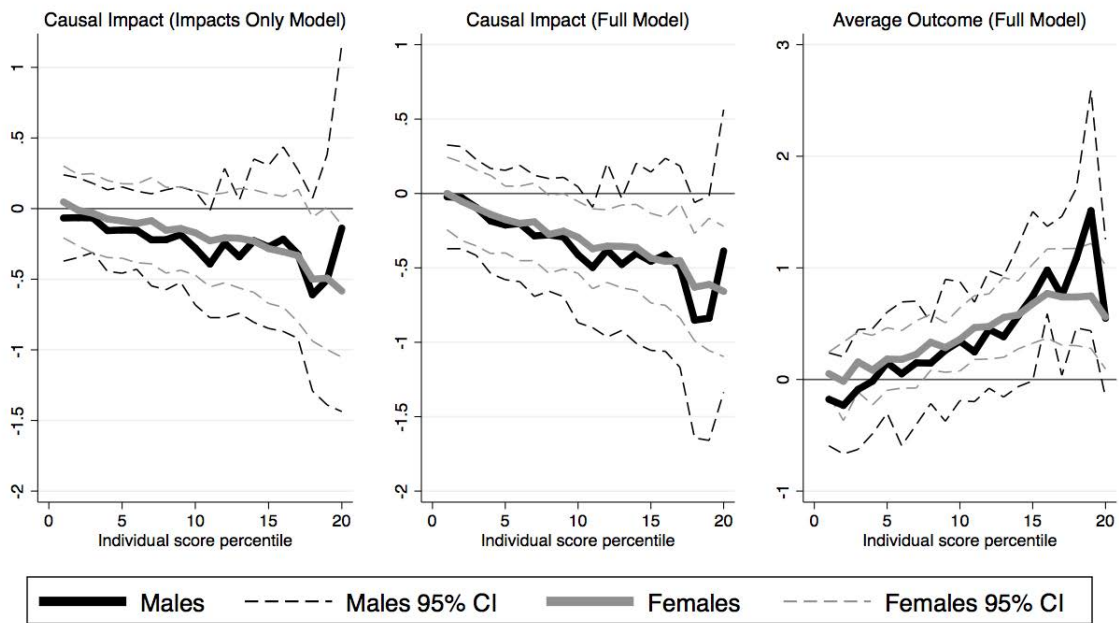
Notes: The connected lines represent the estimated coefficients, computed separately for each (SEA score ventile) \times (gender) cell for two different models: Impacts Only Model (left panel), and Full Model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure D4: No Dropout by Age 14



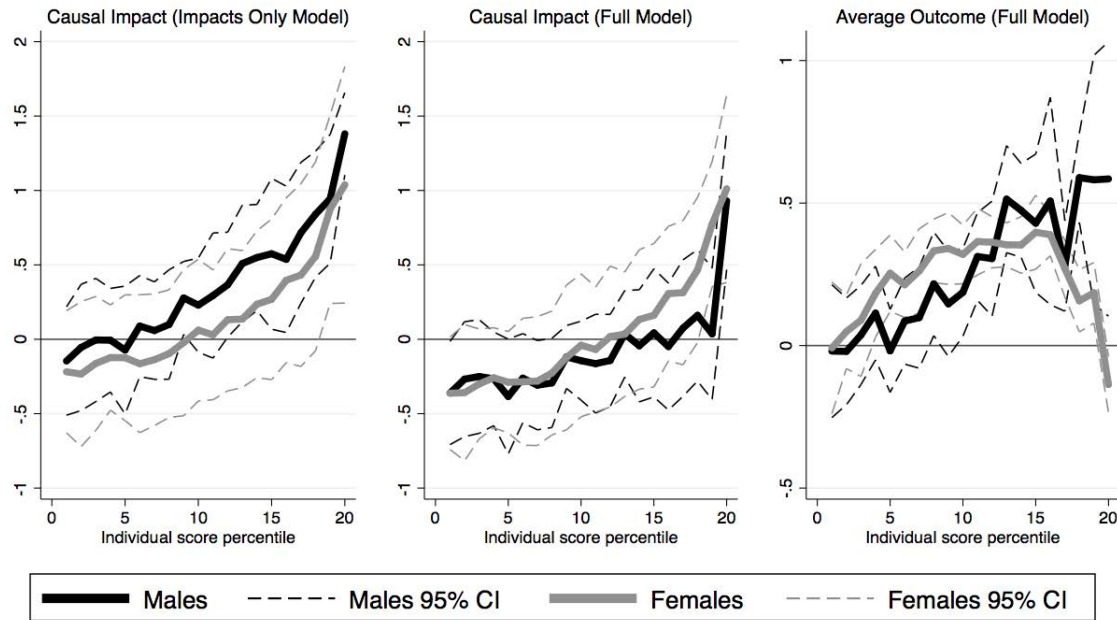
Notes: The connected lines represent the estimated coefficients, computed separately for each (SEA score ventile)×(gender) cell for two different models: Impacts Only Model (left panel), and Full Model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure D5: No Live Birth by Age 19



