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
Northwestern University

Francisco Pardo
Inter-American Development Bank

October 18, 2019

*Beuermann: Inter-American Development Bank, 1300 New York Avenue, NW, Washington DC 20577 (e-mail: dietherbe@iadb.org); Jackson: Northwestern University, 2120 Campus Drive, Evanston IL 60208 (e-mail: kirabo-jackson@northwestern.edu); Navarro-Sola: Northwestern University, Department of Economics, 2211 Campus Drive, Evanston IL 60208 (e-mail: navarrosola@u.northwestern.edu); Pardo: Inter-American Development Bank, 1300 New York Avenue, NW, Washington DC 20577 (e-mail:franciscopa@iadb.org). 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 assembly the administrative datasets used in the study. We are indebted to Chief Education Officer Harrilal Seecharan of the Trinidad and Tobago Ministry of Education for his continuous support. We would also like to thank Ria Boafo, Lisa Henry-David, Shalini Maharaj, Brenda Moore, and Peter Smith of the Trinidad and Tobago Ministry of Education for granting the facilities to access the educational data needed for the study, their assistance, and their generosity. We would like to 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 the access to the 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 formal employment participation records maintaining individual confidentiality. Tatiana Zarate and Diego Zuniga provided excellent research assistance. This paper benefited from comments by seminar participants at the NBER, MIT, Zurich, Lund, and Wharton. The statements made and views expressed are solely the responsibility of the authors.

Abstract

Is a school's impact on high-stakes test scores a good measure of its overall impact on students? Do parents value school impacts on tests, longer-run outcomes, or both? To answer the first question, we exploit quasi-random school assignments and data from Trinidad and Tobago. We construct exogenous instruments for *each* individual school and estimate the causal impacts of individual schools on several short- and longer-run outcomes. Schools' impacts on high-stakes tests are weakly related to impacts on low-stakes tests, dropout, crime, teen motherhood, and formal labor market participation. To answer the second question, we link estimated school impacts to parents' ranked lists of schools. We propose a modified mulitnomial logit model that allows one to infer preferences for school attributes even in some settings where choices are strategic. Parents of higher-achieving students value schools that improve high-stakes test scores conditional on average outcomes, proximity, and even peer quality. Parents also value schools that reduce crime and increase formal labor market participation. Most parents' preferences for school impacts on labor-market and crime outcomes are, as strong, or stronger than those for test scores. These results provide a potential explanation for recent findings that parent preferences are not strongly related to test-score impacts. They also suggest that evaluations based solely on test scores may be very misleading about the welfare effects of school choice. (JEL I20, J0)

I Introduction

Is a school's causal impact on test scores a good measure of its overall impact on students? Do parents value schools that have positive causal impacts on high-stakes standardized tests? Do parents value school effectiveness 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. To address the second and third questions, we merge our estimated causal impacts to parents' school rankings and explore whether parents value schools causal impacts on these multiple dimensions – providing the first exploration into whether parents value school effectiveness 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), 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 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 lead to greater efficiency in educational production (Friedman 1955; Chubb and Moe 1990). However, if parents are unable to discern school effectiveness, 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 effectiveness on test scores (MacLeod and Urquiola 2018; Beuermann and Jackson 2018). The few studies that directly examine parent preferences for school effectiveness (as opposed to peer quality or average outcomes) conclude that parents do not value school effectiveness *per se* (Rothstein 2006, Abdulkadiroğlu

[1] We 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 (2018) 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 that 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. Relatedly, 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 fully 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."*

et al. 2019b) – casting doubt on the likely efficacy of school choice. However, this may not be the final word on this issue for two reasons. First, existing work linking parent choices to measures of school effectiveness rely on strong selection on observables assumptions and may therefore not uncover individual schools' causal impacts.[3] Second, 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 value schools that have causal impacts on test scores and *also* key nonacademic outcomes.

We use data on the full population of 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, we implement an Instrumental Variables (IV) model similar to Kirkeboen et al. (2016). 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). We implement several tests to validate our approach.

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 with limited choices (as in our context), applicants may engage in strategic behavior by accounting for the likelihood of admission when making choices (Chade and Smith, 2006). Because the nature of any strategic choices is known, and we can obtain estimates of ad-

[3]For example, Rothstein (2006) defines "effectiveness" as the between-school differences in student performance that cannot be attributed to differences in student body composition. Abdulkadiroğlu et al. (2019b) control for variables such that the school assignment is random, but estimate impacts of the attended school so that *"noncompliance with the assignment mechanism, are presumed to be unrelated to potential outcomes."* In neither of these two cases is estimated school effectiveness based on quasi-random variation so that the estimates may not reflect schools' causal impacts.

mission probabilities using historical data, we are able to condition on admission probabilities and uncover true preferences in our context. We implement a modified multinomial logit model (McFadden, 1973) to estimate parent preferences for schools based on observed peer quality, proximity, average school characteristics, admission probabilities, and estimated school impacts.

Schools have meaningful effects on an array of outcomes. Going from a school at the median of the impact distribution to one at the 85^{th} 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. Notably, the correlations between school impacts on high-stakes tests and other outcomes are modest. For example, the correlation between school impacts on high-stakes and non-dropout is 0.10, and the correlation between a school's impact on high-stakes tests and being formally employed is 0.08. Based on additional empirical tests, we find that these low correlations are not mainly due to measurement errors in the school estimates, or different schools enrolling different student populations who may be marginal for different outcomes. Rather, the patterns suggest that school output is multidimensional such that (a) schools that improve academic skills are not necessarily those that improve broader adult well-being (which parents may value), and (b) the schools that improve the outcomes valued by some parents may not be the schools that improve the outcomes valued by others.

In our analysis of parental preferences, we replicate several key results of existing studies. Parents express greater preference for proximity to home, better peers, and higher-performing schools (Burgess et al. 2015; Abdulkadiroğlu et al. 2019b). We also find that parents value schools that are safe (as measured by fewer students who are arrested as teens) and schools with lower teen live birth rates. This 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)). As in Hastings et al. (2006), we find that preferences for high-achieving schools are stronger for parents of higher-achieving students. Looking at test-score impacts, parents of higher-achieving children have strong preferences for schools with high causal effects on high-stakes exams. This pattern is robust to controls for peer quality and average outcomes– suggesting that some parents *can and do* disentangle effective schools from schools with strong performance. Our results are not entirely at odds with Abdulkadiroğlu et al. (2019b) because conditional on average outcomes and peer quality, only parents of the most able 25 percent of students exhibit strong preferences for schools that improve high-stakes exam scores.

A key innovation of our work is to measure preferences for schools' causal impacts on nonacademic outcomes. As with high-stakes exams, parents of high-achieving students are sensitive to school effectiveness on (reduced) arrests. This preference for arrest value-added is stronger for parents of males. Indeed, parents of males at almost all achievement levels value a one standard deviation increase in teen arrest value-added more than that for high stakes exams. Unlike preferences for crime or exams (which are only present for parents of high-achieving children), we find strong preferences for schools that raise formal labor market participation among all groups (including low-achieving children). Parents of males in the bottom 70 percent of the incoming achievement distribution (i.e., most males) value a one standard deviation increase in labor-market value added more than that for high-stakes exams. This is also true for parents of females in the bottom half of the incoming achievement distribution. Our results indicate that most parents value school impacts on non-academic outcomes *more* than they value test score impacts. Given that schools that 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 contribute to existing literatures in many ways. We build on the school quality literature by examining school impacts on high-stakes tests, low-stakes tests, dropout, teen motherhood and labor-market participation simultaneously. We present the first analysis of the relationships between school impacts on all these different outcomes – providing direct evidence of the multidimensionality of school output.[4] Our findings have important policy implications because test-score measures of school quality (which are not strongly related to impacts in other dimensions) are increasingly used for policy decisions. We also contribute to the work on parental preferences. First, we provide the first study of parental preferences for school effectiveness on non academic outcomes such as fertility, crime, and labor market participation. Second, we rely on clean exogenous variation to identify school's causal impacts. By using estimates of schools' causal impacts on a wide set of outcomes, we shed new light on parental preferences for schools' causal impacts on both academic and nonacademic outcomes. We show that parents 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.[5] Regarding the school choice literature, our results suggest that evaluations of school choice based solely on test scores 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.

[4] Dobbie 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.

[5] This main point is also suggested in Beuermann and Jackson (2018) who examine the short and long run impacts of attending a preferred secondary school in Barbados.

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 are two types:[6] 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 158 public secondary schools during our study period. Among these, 44 were Government Assisted schools. 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 students in eight subjects.[7] NCSE performance has no consequences in terms of school progression or admission to tertiary education institutions. Therefore, the NCSE is 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 language 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 subjects approved. Consequently, CSEC performance is a key determinant of both progression into tertiary education and future employment prospects. The CSEC is a high-stakes examination. 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 accepted as a general admission requirement to British higher education institutions. The post-secondary qualification of a CAPE Associate's Degree is awarded after passing seven CAPE units including Caribbean and Communication Studies. Finally, students

[6] Private secondary schools serve a very small share of the student population (about 3.4 percent).

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

who obtain the highest achievable grade in eight CAPE units (including Caribbean and Communication Studies) 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 the full population of 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, the census tract of residence, religious affiliation, SEA scores, the ranked list of secondary schools preferences, 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.[8] 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.[9] The CAPE data are available for years 2005 through 2016, and are linked to the 1999 through 2009 SEA cohorts.[10] 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 of Trinidad and Tobago. These data include all persons who were formally employed and, therefore, contributing to the na-

[8]We matched 97.44 percent of all NCSE individual records to the SEA data.

[9]We matched 96.31 percent of all CSEC individual records to the SEA data. The non-match rate of 3.69 percent closely mimics the share of students served by private schools (3.4 percent) who would not have taken the SEA.

[10]We matched 96.6 percent of all CAPE individual records to the SEA data.

tional social security system as of May 2017. For each affiliate, the data include the full current name, full original name prior to marriage or any other name changes, date of birth, and gender. To explore formal employment among individuals aged between 27 and 33 years old, 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, 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

We first seek to estimate the causal impacts of individual schools on academic, nonacademic and social outcomes. Consider modelling the outcomes of student i who attended school j as below.

$$Y_{ij} = \Sigma(I_{i,j} \cdot \theta_j^{TOT_0}) + \varepsilon_{ij} \tag{1}$$

In (1), Y_{ij} is the outcome of interest for student i who attended school j. $I_{i,j}$ is an indicator equal to one if student i attended school j and zero otherwise, and ε_{ij} is a mean-zero random error term. The OLS estimate from (1), $\hat{\theta}_j^{TOT_0}$, is the average outcome for all students i who attended school j. When these school-level averages are standardized to be mean zero, $\hat{\theta}_j^{TOT_0}$ is the average outcome for all students who attended school j relative to those who attended the average school. If (a) students were randomly assigned to schools, *and* (b) students who comply with the random assignment are similar in unobserved ways from those who do not (i.e. have similar potential outcomes), then $\hat{\theta}_j^{TOT_0}$ would be an unbiased estimate of the causal impact of attending school j relative to the average school. Because neither of these two conditions is likely to hold empirically, we discuss how we address both of these issues in our identification strategy below.

III.1 Relying on Conditional Random Assignment

Assuming that compliance with the random assignment is random (a nontrivial assumption that we address below), one can uncover the causal impact of attending school j relative to the average school if one conditions on a set of variables such that the school assignment is random. 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. Specifically, 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 filed. 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 now be in the applicant pool for their second choice. This process continues until all school slots are filled or all students are assigned.[11] We refer to this purely 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 administra-

[11] See 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.

tor 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. 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.

III.1.1 Identifying Variation

This assignment mechanism generates two sources of exogenous variation (Jackson 2010; Angrist and Rokkanen 2015). To see this, consider this scenario: 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 (see Figure 1). Among those with school 1 as the top choice, 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, 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.

Figure 2 illustrates the discontinuity in preferred school attendance as one's score goes from below to above the cutoff. To show this in a single figure, we recenter incoming scores on the cutoff for the preferred school, and stack the applicant pools to all schools across all SEA cohorts. While compliance with the algorithm-based assigned school is not perfect, there is a sharp discontinuity in preferred school attendance as a result of scoring above the school assignment cutoff. We show in Appendix A that, conditional on smooth functions of incoming SEA scores, scoring above the school assignment cutoff is a credible source of exogenous variation in attending a preferred school. 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.

Importantly, the RD variation is not all the identifying variation that is embedded in the assignment mechanism. To see this, 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 93, the two choice groups 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. Concretely, 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. With different groups of students with different choices, one can estimate impacts for other sets of schools (say schools 3, 4, and 5). So long as there is sufficient overlap in the schools listed across the several choice groups, one can compare each school to every other school.[12] This is the approach we employ. This logic was first laid out in Jackson (2010) and is similar to that employed in Kirkeboen et al. (2016). We show that our estimates are robust to relaxing this additive assumption in Section III.1.4.

III.1.2 Relying Only on the Identifying Variation

Based on the algorithm, two students with the same set of school choices will only be tentatively assigned to different schools because one scored above the cutoff for a desired school while the other did not. As such, conditional on smooth flexible functions of incoming SEA scores, students initial tentative assignments are as good as random if locations of the test score cutoffs are exogenous to other student characteristics. In Appendix A we present the standard battery of empirical tests to support the exogeneity of the test score cutoffs. Also among students with the same test scores, some may be assigned to different schools due to differences in their choices. Using these two sources of variation, so long as choices and scores are additively separable, the average outcomes of all students tentatively assigned to a given school (conditional on their school choices, sex, district fixed effects, religion fixed effects and smooth functions of their incoming SEA scores) provide a credible estimate of the causal effect of *being tentatively assigned* to that school.

One can obtain such school-level averages by estimating (2) by Ordinary Least Squares (OLS).

$$Y_{i\tau ct} = \Sigma(I_{i,\tau} \cdot \theta_{\tau}^{ITT}) + f(SEA_i) + \lambda_c + \mathbf{X}'_{\mathbf{it}}\delta + C_t + \varepsilon_{i\tau ct} \quad (2)$$

In (2), $Y_{i\tau ct}$ 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 *tentatively*

[12]Suppose 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 (see Mansfield 2015).

assigned to school τ. Attended school j and assigned school τ are the same for those who comply with the quasi-random assignment. $f(SEA_i)$ is a 5th-order polynomial of the incoming SEA score. **Xit** is a vector of individual-level baseline characteristics (measured at SEA registration) including sex, district of residence fixed effects, and religion fixed effects. C_t denotes SEA cohorts 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.[13] The choice-group indicators allow for the difference-in-difference identification away from the cut-offs (outlined above).[14] The estimated $\hat{\theta}_{\tau}^{ITT}$s from (2) 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. Because not all students who are tentatively assigned to school τ actually attend school τ, this is an Intent-to-Treat (ITT) estimate.

Under the identifying assumptions, the deferred acceptance *assigned* school is conditionally unrelated to unobserved characteristics of the student. But because not all students attend the school to which they are assigned, it is not the case that conditional on choices and test scores, the school *attended* is unrelated to unobserved characteristics of the student. However, if compliance with the initial assignment were random (or perfect), then one could replace the assigned school indicator with an attended school indicator and the resulting estimate would be a valid Treatment-on-the-Treated (TOT) estimate.[15] This is the approach taken in Abdulkadiroğlu et al. (2019b) where "*noncompliance with the assignment mechanism, are presumed to be unrelated to potential outcomes.*" We refer to this estimate as $\hat{\theta}_{j}^{TOT_{OLS}}$. To address concerns that complaince may be non-random, we employ an instrumental variables approach that builds on the Abdulkadiroğlu et al. (2019b) approach and *also* accounts for non-random compliance with the initial assignment.