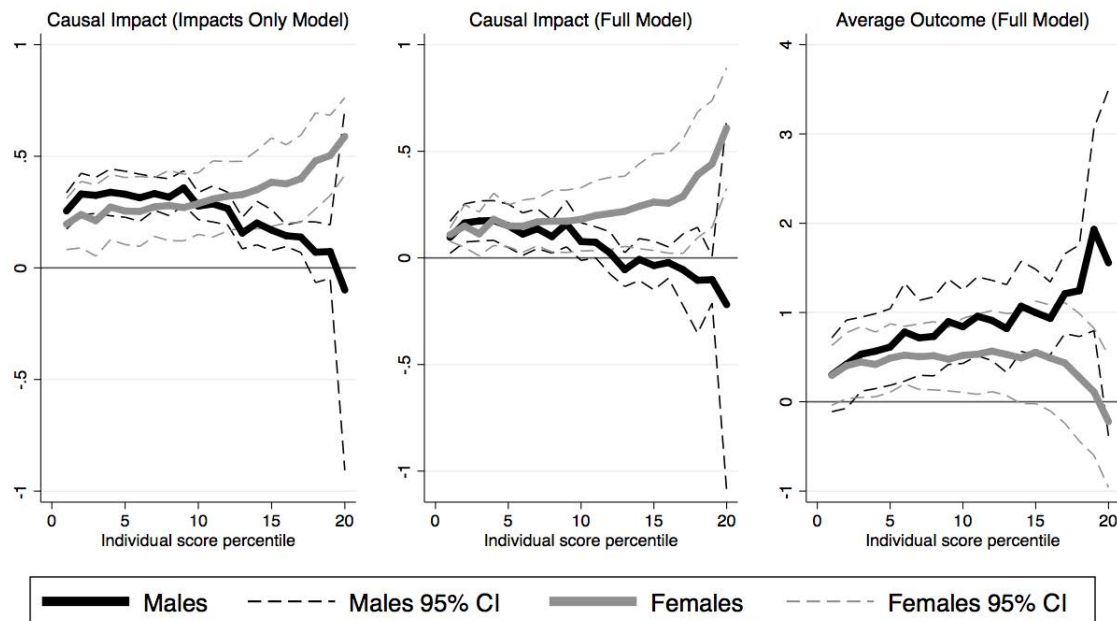
Notes: The connected lines represent the estimated coefficients, computed separately for each (SEA score ventile)×(gender) cell for two different models: Impacts Only Model (left panel), and Full Model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure D6: Not Arrested by Age 18



Notes: The connected lines represent the estimated coefficients, computed separately for each (SEA score ventile)×(gender) cell for two different models: Impacts Only Model (left panel), and Full Model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

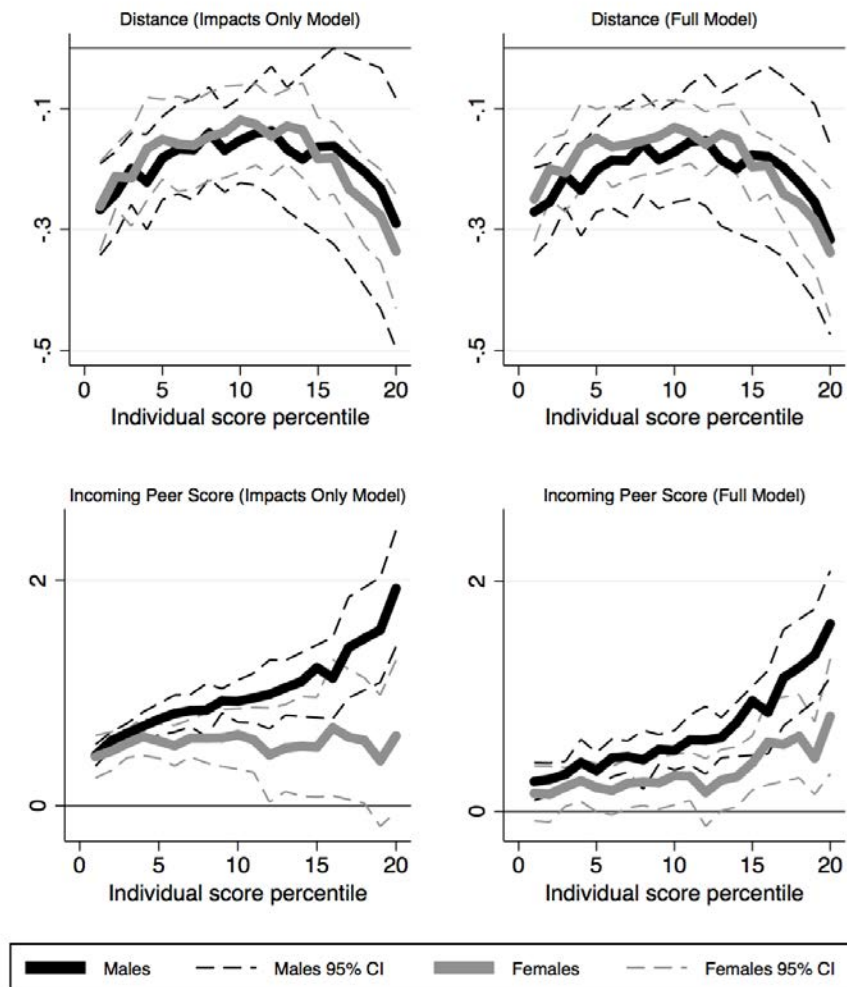
Figure D7: Formally Employed at 27+ Years Old



Notes: The connected lines represent the estimated coefficients, computed separately for each (SEA score ventile)×(gender) cell for two different models: Impacts Only Model (left panel), and Full Model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

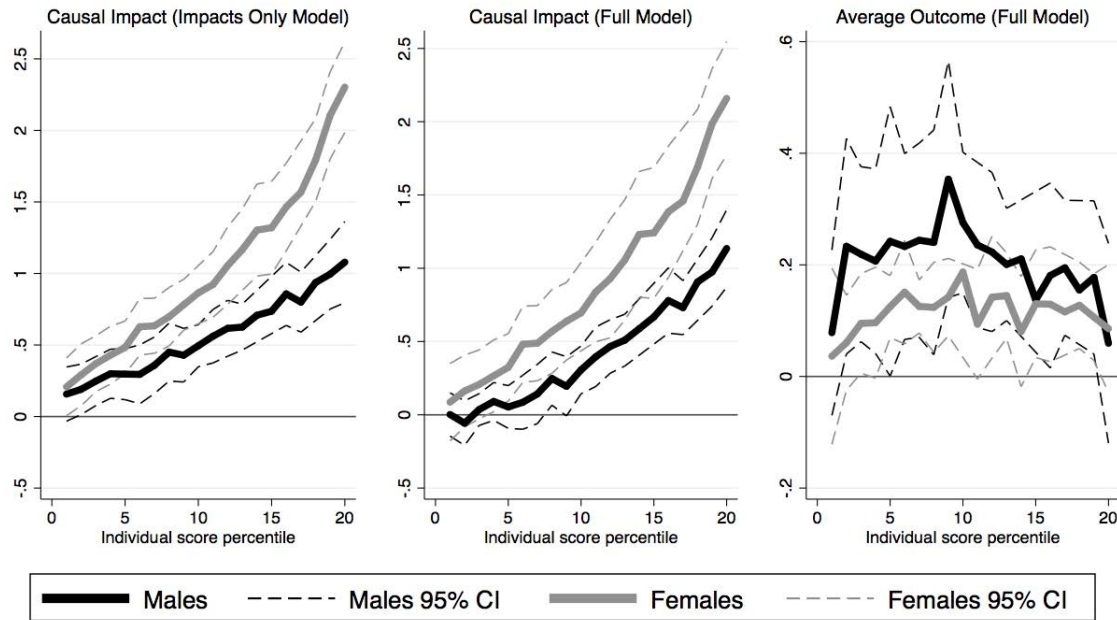
Appendix E: Choice Model using Rank Ordered Logit