[13]In most years, students listed up to 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.

[14]In principle one could efficiently rely only on the discontinuity variation by using the "tie breaking" approach proposed in Abdulkadiroğlu et al. (2019a). 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.

[15]That is one could estimate,

$$Y_{ijct} = \Sigma(I_{i,j} \cdot \theta_{j}^{TOT_{OLS}}) + f(SEA_i) + \lambda_c + \mathbf{X}_{\mathbf{it}}'\delta + C_t + \varepsilon_{ijct} \tag{3}$$

If one assumes random compliance, then $\hat{\theta}_{j}^{TOT_{OLS}}$ from (3) will be an unbiased TOT estimated effect.

III.1.3 Using Instruments to Account for Nonrandom Compliance

To obtain clean school impacts that are not biased by nonrandom compliance we use the rule-based school assignments as instruments for actual school attendance.[16] 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.[17] While not perfect, 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:

$$Y_{ijct} = \Sigma(I_{i,j} \cdot \theta_j^{TOT_{IV}}) + f(SEA_i) + \lambda_c + \mathbf{X}'_{\mathbf{it}}\delta + C_t + \varepsilon_{ijct} \tag{4}$$

$$I_{i,j} = \Sigma(I_{i,\tau} \cdot \pi_{\tau j}) + f(SEA_i) + \lambda_c + \mathbf{X}'_{\mathbf{it}}\delta + C_t + \upsilon_{ijct}, \quad \text{for each } j \in J \tag{5}$$

The endogenous variables in the second stage equation (4) 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 the time of writing the CSEC exams. 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 (5). Since the $\hat{\theta}_j^{TOT_{IV}}$ 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' casual 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-

[16]Noncompliance with the initial assignment is possible in our setting 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 $\hat{\theta}_j^{TOT_{OLS}}$ biased.

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

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.1.4 Validating The Instrumental Variables Estimates

Our estimation approach assumes that the impacts of choices and scores are approximately additive. To assuage concerns that our estimates are sensitive to this assumption, we validate our school estimates using only the local RD variation through the cutoffs that do not rely on the additivity assumption. Specifically, for each school in each year, we can estimate the RD impact 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 can also obtain an estimate of the impact of scoring above the cutoff on the "value-added" of the attended school relative to that of the same set of counterfactual schools. If our school IV estimates reflect the causal impact of attending school j relative to the average school, then the RD impact on actual outcomes should be similar to the RD impact on our estimated IV school impacts.[18] To show this, we compare the RD impacts on actual outcomes to the RD impacts on the estimated IV school impacts. Each estimate is an RD effect of scoring above the cutoff for each school in each year (using the optimal bandwidth from Imbens and Kalyanaraman (2012)). Following Hastings et al. (2015), to account for noisiness in the RD estimates, we weight each estimate by the inverse of its squared standard error. The scatter-plot is shown in Figure 3. The estimated slope coefficient is 1.01, and the test that this is different from 1 yields a p-value of 0.974. In contrast to the IV estimates, the slope for the OLS school estimates (as used in Abdulkadiroğlu et al. (2019b)) is 1.19 (although one cannot reject that this slope is equal to 1). Because the RD validation may be a noisy test, we also assess whether the OLS school estimates are consistent with the IV school estimates by regressing the latter on the former. This yields an estimated slope coefficient of 1.18 and one can reject that the OLS and IV school estimates are the same in expectation.[19] This indicates that (a) our IV estimates are valid under the weaker RD identifying assumptions and (b) the non-random compliance with the initial assignment leads to

[18] A similar test was implemented in Hastings et al. (2015) and Beuermann and Jackson (2018). 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.

[19] These two scatter-plots can be seen in Appendix Figure C1. Any estimation error in the OLS estimates would attenuate the slope toward zero, so that 1.18 is likely an *upper* bound.

bias in the OLS estimates. This underscores the importance of exploring whether parents value unbiased estimated school impacts.

IV Magnitude of the School Effects

To assess the magnitude of the school impacts on each outcome, we estimate the standard deviation of these impacts. 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 (6), we uncover Maximum Likelihood estimates of the variance of the persistent school impacts ($\sigma^2_{\theta_j^{TOT}}$) and of the transitory school impacts ($\sigma^2_{\mu_{jp}}$). Table 2 reports estimates of the standard deviation of the persistent school impacts for each outcome along with their 95 percent confidence intervals. To be conservative, we exclude outlier schools with estimated impacts lying 4σ away from the median school.[20]

$$\begin{bmatrix} \theta_j^{TOT} \\ \mu_{jp} \end{bmatrix} \sim N\left(0, \begin{pmatrix} \sigma^2_{\theta_j^{TOT}} I_J & 0 \\ 0 & \sigma^2_{\mu_{jp}} I_M \end{pmatrix}\right) \tag{6}$$

Low-stakes exams (age 14): 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.

High-stakes exams (age 16 through 19): The persistent school effects for the high-stakes dimension have a standard 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). It is also interesting to note that the magnitude of the school impacts on high-stakes and low-stakes tests are very similar. We will explore the extent to which these effects are correlated in Section IV.1 below.

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

[20]In Appendix C, we also show estimates of the transitory school impacts, and also estimates with and without removing outliers.

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 85^{th} 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 1σ 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.[21]

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. The estimates indicate that going from a school at the median to one at the 85^{th} percentile of the impact distribution would reduce teen live births by 3.5 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.[22] Given that the teen live birth rate is around 10 percent on average, these represent economically important relative impacts.

Crime: We present estimated school impacts on having any arrest by the age 18 in the lower panel (as opposed to the crime index in the top panel). Schools have meaningful and statistically significant impacts on arrests. The standard deviation of the persistent school effects is 0.016 which means that being assigned to a school at 85^{th} 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.

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). The estimates suggest that going from a school at the median of the impact distribution to one at the 85^{th} percentile would increase the likelihood of being formally

[21]The 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 2σ above the typical school.

[22]In 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 (2018) find that attending an elite school in Barbados decreases the teen motherhood rate by about 6 percentage points.

employed by about 3.6 percentage points. This effect is economically meaningful.

The fact that schools have economically meaningful effects 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 schools 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 reforms (e.g. No Child Left Behind) 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 school impacts on important dimensions of quality. The top panel of Table 3 presents raw correlations between estimated school impacts on the six outcomes. Note that the indexes were computed in a separate factor model so that these can be correlated. To aid interpretation, all outcomes are coded so that higher values reflect better outcomes. The correlations between school impacts across dimensions are relatively small – suggesting that different schools are good at improving different student outcomes. To examine whether high stakes tests measure impacts on other outcomes, we focus on the correlations between high-stakes exam impacts and other outcomes. These are generally low (see Figure 4). The correlation between school impacts on the high-stakes and low-stakes exam indexes is only 0.10, the correlation with dropout is 0.10, and the correlation with being formally employed is 0.08 (Table 3, column 1). 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 (0.24 and 0.26 respectively). This suggests that schools that improve high-stakes exams performance also tend to improve these outcomes, on average, but that only about 6.8 percent (i.e. 0.26 x 0.26 = 0.068) 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).

Can estimation errors explain these low correlations? The raw correlations previously discussed, are between two estimated school effects. Let the correlation between estimated school effect for outcome z and outcome z' be $\hat{\rho}_{z,z'}$. These raw correlations may be understated due to estimation errors in each school effect. For each outcome z, we can compute a reliability coefficient which is equal to the ratio between the variance of the true effects and that of the estimated effects, $R_z = var[\theta^{TOT}_{zj}]/var[\hat{\theta}^{TOT}_{zj}]$. If the estimation errors are random, then the correlation between the real effects ($\rho_{z,z'}$) is the raw correlation divided by the geometric mean of the reliability coefficients of the two estimates (i.e., $\rho_{z,z'} = \hat{\rho}_{z,z'}/\sqrt{R_z R_{z'}}$) (Spearman, 1904). We compute estimates of the reliability coefficient for each outcome, $\hat{R}_z$, as the ratio between (a) the maximum likelihood estimate of the variance of the persistent school effects and (b) the raw variance of the estimated school impacts. We then use these estimates to report disattenuated correlations (i.e., $\hat{\rho}_{z,z'}/\sqrt{\hat{R}_z\hat{R}_{z'}}$).

The disattenuated correlations are reported in the lower panel of Table 3. It is worth noting that these adjusted correlations rely on the assumption that the estimation errors are uncorrelated across outcomes. However, if the estimation errors are positively correlated across outcomes (as is very possible) these adjusted correlations may overstate the true relationship. We focus attention on the correlations involving the high stakes value-added. As one can see, even after accounting for measurement errors, the correlations with many of the other outcome impacts are modest. While the adjusted correlation between teen pregnancy and the high stakes exams is 1 (which can happen if the estimation errors are positively correlated across outcomes), the correlations with low stakes exams, dropout, and labor market participation are all below 0.3. In sum, while noise in our school estimates may explain some of the modest correlations between school impacts on high stakes exams and impacts other outcomes, it cannot explain most of it.

Are these relationships due to different students being marginal for different outcomes? If the same school has the same effect on all students, then the low correlations would be *prima-facie* evidence that schools that raise test scores are not those that improve other outcomes. However, these low correlations could also 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.[23]

[23]Imagine 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

Ideally, to isolate differences in school impacts from differences in the responsiveness of the students they admit to improvements in each outcome, one would want to estimate school impacts on similar students. To approximate this approach, we estimate the value added of each school on each outcome for a representative student. So long as the estimated effect is correct for the same representative student across schools, the underlying correlations should be largely unaffected by differences in the margins of ability that influence different outcomes. Specifically, where pct_i is the percentile of student i in the SEA distribution, we estimate each school's TOT effects while weighting each observation by $1/(1+(X-pct_i)^2/100)$. 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 will result in school effect estimates that are largely representative of students at the X^{th} percentile of the incoming ability distribution. As such, any reduction in the correlation across outcomes due to differential ability margins would be reduced for the weighted estimates.

We present the correlations between the value added estimates of high-stakes and other outcomes when we center the weights at the 25^{th} percentile, the median, and the 75^{th} percentile of the SEA distribution in Table 4.[24] These correlations suggest that there is little indication that this is a significant factor. 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 0.1 to 0.24, with formal employment increases from 0.08 to 0.17, and with no teen motherhood increases slightly from 0.24 to 0.28. Using weights centered on the median increase the correlations between high stakes test impacts and no dropout from 0.1 to 0.19. Taken together, the results do support the notion that these low correlations are due in part to different schools serving students who are marginal for different outcomes. However, even using this re-weighting, none of the correlations is greater than 0.3 – suggesting that much of the low correlation reflect different schools having impacts on different outcomes.

Overall, the patterns in Tables 3 and 4 indicate that school 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 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. (2019b) 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.

even though they would have the same effect had they admitted the same students.

[24] Appendix Table B2 presents weighted correlations between value added estimates of all other outcomes.

V Estimating Preferences for Schools

In this section we will examine the extent to which parents choose schools based on their causal impacts, and explore the extent to which they value school impacts on outcomes other than high-stakes tests. To this aim, we derive the choice probabilities using the school rankings submitted during the application process. Based on these choice probabilities, we use a modified exploded multinomial logistic model (also known as rank-ordered multinomial logit) to estimate preferences for different school attributes.

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:

$$U_{ij} = U(X_i, Z_{ij}, \varepsilon_{ij}) = \delta(X_i, Z_{ij}) + \varepsilon_{ij} \tag{7}$$

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 $U_{ij}^{r_{i\emptyset}}$ 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:

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

While one could rely only on comparisons within the set of submitted choices to infer preferences, such comparison can be misleading 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 the set of choices made (or at least one of the choices made) against all the possible choices that could have been made. Following Hastings et al. (2009) and Abdulkadiroğlu et al. (2019b) we 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 $U_{ij}^{r_{i1}} > U_{ik}\ \forall k \neq j \in J$ (Roth and Oliveira Sotomayor 1992).[25] 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.[26] 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

> $U_{ij}^{r_{i1}}|p_{ij} > U_{ik}^{r_{i\emptyset}}|p_{ik}\forall k \neq j \in J$: 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 (8) below.

$$\begin{gathered} Prob[r_{i1}, r_{i2}, \ldots, R_i] = Pr\big[(U_{ij}^{r_{i1}}|p_{ij} > U_{ik}^{r_{i\emptyset}}|p_{ik}\forall k \neq j \in J) \\ \cap(U_{ij}^{r_{i1}} > U_{ik}^{r_{im}}, 1 < m, \forall m \in \{2, \ldots, R_i\}) \cap \ldots \cap (U_{ij}^{r_{iR_i-1}} > U_{ik}^{r_{iR_i}})\big] \end{gathered} \tag{8}$$

V.2 Modified Exploded Multinomial Logistic Model

Equation (8) 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

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

[26] When there are finite choices, 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).

Hastings et al. (2009) and Abdulkadiroğlu et al. (2019b), we assume that the choices are rational and we parametrize δ_{ij} as a linear-in-parameters function of the school characteristics

$$U_{ij} = \beta' Z_{ij} + \varepsilon_{ij} \tag{9}$$

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 (8)) 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 $\{r_{i1}, r_{i2}, \ldots, R_i\}$ is:

$$Prob[r_{i1}, r_{i2}, \ldots, R_i] = \frac{\exp(\beta_1' Z_{ij}^{r_{i1}} + \pi p_{ij})}{exp(\beta_1' Z_{ij}^{r_{i1}} + \pi p_{ij}) + \sum_{k=1}^{J-R_i} \exp(\beta_1' Z_{ik}^{r_{i\emptyset}} + \pi p_{ik})}$$
$$\cdot \frac{\exp(\beta_1' Z_{ij}^{r_{i1}})}{\sum_{k=1}^{R_i} \exp(\beta_1' Z_{ik}^{r_{ik}})} \cdots \frac{\exp(\beta_1' Z_{ij}^{r_{iR-1}})}{\exp(\beta_1' Z_{ij}^{r_{iR-1}}) + \exp(\beta_1' Z_{ik}^{r_{iR_i}})}$$

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

$$\log L(\beta) = \sum_{i=1}^{N} \log l_i(\beta) = \sum_{i=1}^{N} \log \left(Prob[r_{i1}, r_{i2}, \ldots, R_i] \right). \tag{10}$$

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. The rest of the pseudo-observations in the model follow the exploded logit model template, where the individual chooses the k^{th} ranked school from the remaining alternatives in her ranking. 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.[27] These two modifications to the conventional multinomial logit model allows us

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

to anchor each individual's choice set while also explicitly accounting for strategic behaviors.[28]

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 unknown or 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 82^{nd} percentile scored above the cuttoff for that school and all students above the 92^{nd} 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 uncertainly for any student regarding their exact score, we do not use the actual likelihood of admission based on their score but rather 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 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 one 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 Preference Parameter Estimates

We examine whether parents value school impacts on academic and non-academic dimensions, above and beyond easily observed school attributes. Our full estimation sample includes 312,420 households making school choice decisions. We estimate choice models separately for each (SEA score ventile)×(gender) cell to allow preferences to vary based on the student's gender and the incoming ability. All specifications include control variables for whether the secondary school is on

[28] Our 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. (2019b) 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.

the same island, whether it is all-girls, whether it is all-boys, and the estimated "rough" likelihood of admission to that schools (but only when comparing the top choice to all unranked schools). 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 investigate the importance parents give to each school attribute by assessing 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 value-added model (which includes the value-added estimates for all outcomes, peer quality, and log distance), and (b) a full model that includes the value-added estimates for all outcomes, the school-level averages for all the outcomes, peer quality, and log distance.