Figure E1: Distance and Peer Quality



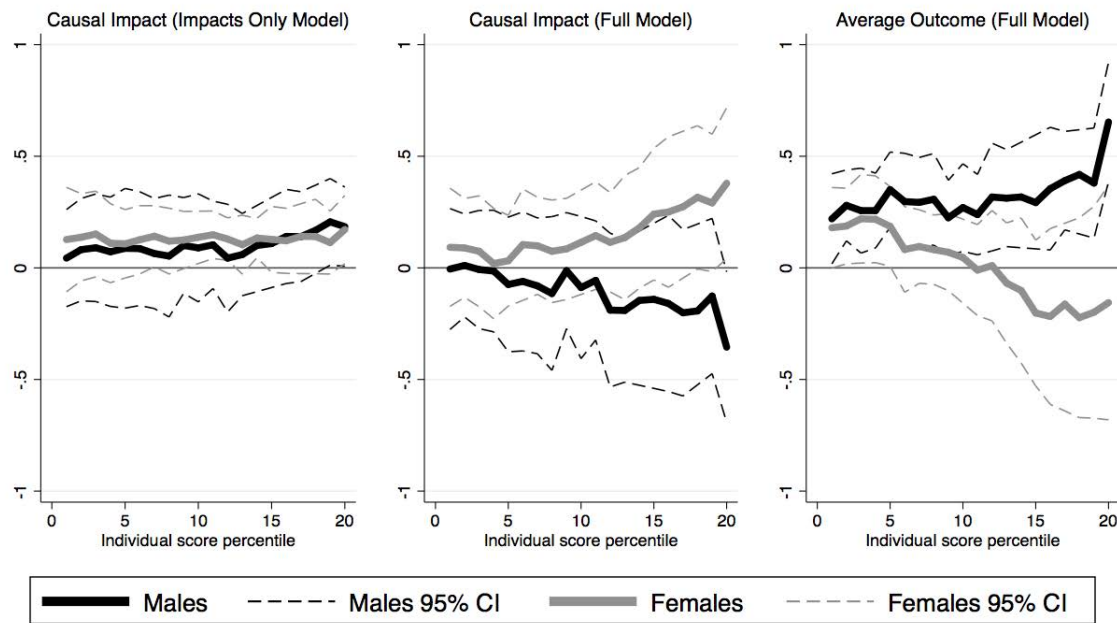
Notes: The connected lines represent the estimated coefficients, computed separately for each (SEA score ventile) \times (gender) cell for two different models: Impacts Only Model (left panel), and Full Model (right panel). The dashed lines represent the associated 95% confidence intervals.

Figure E2: High-Stakes Index



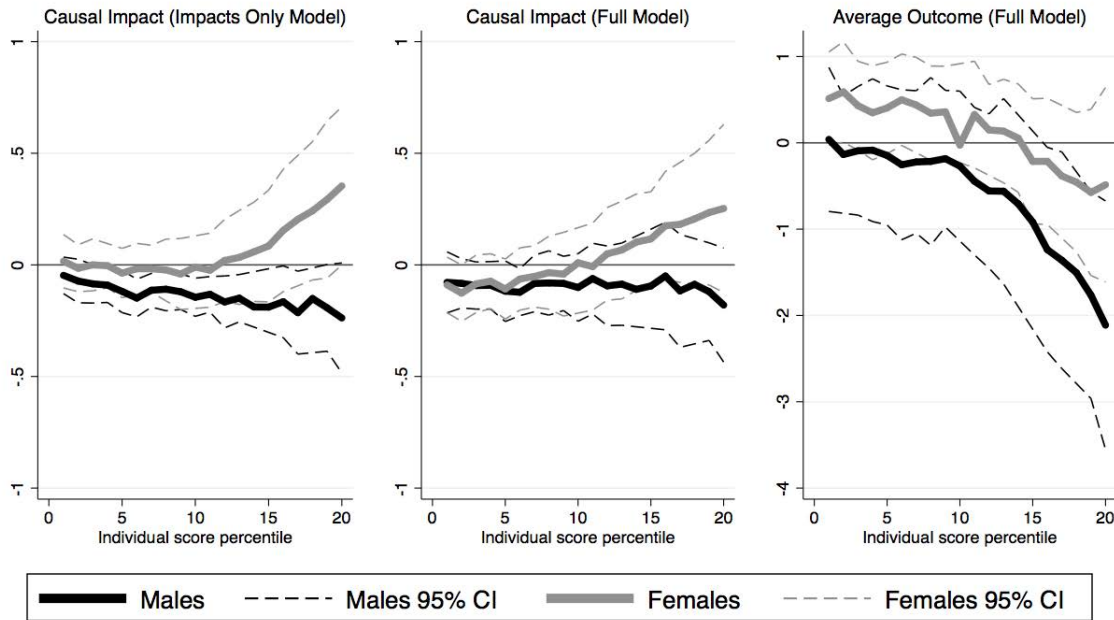
Notes: The connected lines represent the estimated coefficients, computed separately for each (SEA score ventile) \times (gender) cell for two different models: Impacts Only Model (left panel), and Full Model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure E3: Low-Stakes Index



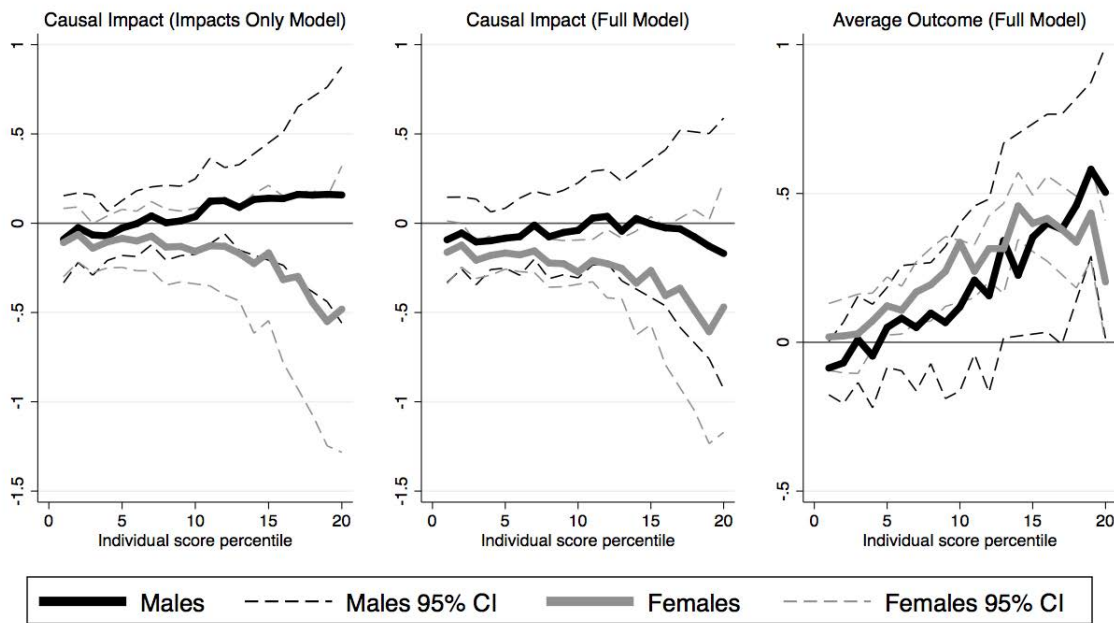
Notes: The connected lines represent the estimated coefficients, computed separately for each (SEA score ventile) \times (gender) cell for two different models: Impacts Only Model (left panel), and Full Model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure E4: No Dropout by Age 14