Before discussing the school impacts, we first 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 point estimate on distance is negative and one can reject that the effect is zero at the 0.001 level. Second, parents of lower-achieving students are more responsive to distance than those of higher-achieving students. It is also worth noting that the patterns for the value-added 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 generally similar within each ability 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 value-added 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 preferences for peer quality, but higher-ability boys are more responsive to peer quality than low-achieving boys (while there is no 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 school choice studies (e.g., Hastings et al. 2005; Abdulkadiroğlu et al. 2019b). A comparison of the coefficients on peer quality and proximity implies that, on average, increasing peer quality by about 0.67 standard deviations (roughly the difference between a student's top choice school and the third choice school) is valued about the same as doubling the distance between the primary and secondary school. That is, parents are willing to travel about 1.5 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* 9 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 in this effect. Parents of males and females in the bottom decile of the test score distribution would be willing to increase travel distance by about 25 percent to attend a school with 1 standard deviation higher peer quality. While this is similar for high-achieving females, males in the top decile would be willing to increase their distance more than threefold (from a base of 6 kilometers) to attend a school with one standard deviation higher-achieving peers.

As discussed in Section V.3, a key conditioning variable in our analysis is the admission probability. 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 value-added only model the point estimates for all groups are positive – that is, conditional on peer quality and proximity and value-added on key outcomes, parents are more likely to chose 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

Having established that parents in Trinidad and Tobago make similar choices to parents in other settings, and that parents 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 value-added 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. The figure reveals three key patterns: (1) School impact on high stakes exams influences the choices of many parents, but there is considerable heterogeneity in parental preference for school impacts on high stakes exams, (2) Parents of high-achieving students have stronger preferences for schools with larger causal impacts on the high-stakes exam performance than parents of lower-achieving students, and (3) at all levels of incoming achievement, parents of girls value school impacts on high-stakes exams more than 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-value<0.001), the figure also reveals that parents of children with low incoming scores *do 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 value-added – in fact the point estimates are slightly negative.

To put these point estimates in perspective, we compare the coefficients on high-stakes value-added to that of the natural log of distance. For those in the top decile of incoming test scores, parents of girls (conditional on average peer scores) would be willing to travel more than three times farther, or about 18 kilometers farther, to attend a secondary school at the 85^{th} percentile of the high-stakes test scores effectiveness distribution than one at the median. This is more than the width, and just under half the length, of mainland Trinidad. 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 85^{th} 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 schools value-added on high-stakes exams.

As we control for school averages (i.e average high-stakes exam scores and that for other outcomes), the importance of high-stakes exam value-added 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 value-added remains positive and highly statistically significantly different from zero for parents of high-achieving children– suggesting that these parents, on average, can distinguish value-added from selection. This result stands in contrast to Abdulkadiroğlu et al. (2019b) who find that after conditioning on peer quality, parents do not appear to value school effectiveness. We discuss possible reasons for these differences in Section V.5. Looking to 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 (2018) who argue that parents may value schools with better average outcomes. However, our results do suggest that (at least in our setting) parents preferences for 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 prefer schools that raise high-stakes exam performance, there is little evidence that parents value schools that improve low-stakes exam performance.[29] Specifically, the results in (Figure 8, left and middle) show insignificant effects across the entire incoming ability distribution for both boys and girls. Moreover, these point estimates are much smaller in magnitude than those for high-stakes exam value-added – further evidence that parents school choices are sensitive to schools' high-stakes value added but not to their low-stakes value added. Looking at average outcomes, parents are largely unresponsive 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 exams are not. As such, the results are consistent with parents being able to discern school impacts on high-stakes exams but not on low-stakes exams. They are, of course, 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 in (Figure 9, left and middle panels) suggest that parents are no more likely to chose schools that causally reduce dropout. For none of the gender-by-ability cells can one reject that the coefficient for dropout value-added is zero. However, looking at dropout *rates*, the patterns suggests that low-achieving male and female students prefer schools with low average dropout rates. See the positive and significant estimated coefficients at the bottom of the ability distribution in the right panel. For females, the point estimates are positive for most of the ability distribution and one can reject that the effects are zero for females in the bottom third of the incoming ability 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 pregnancy 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 schooling decisions. Recall that all variables are coded so that positive values in-

[29] 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.

dicate better outcomes. We start with teen motherhood. The patterns in Figure 10 reveal little preference of parents for schools that reduce teen motherhood. In the value-added model the 95 percent confidence interval for all of the estimates include zero – indicating little systematic relationship between how parents rank schools and their impacts on teen pregnancy. In the full model that also includes the teen pregnancy 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. We take this as weakly suggestive evidence that females may avoid schools that have high pregnancy value-added.

While the evidence on parental preferences for teen pregnancy value added is mixed, parents do have strong preferences for schools with low average teen motherhood rates. Between two schools with the same average academic performance and value-added 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 value school safety (or low risk schools) is not new, but this is the first study to document this rigorously in a discrete-choice framework. Looking at the point estimates, the average parent would be willing to increase their distance by about 42 percent to send their child to a school that was at the 85^{th} percentile of the average (non) teen motherhood distribution versus one at the median. These preferences are even stronger for parents of high-achieving children. Parents of females in the top decile of the achievement distribution would be willing to increase their distance by about 60 percent to attend a school at the 85^{th} percentile of the (non) teen motherhood distribution versus one at the median. Parents of males in the top decile of the achievement distribution would be willing to more that double their distance to attend a school at the 85^{th} percentile of the (non) teen motherhood distribution versus one at the median. In sum, while there is little evidence that parents value school impacts on teen motherhood (perhaps because they cannot observe or infer it), they do value schools that have low pregnancy rates. This likely reflects parental preferences for safety.

Another important non-academic measure is teen arrests. Unlike teen motherhood, the results in Figure 11 reveal that parents are responsive to school impacts on reduced teen arrests. This is driven largely by parents of males and parents of high-achieving students (left panel). Males of almost all ability levels prefer schools that reduce crime, and this preference is stronger for the higher-ability males. Indeed, one can reject that this effect is zero for males above the 40^{th} percentile of the incoming achievement distribution. In contrast, while the parents of higher-ability females appear to prefer schools that reduce crime, one cannot reject null impacts for girls in most ability groups.

In sum, while there is clear evidence that boys value schools that reduce crime, this is not so for girls. 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 have the strongest preference for crime value added. Specifically, parents of students (both boys and girls) in the top decile would be willing to increase their distance by about 170 percent (roughly 13 kilometers) in order to send their child to a school that was at the 85^{th} percentile of the (non) teen arrest value-added distribution versus one at the median. In fact, a comparison of the estimates for high stakes value added reveal that for the parents of males below the 85^{th} percentile, impacts on arrests matter more 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 preferences for low-risk schools, parents of all ability levels prefer 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, show that parents clearly prefer schools with a lower prevalence of risky behaviors (conditional on peer achievement and average test score outcomes). The results also indicate that parents do 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 labor market participation at ages 27 and older. All parents value school impacts on employment (Figure 12, left). On average, parents of students would be willing to increase their distance by about 40 percent to send their child to a school that was at the 85^{th} percentile of the labor market employment value-added distribution versus one at the median. These patters are noteworthy for two reasons. First, this is the first direct demonstration that parents may value schools that have causal impacts on labor market participation (above and beyond peer quality and impacts on academic outcomes). Second, even though parents of children in the bottom half of the incoming ability distribution do not value school impacts on test scores, they do value school impacts on formal labor market participation – suggesting that for more than half of the population school impacts on labor market participation matter more than impacts on high stakes exams. The effects of labor market value added are largely similar in models that include school average labor market participation rates (Figure 12, middle).

As with the other dimensions, (Figure 12, right) reveals that most parents also value school average labor market employment rates. On average, parents of students would be willing to increase their distance by about 50 percent to send their child to a school that was at the 85^{th} percentile of the average labor market employment distribution versus one at the median. Interestingly, the preference for schools with higher average employment rates is larger for parents of children at the lower end of the incoming ability distribution.

In sum, these results provide evidence that parents 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 value 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.[30]

Even though we do not find that parents value school impacts on all outcomes, we do find evidence that parents prefer schools with lower dropout rates, lower teen arrest rates, lower teen pregnancy rates and higher future labor market participation. It is important to note that valuing schools' average outcomes (but not value-added) is not necessarily irrational, because schools with higher average outcomes may confer benefits to students that are not measured in our data.[31] Overall, parents appear to be relatively sophisticated in their understanding of school quality.

V.5 Discussion of Parental Preference Results

One of our key findings is that parents value school effectiveness on multiple outcomes above and beyond peer quality and average outcomes. The only other paper to formally test this notion is Abdulkadiroğlu et al. (2019b) who find that parents do prefer school 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 parental preferences for schools reflect effects beyond academics – this is very important given that many school choice evaluations use test scores alone.

Another potential explanation for differences between our work and Abdulkadiroğlu et al. (2019b) 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., Besedeŝ et al. (2012)) and less likely to make the optimal choice (e.g., Schram and

[30] 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 value-added estimates 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 value added and leave-year-out school average outcomes are qualitatively similar, and our central conclusions unchanged.

[31] Indeed, MacLeod and Urquiola (2018) present a model that rationalizes preferences for schools' absolute achievement when this attribute serves as a signal that improves labor market matching.

Sonnemans (2011)). Abdulkadiroğlu et al. (2019b) 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 (2018)). Indeed, within the studied context, school value-added may be easier to infer for relatively sophisticated parents. In Trinidad and Tobago, 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 value-added (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 (ie. those of more able students) can discern school effectiveness (conditional on average outcomes) while some other studies do not find so.

As previously suggested, because the average incoming scores of students is known in Trinidad and Tobago, sophisticated parents may be able to infer value-added by comparing average outcomes for schools with similar incoming student populations. 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 value-added 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 value schools that have larger positive causal impacts on high-stakes tests. However, they also value schools that decrease crime and increase labor market participation. Importantly, parents value schools that 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 – which is a key requirement for the potential benefits of school choice (Friedman, 1955). The fact that parents do not *only* prefer 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 that fact that parental preferences for schools are not strongly related to school's test score impacts (MacLeod and Urquiola, 2018). 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 prefer high value-added schools, we do find heterogeneity in this effect. Parents value effectiveness 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 value-added 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).[32] 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.

[32]A recent experimental study in Chile (Allende et al. 2019) find that a personalized information intervention led parents to chose higher value added schools 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)
Panel A: SEA data (cohorts: 1995 - 2012)					
Female (%)	50.38			53.52	47.24
	(50.00)			(49.88)	(49.92)
Admitted cohort size	217.72	220.55	215.06	220.55	215.06
	(165.11)	(163.97)	(166.12)	(163.97)	(166.12)
Standardized SEA score	0.00	-0.13	0.12	0.70	-0.70
	(1.00)	(1.04)	(0.94)	(0.62)	(0.79)
Individuals	312,420	155,017	157,403	156,168	156,252
Panel B: NCSE data (linked to SEA cohorts: 2006 - 2012)					
Took NCSE (%)	90.48	88.72	92.26	95.27	83.24
	(29.34)	(31.64)	(26.72)	(21.22)	(37.35)
Standardized NCSE score	0.00	-0.23	0.22	0.31	-0.55
	(1.00)	(1.00)	(0.95)	(0.89)	(0.95)
Individuals	108,097	54,209	53,888	65,084	43,013
Panel C: CSEC data (linked to SEA cohorts: 1995 - 2011)					
Took at least 1 subject (%)	77.58	72.24	82.85	89.27	66.13
	(41.70)	(44.78)	(37.70)	(30.95)	(47.33)
Number of subjects passed	3.28	2.66	3.89	4.93	1.66
	(3.11)	(2.98)	(3.11)	(2.94)	(2.31)
Qualified for tertiary (%) *	36.00	28.90	42.97	59.75	12.70
	(48.00)	(45.33)	(49.50)	(49.04)	(33.30)
Individuals	296,838	147,212	149,626	146,950	149,888
Panel D: CAPE data (linked to SEA cohorts: 1999 - 2009)					
Took at least 1 unit (%)	20.58	16.01	25.10	34.84	4.65
	(40.43)	(36.67)	(43.36)	(47.65)	(21.05)
Number of units passed	1.45	1.11	1.79	2.50	0.28
	(2.98)	(2.66)	(3.22)	(3.58)	(1.38)
Earned Associate Degree (%)	15.35	11.38	19.28	26.85	2.50
	(36.05)	(31.75)	(39.45)	(44.32)	(15.60)
Earned scholarship (%)	0.99	0.69	1.29	1.88	0.01
	(9.92)	(8.30)	(11.29)	(13.57)	(0.86)
Individuals	198,762	98,810	99,952	104,897	93,865
Panel E: Criminal records (linked to SEA cohorts: 1995 - 2010) - in percent					
Arrested by 18	3.28	5.81	0.80	1.73	4.77
	(17.82)	(23.39)	(8.89)	(13.04)	(21.31)
Individuals	281,578	139,581	141,997	137,693	143,885
Panel F: Birth records (linked to SEA cohorts: 2004 - 2010) - in percent					
Live birth by 19			9.88	6.49	14.97
			(29.84)	(24.64)	(35.68)
Individuals			42,167	25,337	16,830
Panel G: Labor market data (linked to SEA cohorts: 1995 - 2002) - in percent					
Formally employed	75.67	79.58	71.87	78.35	73.79
	(42.91)	(40.31)	(44.96)	(41.19)	(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]	
Standardized outcomes			
Low-Stakes Index	0.412	0.359	0.474
High-Stakes Index	0.393	0.347	0.446
No Dropout by 14	0.114	0.097	0.135
No live birth by 19	0.115	0.067	0.197
Not arrested by 18	0.089	0.076	0.103
Formally employed	0.082	0.067	0.100
Binary outcomes			
No Dropout by 14	0.036	0.030	0.042
No live birth by 19	0.035	0.020	0.060
Not arrested by 18	0.016	0.014	0.019
Formally employed	0.036	0.029	0.044

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).

Table 3: Raw and Disattenuated Correlations Between School Impacts

	High-Stakes Index	Low-Stakes Index	No Dropout by 14	No live birth by 19	Not arrested by 18	Formally employed 27+
Raw correlations						
High-Stakes Index	1.00					
Low-Stakes Index	0.10	1.00				
No Dropout by 14	0.10	0.05	1.00			
No live birth by 19	0.24	-0.07	-0.19	1.00		
Not arrested by 18	0.26	-0.03	0.23	-0.39	1.00	
Formally employed 27+	0.08	-0.04	-0.22	0.19	-0.03	1.00
Disattenuated correlations						
High-Stakes Index	1.00					
Low-Stakes Index	0.15	1.00				
No Dropout by 14	0.18	0.09	1.00			
No live birth by 19	1.08	-0.35	-1.07	1.00		
Not arrested by 18	0.51	-0.06	0.58	-2.44	1.00	
Formally employed 27+	0.22	-0.11	-0.75	1.64	-0.12	1.00

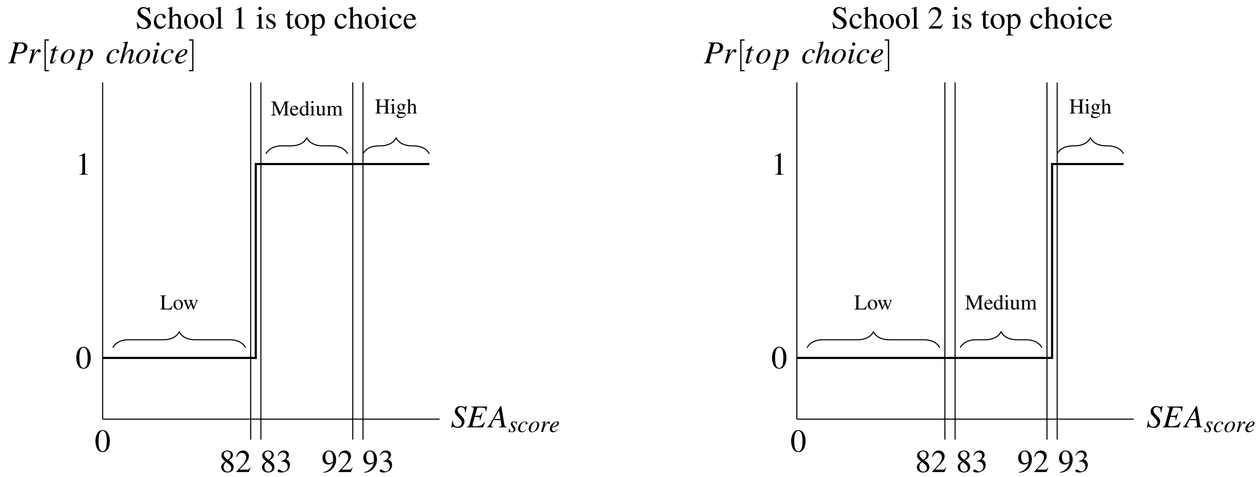
Notes: This table reports raw and disattenuated correlation coefficients of the estimated school impacts, $\hat{\theta}_j^{TOT_{IV}}$, across the different dimensions measured. We compute estimates of the reliability coefficient for each outcome, $\hat{R}_z$, as the ratio between (a) the maximum likelihood estimate of the variance of the persistent school effects and (b) the raw variance of the estimated school impacts. We then use these estimates to report disattenuated correlations (i.e., $\hat{\rho}_{z,z'}/\sqrt{\hat{R}_z\hat{R}_{z'}}$) in the lower panel.

Table 4: Weighted Correlations Between High-Stakes and Other School Impacts Impacts

	Regression Weights Centered on			
	Average	25pct	50pct	75pct
	High-Stakes Index			
High-Stakes Index	1.00	1.00	1.00	1.00
No Dropout by 14	0.10	-0.11	0.19	0.02
Low-Stakes Index	0.10	0.24	0.16	0.11
Formally employed 27+	0.08	0.17	-0.04	0.11
No live birth by 19	0.24	0.28	0.22	0.02
Not arrested by 18	0.26	0.25	0.08	0.19

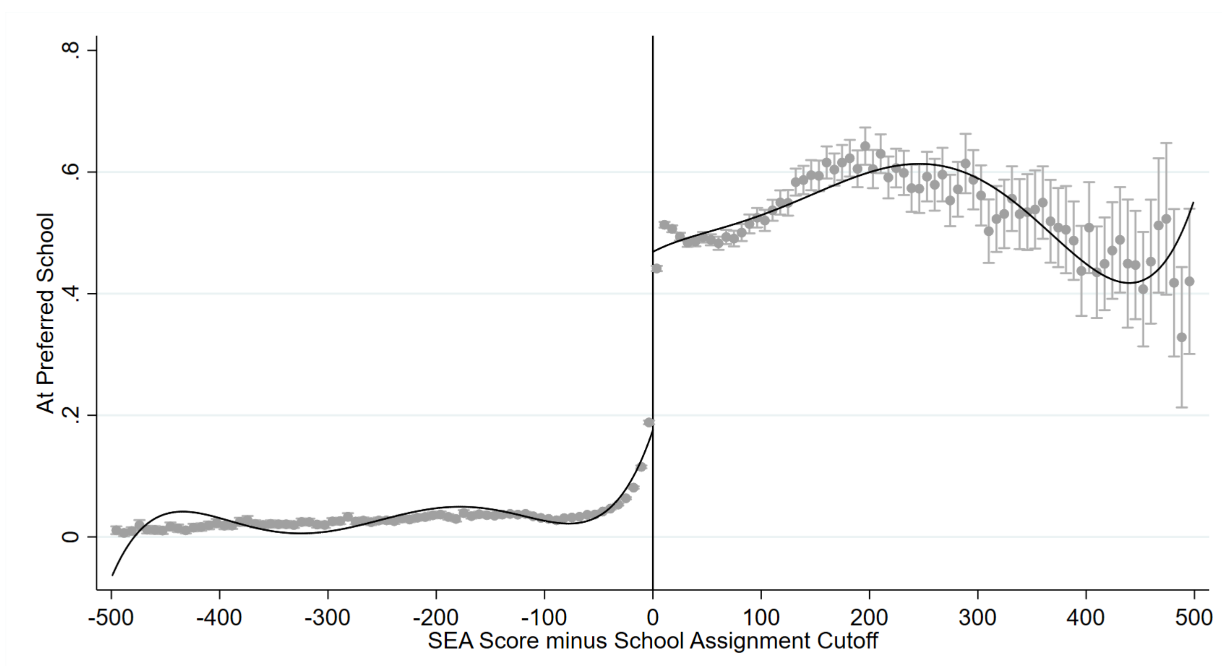
Notes: This table reports correlation coefficients of the estimated school impacts, $\hat{\theta}_j^{TOT_{IV}}$, between high-stakes and the different dimensions measured. Results are reported for the school impacts estimated with each student having equal weights (average) as well as student's weights being centered around different SEA score percentiles to test for effect heterogeneity across students. The weights used in each case were $w_i = \frac{1}{1+\frac{(score_{pct}-pct_i)^2}{100}}$ were $score_{pct}$ is the students score percentile and pct_i the percentile around which the weights are centered (25, 50 or 75)

Figure 1: Exemplar of Variation



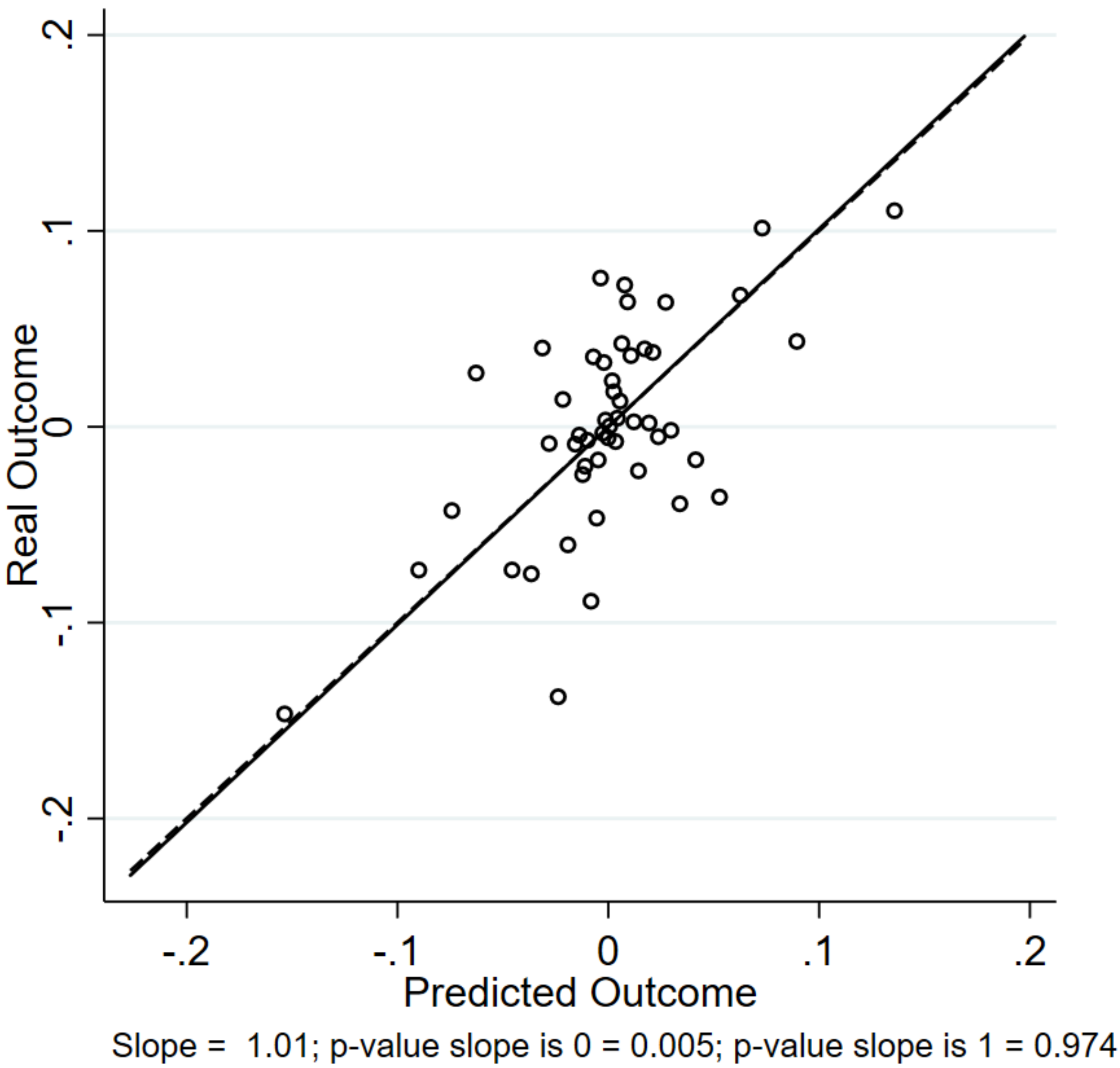
Notes: The Y-axis represents the probability of student *i* being assigned to her top choice. The X-axis represents the student's SEA score. This figure shows the two different sources of variation, that is, the RD variation given by the scores when comparing the Low scoring group to the Medium scoring group in the left panel (or the Medium to the High scoring group in the right panel) and the variation given by different choices, that is, comparing the Medium scoring group between both panels.

Figure 2: Discontinuity in Preferred School Attendance Through Assignment Cutoffs



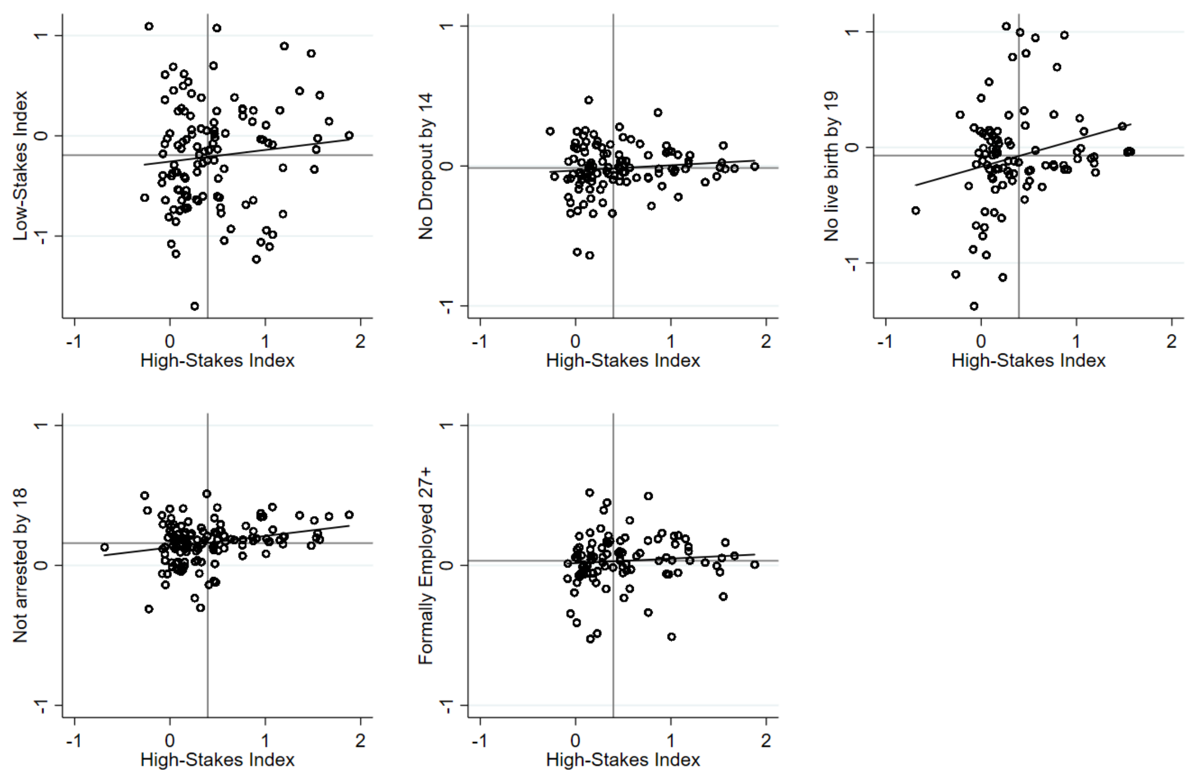
Notes: The Y-axis represents the likelihood of preferred school attendance (i.e. the school where the student was enrolled at the time of taking the CSEC examinations). The X-axis is the SEA score relative to the deferred acceptance rule-based assignment cutoff. The circles are means corresponding to 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 3: Predicted Cutoff Effects versus Actual Cutoff Effects



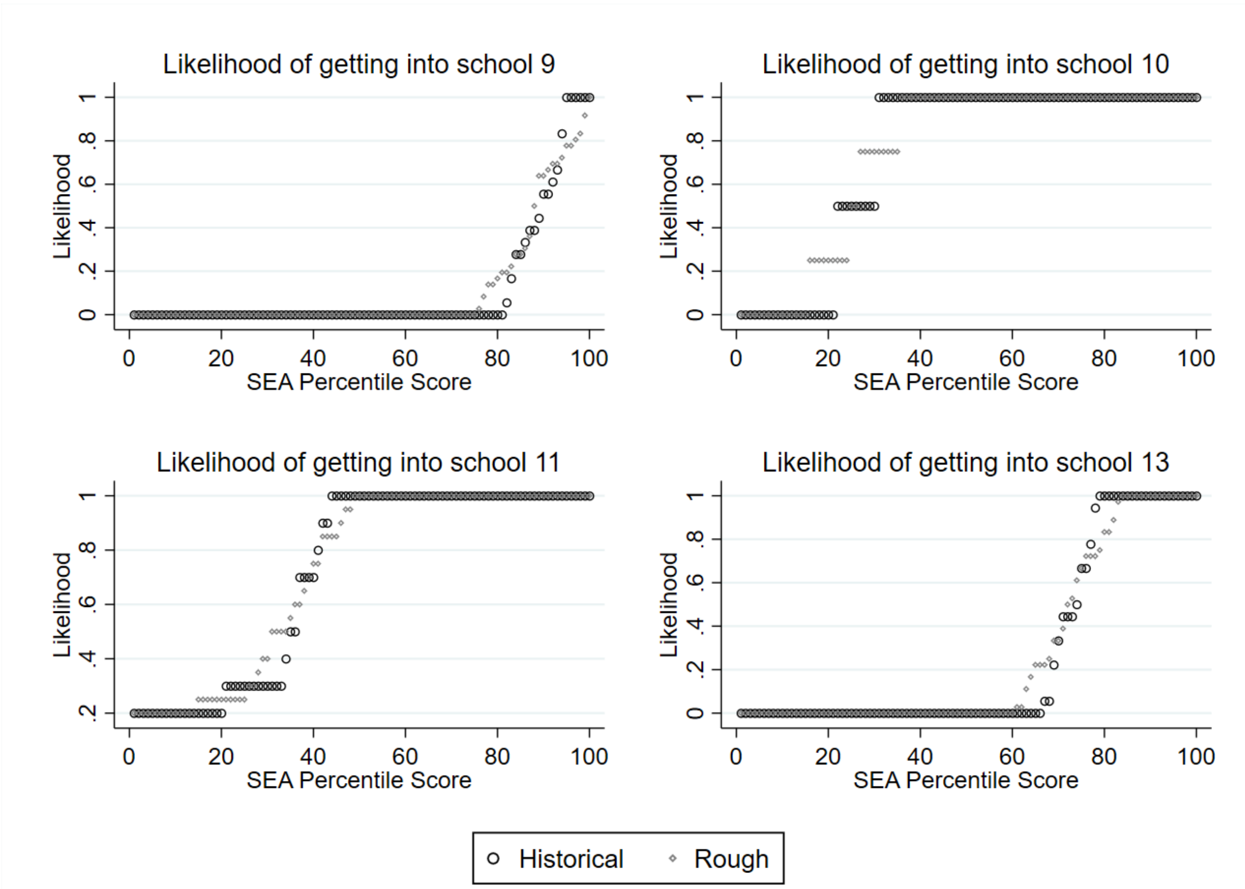
Notes: The X-axis represents the estimated coefficients on an indicator for scoring above the rule-based cutoff resulting from an RD model that controls for a fifth degree polynomial of the SEA score, gender, district of residence at SEA registration, and religion; estimated for each school *j* and for each outcome where the estimated TOT school impacts ($\hat{\theta}_j^{TOT_{IV}}$) 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 connected lines represent lineal fits; 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. All schools estimated effects are weighted by the inverse of the squared standard error of each estimated coefficient. Outliers above 4 standard deviations away from the mean were removed.