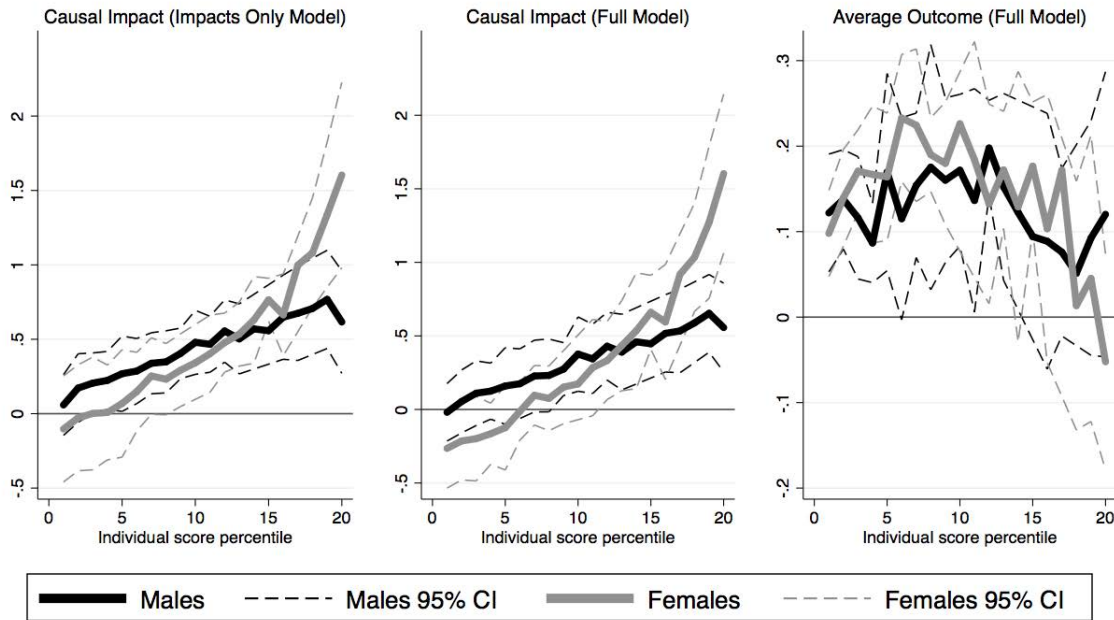
Notes: The connected lines represent the estimated coefficients, computed separately for each (SEA score ventile) \times (gender) cell for two different models: Impacts Only Model (left panel), and Full Model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure E5: No Live Birth by Age 19



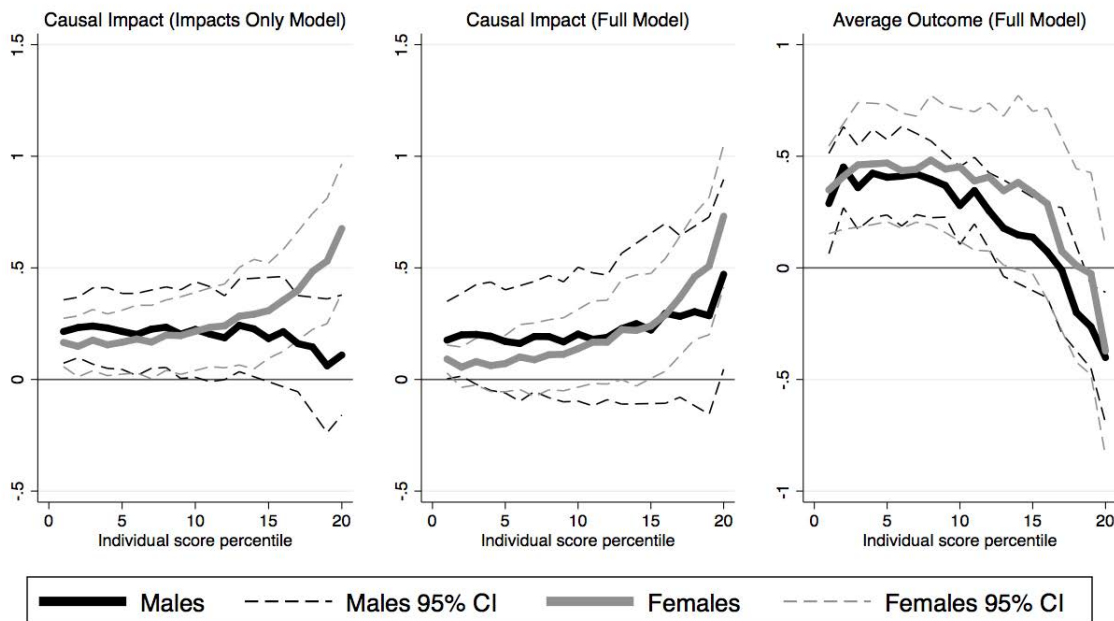
Notes: The connected lines represent the estimated coefficients, computed separately for each (SEA score ventile) \times (gender) cell for two different models: Impacts Only Model (left panel), and Full Model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure E6: Not Arrested by Age 18



Notes: The connected lines represent the estimated coefficients, computed separately for each (SEA score ventile)×(gender) cell for two different models: Impacts Only Model (left panel), and Full Model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure E7: Formally Employed at 27+ Years Old



Notes: The connected lines represent the estimated coefficients, computed separately for each (SEA score ventile)×(gender) cell for two different models: Impacts Only Model (left panel), and Full Model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.