Figure 4: Relationships Between High-Stakes and Other Impacts



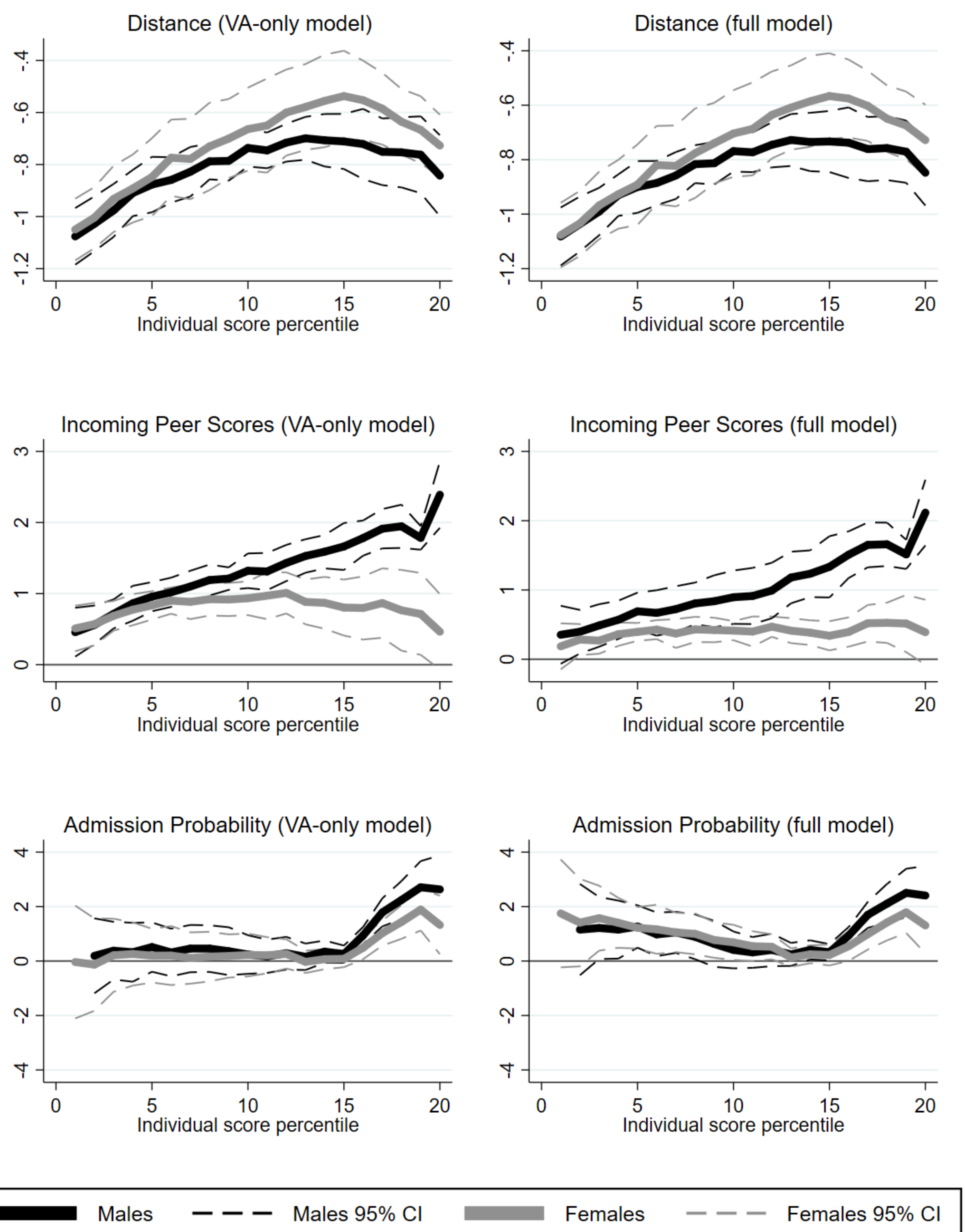
Notes: The X-axis represents the school TOT impacts ($\hat{\theta}_j^{TOT_{IV}}$) on the high-stakes index. The Y-axis represents the school TOT impacts on the other outcomes. A linear fit is added to each plot. It can be seen that the estimated school impacts $\hat{\theta}_j^{TOT_{IV}}$ for each school on the high-stakes index compared to all other outcomes have low correlations.

Figure 5: Estimated Admission Probabilities for Selected Schools



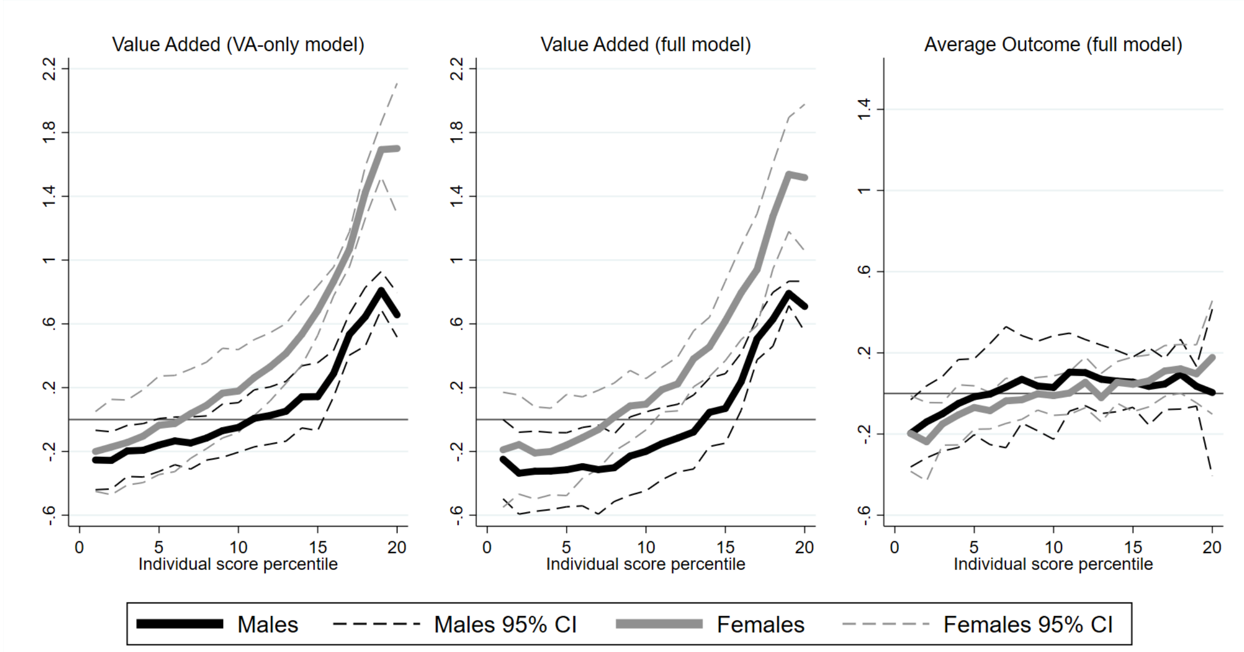
Notes: The X-axis represents the SEA score percentile. The Y-axis represents the likelihood of being assigned to a given school. School 9 is a very selective secondary school, where the school algorithm-based assignment cutoff has always been above the 82^{nd} percentile. School 10, on the other hand, is a less selective school where students above the 30^{th} 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



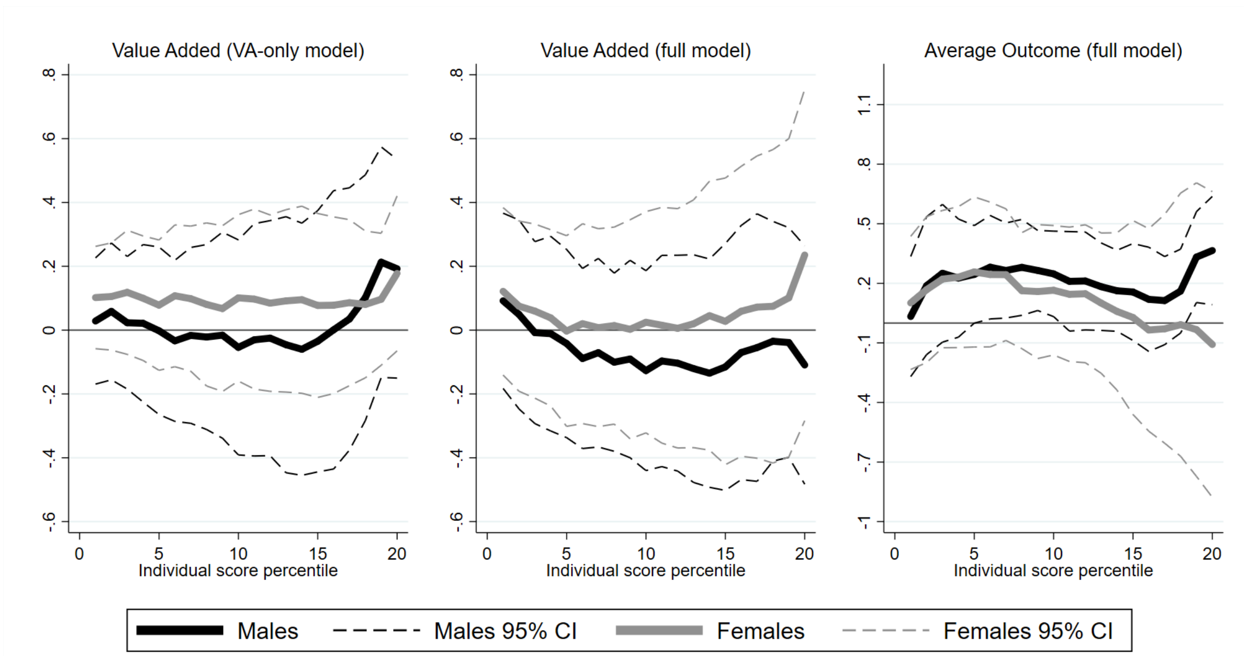
Notes: The connected lines represent the estimated coefficients for the outcome, computed separately for each (SEA score ventile)×(gender) cell for two different models: VA-only model (left panel), and full model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure 7: High-Stakes Index



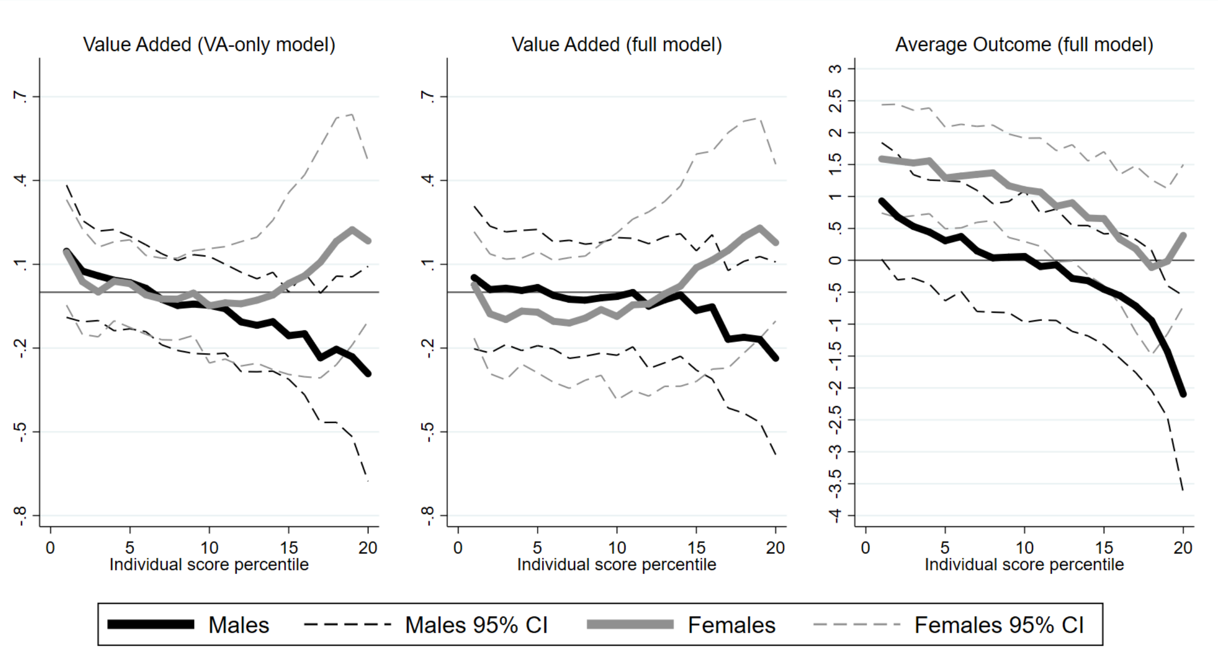
Notes: The connected lines represent the estimated coefficients for the outcome, computed separately for each (SEA score ventile)×(gender) cell for two different models: VA-only model (left panel), and full model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure 8: Low-Stakes Index



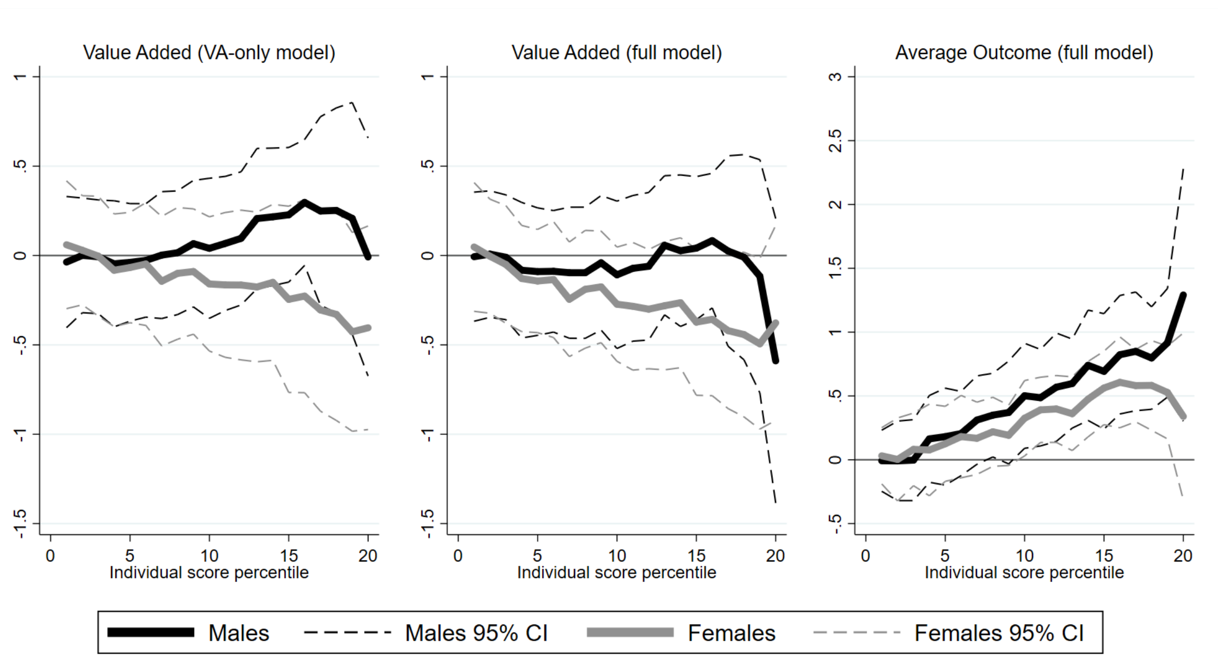
Notes: The connected lines represent the estimated coefficients for the outcome, computed separately for each (SEA score ventile)×(gender) cell for two different models: VA-only model (left panel), and full model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure 9: No Dropout by Age 14



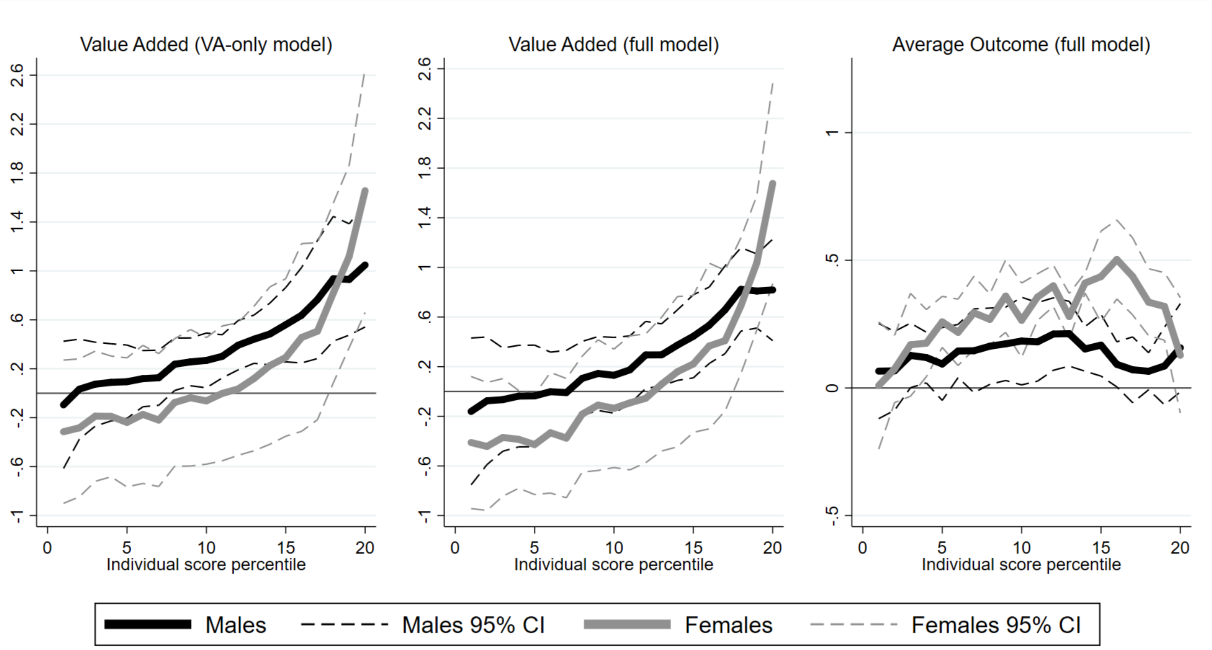
Notes: The connected lines represent the estimated coefficients for the outcome, computed separately for each (SEA score ventile)×(gender) cell for two different models: VA-only model (left panel), and full model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure 10: No Live Birth by Age 19



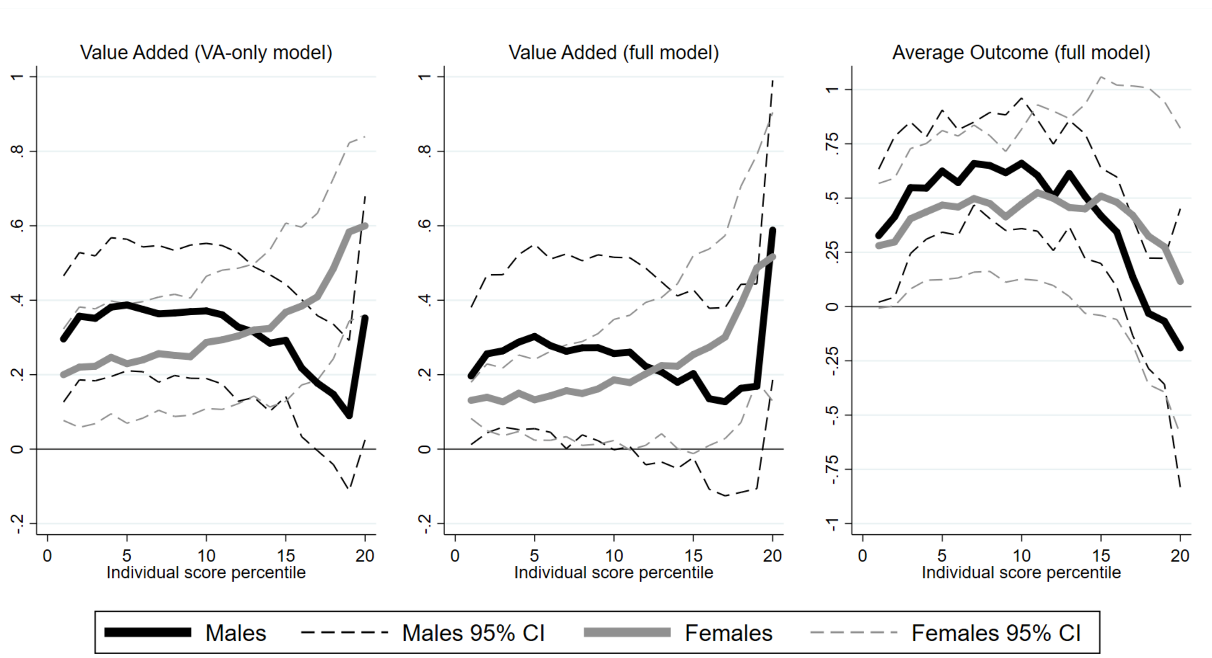
Notes: The connected lines represent the estimated coefficients for the outcome, computed separately for each (SEA score ventile)×(gender) cell for two different models: VA-only model (left panel), and full model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure 11: Not Arrested by Age 18



Notes: The connected lines represent the estimated coefficients for the outcome, computed separately for each (SEA score ventile)×(gender) cell for two different models: VA-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 for the outcome, computed separately for each (SEA score ventile)×(gender) cell for two different models: VA-only model (left panel), and full model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

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., 2019a.

Atila Abdulkadiroğlu, Parag Pathak, Jonathan Schellenberg, and Christopher Walters. Do Parents Value School Effectiveness? NBER Working Paper 23912, Cambridge, MA, 2019b.

Claudia Allende, Francisco Gallego, and Christopher Neilson. Approximating the Equilibrium Effects of Informed School Choice. *Princeton University Mimeo*, 2019.

Joshua Angrist and Miikka Rokkanen. Wanna Get Away? RD Identification Away from the Cutoff. *Journal of the American Statistical Association*, 2015.

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.

Tibor Besedeŝ, 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 Beuermann and C. Kirabo Jackson. The Short and Long-Run Effects of Attending The Schools that Parents Prefer. NBER Working Paper 24920, Cambridge, MA, 2018.

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.

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.

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 Stabilty 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.

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, 2018.

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.

C. Spearman. The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, 15(1):72, 1904.

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 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 an 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 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 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.[33] Scoring above zero means scoring above the cutoff for a preferred school. Figure **??** in the main text 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.[34] 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 exogenous 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 simuled 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.[35] 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

[33] Specifically, 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.

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

[35] We 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 Appendix Table A2, column (1) 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:

$$Y^{p}_{i\tau t} = \pi \cdot Above_{ijt} + f(SEA_{it}) + C_{jt} + \varepsilon_{ijt} \tag{11}$$

where $Y^{p}_{i\tau t}$ is the predicted outcome for individual i who attended school τ at time t. $Above_{ijt}$ is an indicator for scoring above the cutoff for preferred school j at time t. $f(SEA_{it})$ is a 5th order polynomial of the incoming SEA score net of the cutoff score for preferred school j fully interacted with the $Above_{ijt}$ indicator. C_{jt} is an cutoff fixed effect for applicants to school j 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, Appendix Table A2, column (2) 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 (11) for different bandwidths around the cutoff. Figure A1 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	*p-value*	SEA Cohort	*p-value*
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

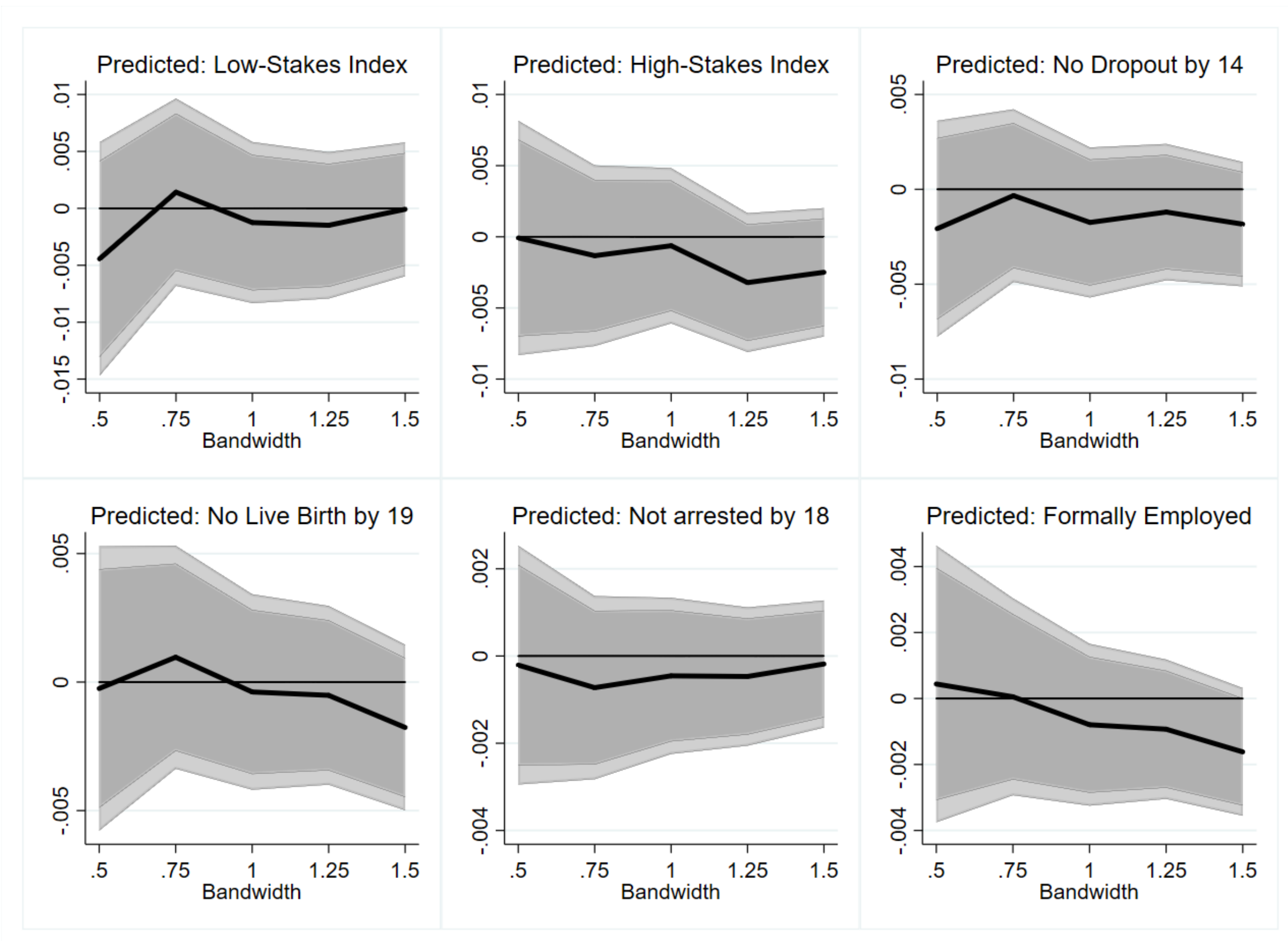
Notes: This table reports *p-values* of differential density tests across school assignment cutoffs for each SEA cohort included in the study.

Table A2. First Stage and Reduced-Form Effects

	Predicted Outcomes			Actual Outcomes			
	Prediction R2	Effect	*p-value*	Effect	*p-value*	Effect	*p-value*
	(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 (11). 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 π=0 shown next to the estimated coefficients.

Figure A1. Reduced-form effects on predicted outcomes by bandwidth



Notes: This figure reports estimated coefficients on 'Above' resulting from equation (11). The estimated coefficients are reported for each bandwidth between +/-0.5sd an +/-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: Weighted Correlations Between School Impacts Impacts

	Regression Weights Centered on							
	Average	25pct	50pct	75pct	Average	25pct	50pct	75pct
	High-Stakes Index				Low-Stakes			
High-Stakes Index	1.00	1.00	1.00	1.00	0.10	0.24	0.16	0.11
Low-Stakes Index	0.10	0.24	0.16	0.11	1.00	1.00	1.00	1.00
No Dropout by 14	0.10	-0.11	0.19	0.02	0.05	0.18	0.18	-0.21
No live birth by 19	0.24	0.28	0.22	0.02	-0.07	0.09	-0.20	-0.06
Not arrested by 18	0.26	0.25	0.08	0.19	-0.03	0.18	-0.14	-0.06
Formally employed 27+	0.08	0.17	-0.04	0.11	-0.04	0.00	-0.03	0.07

	Regression Weights Centered on							
	Average	25pct	50pct	75pct	Average	25pct	50pct	75pct
	No Dropout by 14				No live birth by 19			
High-Stakes Index	0.10	-0.11	0.19	0.02	0.24	0.28	0.22	0.02
Low-Stakes Index	0.05	0.18	0.18	-0.21	-0.07	0.09	-0.20	-0.06
No Dropout by 14	1.00	1.00	1.00	1.00	-0.19	-0.07	0.17	0.12
No live birth by 19	-0.19	-0.07	0.17	0.12	1.00	1.00	1.00	1.00
Not arrested by 18	0.23	0.06	0.26	-0.04	-0.39	-0.18	-0.29	-0.17
Formally employed 27+	-0.22	-0.15	-0.38	-0.14	0.19	0.03	0.15	-0.06

	Regression Weights Centered on							
	Average	25pct	50pct	75pct	Average	25pct	50pct	75pct
	Not arrested by 18				Formally employed 27+			
High-Stakes Index	0.26	0.25	0.08	0.19	0.08	0.17	-0.04	0.11
Low-Stakes Index	-0.03	0.18	-0.14	-0.06	-0.04	0.00	-0.03	0.07
No Dropout by 14	0.23	0.06	0.26	-0.04	-0.22	-0.15	-0.38	-0.14
No live birth by 19	-0.39	-0.18	-0.29	-0.17	0.19	0.03	0.15	-0.06
Not arrested by 18	1.00	1.00	1.00	1.00	-0.03	0.05	-0.12	0.01
Formally employed 27+	-0.03	0.05	-0.12	0.01	1.00	1.00	1.00	1.00

Notes: This table reports correlation coefficients of the estimated school impacts, $\hat{\theta}_j^{TOT}$, across the different dimensions measured. Results are reported for the school impacts estimated with each student having equal weights (average) as well as student's weights being centered around different SEA score percentiles to test for effect heterogeneity across students. The weights used in each case were $w_i = \frac{1}{1+\frac{(score_{pct}-pct_i)^2}{100}}$ were $score_{pct}$ is the students score percentile and pct_i the percentile around which the weights are centered (25, 50 or 75)

Appendix C: Validation and Magnitude of School Impacts

Validation of school impacts using exogenous cutoff variation

Existing papers that have explored parental preferences for school effectiveness 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.[36] 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 j, the causal effect of being tentatively assigned to school j is simply the effect of scoring above the admission cutoff for school j (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.[37]

To obtain the reduced-form Regression Discontinuity (RD) effect of being tentatively assigned to any school j for each outcome, we estimate RD models for each outcome among all applicants to school j.[38] Under the RD identifying assumptions, the reduced-form effect of being tentatively assigned to school j on outcome Y, is captured by estimating the equation below.

$$Y_{i\tau} = Above_{ij} \cdot \gamma_j + f(SEA_i) + \mathbf{X_i}'\delta + \varepsilon_{i\tau} \quad (12)$$

Where $Y_{i\tau}$ is the outcome of student i who attended school τ, and $Above_{ij}$ is an indicator for scoring above the algorithm-based assignment cutoff for school j. Among those who comply with the cutoff, τ=j. The parameter γ_j captures the difference in outcomes (all else equal) between those exogenously assigned to a preferred school j (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 j). As such, in the neighborhood of the cutoff,

$$E[\hat{\gamma}_j|X_i, SEA_i] = E(Y_{i\tau}|Above = 1) - E(Y_{iq}|Above = 0) \quad (13)$$

[36]In related work Abdulkadiroğlu et al. (2019b) 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.

[37]For 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.

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

To simplify equation (13), we consider this expression for compliers and for non-compliers. Under the assumption of unconfoundedness (Rubin 1990), it follows that $E[Y_{i\tau} - Y_{iq}|X_i, SEA_i] = \theta_\tau^{TOT} - \theta_q^{TOT}$. That is, if there is no selection on observables, the average difference in outcomes between observationally equivalent individualsablin school and school q reflects the difference in effectiveness between school τ and school q. Among the compliers, school τ is school j if they score above the cutoff.[39] As such, for compliers, $E[\hat{\gamma}_j|X_i, SEA_i] = \theta_j^{TOT} - E[\theta_q^{TOT}]$, where $E[\theta_q^{TOT}]$ is the average effectiveness of the counterfactual schools for the applicants to school j. Among non-compliers, the cutoff does not change the school attended so that $E[\hat{\gamma}_j|X_i, SEA_i] = 0$. It follows that for the average applicant to school j, equation (13) can be written as equation (14) below.

$$E[\hat{\gamma}_j|X_i, SEA_i] = \bar{p}_j \times (\theta_j^{TOT} - E[\theta_q^{TOT}]) \tag{14}$$

In words, in expectation, the estimated effect of scoring above the cutoff for school j is the difference between the impact of attending preferred school j and that of attending the average counterfactual school q, all times the compliance rate ($\bar{p}_j$). 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, $\hat{\theta}_\tau^{TOT_{IV}}$, as below.

$$\hat{\theta}_\tau^{TOT_{IV}} = Above_{ij} \cdot \zeta_j + f(SEA_i) + \mathbf{X_i'}\delta + \varepsilon_{ij} \tag{15}$$

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

$$E[\hat{\zeta}_j|X_i, SEA_i] = \bar{p}_j \times (E[\hat{\theta}_j^{TOT_{IV}}] - E[\hat{\theta}_q^{TOT_{IV}}]) \tag{16}$$

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

Inspection of (14) and (16) reveals that if our treatment on the treated estimated impacts for attended school τ and the average counterfactual school q are unbiased, then by the law of iterated expectations, $E[\hat{\zeta}_j|X_i, SEA_i] = E[\hat{\gamma}_j|X_i, SEA_i]$. In words, if our estimated TOT school impacts ($\hat{\theta}_j^{TOT_{IV}}$) are consistent estimates of the causal effect of attending school j, then for each applicant school, the RD estimates on the actual outcomes and the RD estimates on the TOT school impacts of the attended school, should be equal in expectation.

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 $\hat{\gamma}_j$ and $\hat{\zeta}_j$ for each preferred school j (using the optimal bandwidth from Imbens and Kalyanaraman (2012)). We then regress the former coeffi-

[39]In 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).

cients on the latter while weighting each coefficients by the inverse of its squared standard error. Finally, we test for whether the estimated slope is statistically indistinguishable from 1.

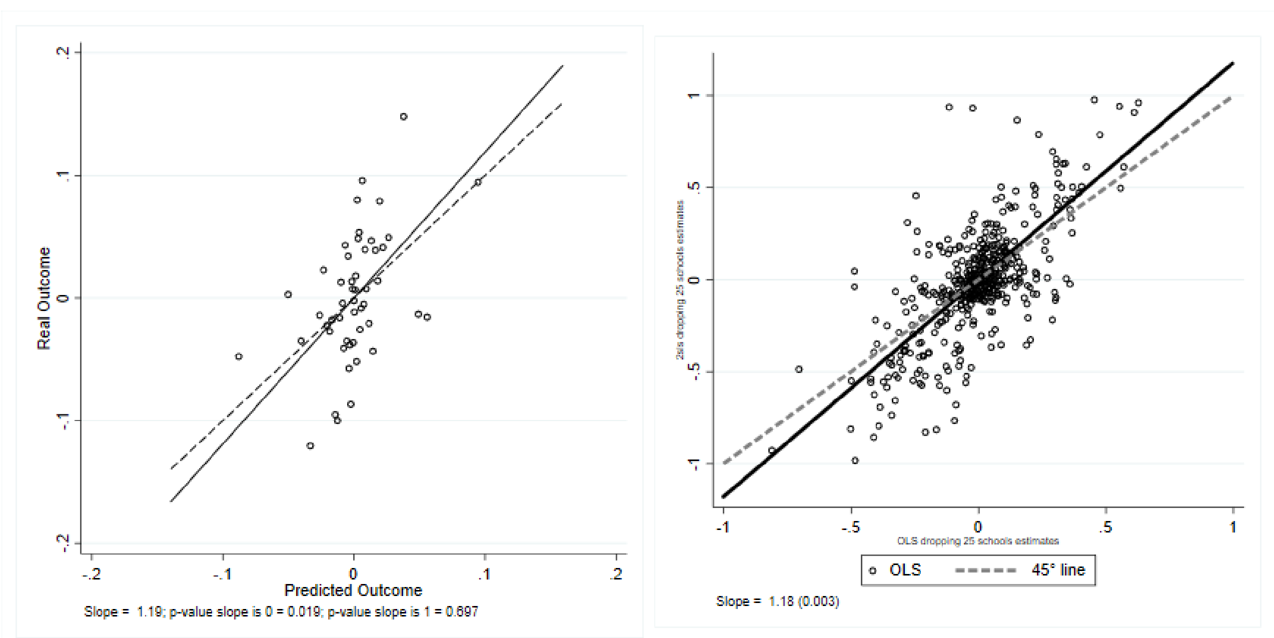
Using this approach, the estimated slope is 1.01. It is statistically different from zero but not from one. Figure 3 in the main text shows scatterplots of the actual and predicted RD estimates for all of the outcomes combined, where all cutoffs are weighted by the inverse of the squared standard error of the coefficient.

A similar procedure can be done for the OLS estimates derived from equation (3) in the main text. In this case, we use the OLS estimates, $\hat{\theta}_{\tau}^{TOT_{OLS}}$, instead of the IV TOT school impacts as below.

$$\hat{\theta}_{\tau}^{TOT_{OLS}} = Above_{ij} \cdot \eta_j + f(SEA_i) + \mathbf{X_i'}\delta + \varepsilon_{ij} \tag{17}$$

The results from this approach are shown in Figure C1, in which the estimated slope is 1.19, 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 are equal to the TOT ones. This model (right panel) yields an almost identical slope of 1.18 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 TOT estimates yield the same effects in expectation. Given that the TOT estimates are almost identical to those implied by the RD model, we take this as evidence of bias in the OLS estimates.

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



Notes: On the left panel, the X-axis represents the estimated coefficients on the 'Above' indicator resulting from model (17); estimated for each school j and for each outcome (estimated school impacts enter as dependent variables). The Y-axis represents the estimated coefficients on the 'Above' indicator resulting from model (12); estimated for each school j and for each outcome (individual level outcomes enter as dependent variables). The connected lines represent lineal fits; 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.

All schools estimated effects are weighted by the inverse of the squared standard error of each estimated coefficient. Outliers above 4 standard deviations away from the mean were removed. On the right panel, the X-axis represents the estimated TOT school impacts from model (4) and (5). The Y-axis represents the estimated OLS school impacts resulting from model (3).

Magnitude of school impacts

This section reports Table C1 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).[40] 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 C1. Standard Deviation of School Impacts

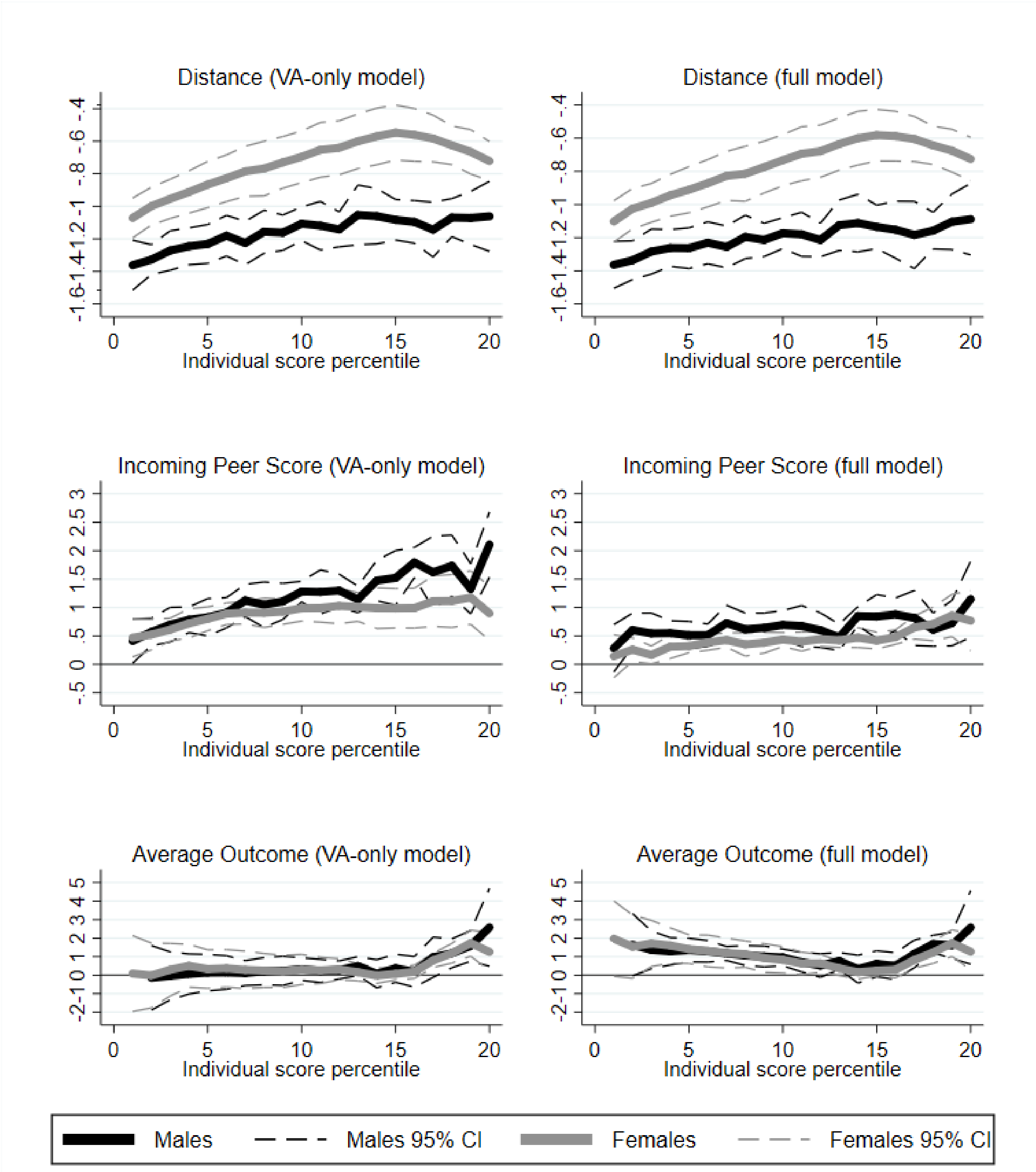
	School Level ($\sigma_{\theta_j^{TOT}}$)			School Level ($\sigma_{\theta_j^{TOT}}$)			School-Year Level ($\sigma_{\mu_{jp}}$)		
Outcome	Size of Impact	[95% CI]		Size of Impact	[95% CI]		Size of Impact	[95% CI]	
Low-Stakes Index									
All Schools	0.433	0.381	0.493	0.418	0.364	0.480	0.164	0.142	0.190
Dropping Outliers	0.427	0.375	0.486	0.412	0.359	0.474	0.149	0.129	0.171
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
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
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
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
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. We took out students who attended schools with low compliance or with outlier value addeds (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 where 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.

[40] For 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.

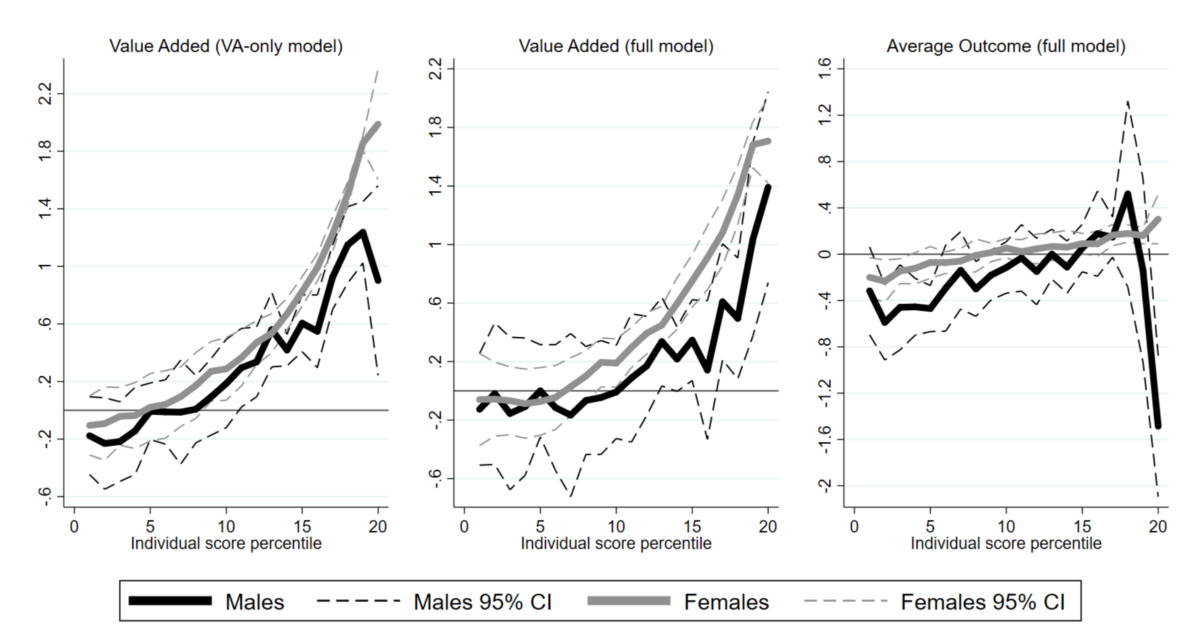
Appendix D: Choice Model using Out of Sample Value-Addeds

Figure D1: Distance, Peer Quality, and Admission Probability



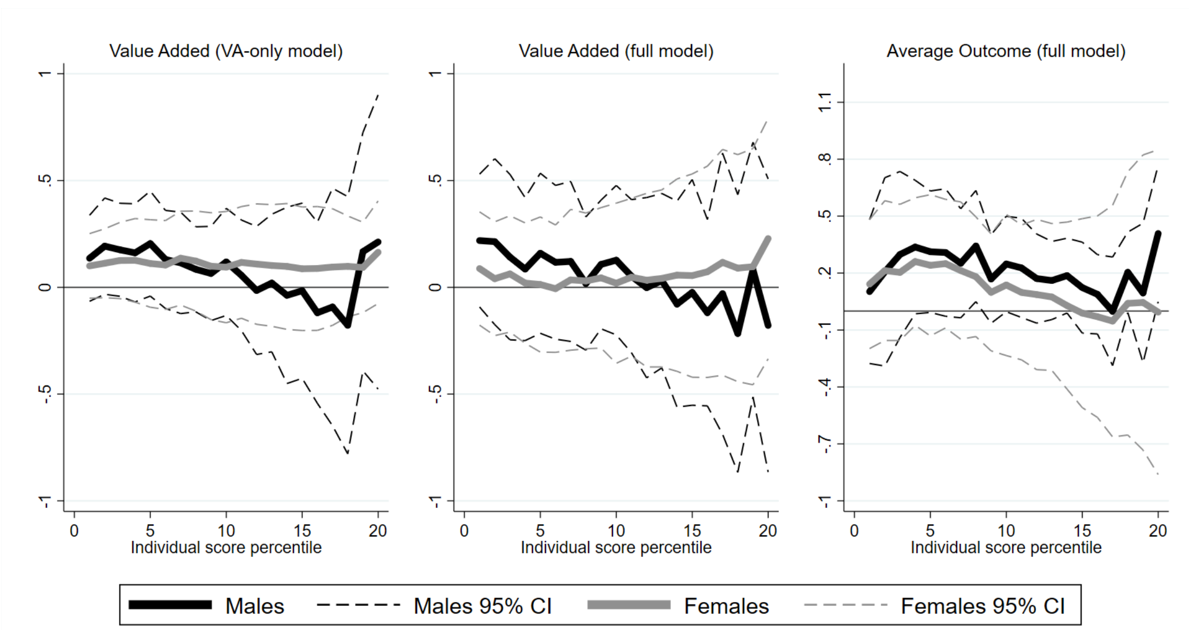
Notes: The connected lines represent the estimated coefficients for the outcome, computed separately for each (SEA score ventile)×(gender) cell for two different models: VA-only model (left panel), and full model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure D2: High-Stakes Index



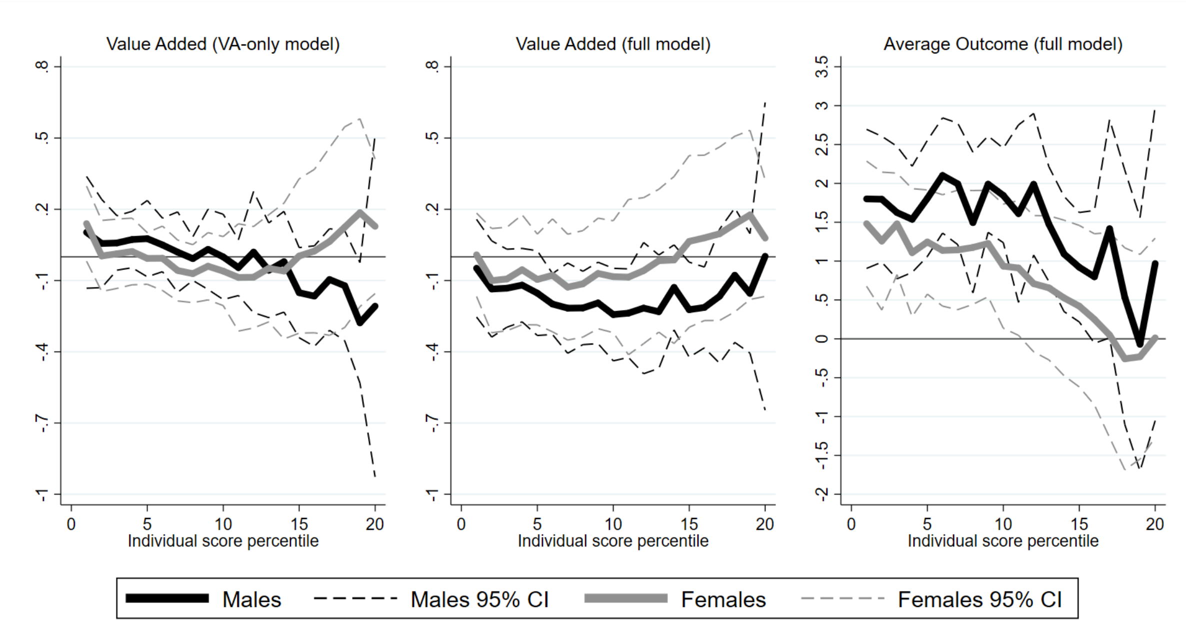
Notes: The connected lines represent the estimated coefficients for the outcome, computed separately for each (SEA score ventile)×(gender) cell for two different models: VA-only model (left panel), and full model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure D3: Low-Stakes Index



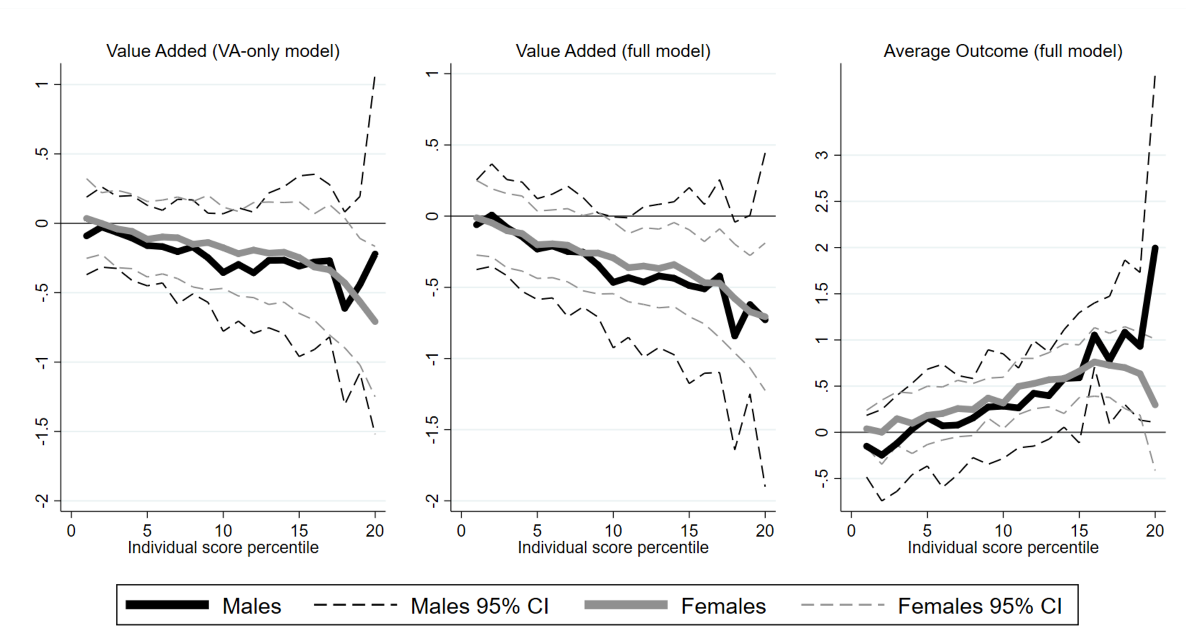
Notes: The connected lines represent the estimated coefficients for the outcome, computed separately for each (SEA score ventile)×(gender) cell for two different models: VA-only model (left panel), and full model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure D4: No Dropout by Age 14



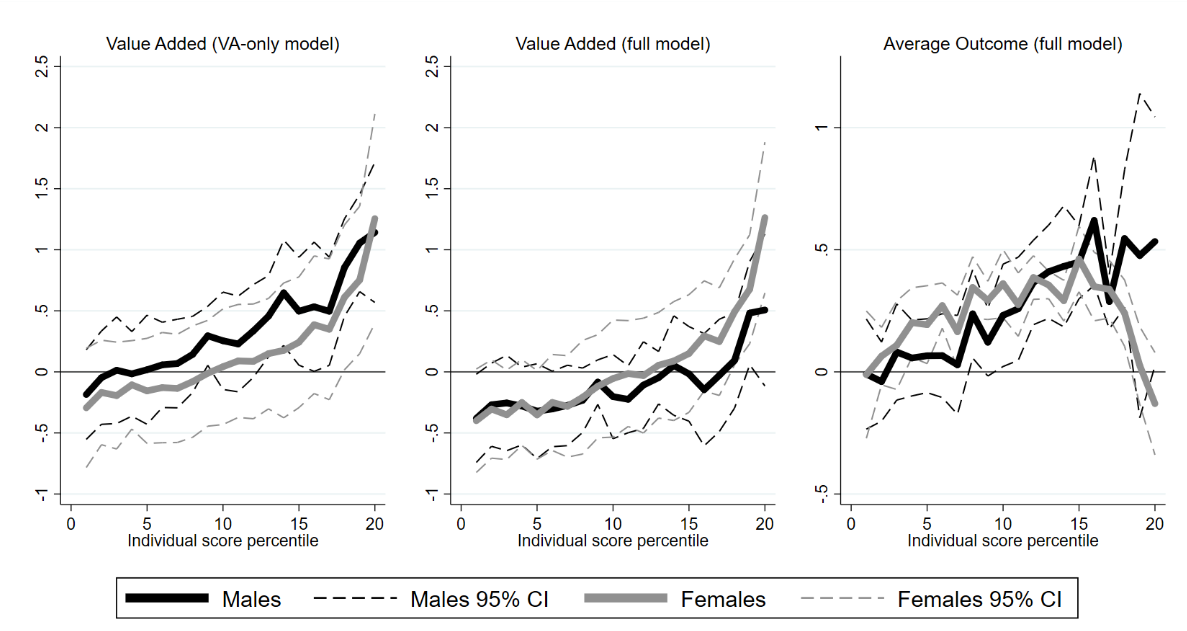
Notes: The connected lines represent the estimated coefficients for the outcome, computed separately for each (SEA score ventile)×(gender) cell for two different models: VA-only model (left panel), and full model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure D5: No Live Birth by Age 19



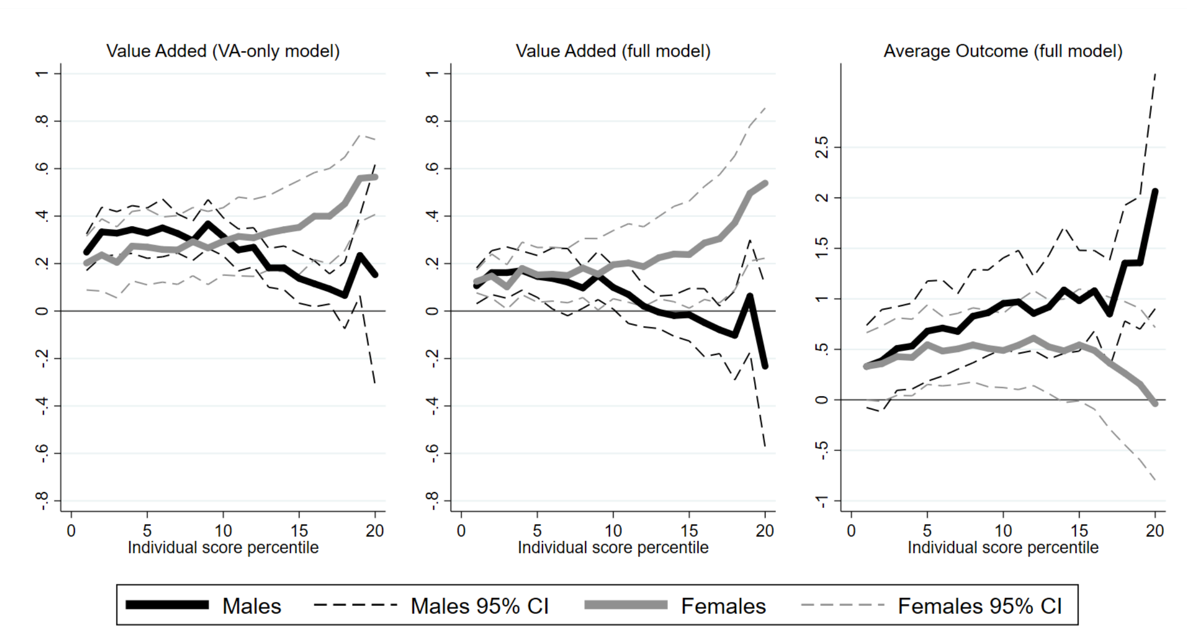
Notes: The connected lines represent the estimated coefficients for the outcome, computed separately for each (SEA score ventile)×(gender) cell for two different models: VA-only model (left panel), and full model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure D6: Not Arrested by Age 18



Notes: The connected lines represent the estimated coefficients for the outcome, computed separately for each (SEA score ventile)×(gender) cell for two different models: VA-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 for the outcome, computed separately for each (SEA score ventile)×(gender) cell for two different models: VA-only model (left panel), and full model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.