Northwestern University

From the SelectedWorks of C. Kirabo Jackson

2018

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

C. Kirabo Jackson, *Northwestern University*
Diether W. Beuermann, *Inter-American Development Bank*
Laia Navarro-Sola, *Northwestern University*
Francisco Pardo, *Inter-American Development Bank*

Available at: https://works.bepress.com/c_kirabo_jackson/39/

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

December 10, 2018

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 high-stakes tests, longer-run outcomes, or both? To answer the first question, we apply quasi-experimental methods to data from Trinidad and Tobago and estimate the causal impacts of individual schools on several 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 and employ discrete choice models to estimate parental preferences. Parents value schools that causally improve high-stakes test scores conditional on average outcomes, proximity, and peer quality. Consistent with parents valuing the multidimensional output of schools, parents of high-achieving girls prefer schools that increase formal labor market participation, and parents of high-achieving boys prefer schools that reduce crime. (JEL I20, J0)

*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 her 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 Bernard Smith and Feyaad Khan for granting us authorization to work in their facilities to match formal employment participation records maintaining individual confidentiality fully protected. Tatiana Zarate and Diego Zuniga provided excellent research assistance. The statements made and views expressed are solely the responsibility of the authors.

I Introduction

Schools aim to provide students with the various skills required to be productive in the labor market, and to contribute positively to society (Goodlad 1984; Hay and Hodgkinson 2006; Postman 1996). However, researchers, practitioners, and policy-makers rely heavily on school's academic performance on standardized tests as a measure of quality. In many settings such as New York City, England, Trinidad and Tobago, and others, schools are ranked based on their average test-score performance and this information is disseminated widely as a means to inform parents. However, this information may be misleading for two reasons. First, because students may select to schools based on ability, average outcomes may not reflect schools' contribution to those outcomes (Abdulkadiroğlu et al. 2014; Lucas and Mbiti 2014). Second, because educational output may be multidimensional (Jackson 2018; Heckman et al. 2006; Kautz et al. 2014; Hanushek 1971), schools that improve test scores may not be those that improve students' life chances. Each of these issues is potentially problematic because (a) policy-makers may be using poor measures of school quality for policy decisions (such as which schools to close, or which schools to emulate, etc.), and (b) parents may be using poor measures of school quality when making investment decisions for their children – undermining many of the potential benefits of school choice (Friedman, 1955).

Because it is difficult to obtain causal estimates of individual schools on both test scores and longer-run outcomes, there is little evidence on the extent to which schools' test score impacts predict improved adult outcomes. Also, because it is difficult to obtain causal estimates of schools linked to data on parent choices, there is little evidence on whether parental preferences for schools are based on schools' average outcomes, schools' impacts on test scores and/or schools' causal impacts on overall well-being. To shed light on these issues, we seek to answer three broad questions: Is a school's impact on test scores a good measure of a school's overall impact on students? Do parents value those schools that have positive causal impacts on student outcomes? Can parents disentangle a school's causal impacts on outcomes from raw school-level averages? Do parents value school impacts on outcomes other than highly publicized high-stakes tests?

This paper proceeds in two parts. First we examine the extent to which school output is multidimensional by estimating school impacts on test scores and a wide array of academic, social, and labor market outcomes. Using these estimates, we assess the extent to which schools vary in their ability to improve different student outcomes. Importantly, we use quasi-random variation to validate our estimated school impacts as reflecting causal impacts. Next, we merge our estimated causal impacts to parents' school rankings and explore whether parents value school impacts on multiple dimensions as opposed to just school's test score effects. We also examine the extent to which parents value schools' causal impacts above and beyond easily observable school attributes such as peer quality and average test scores. Finally, we examine heterogeneity in preferences to

describe the different ways certain kinds of parents may define "good" schools.

To examine these questions we use data on all applicants to public secondary schools in Trinidad and Tobago between 1995 and 2012. These data contain students' identifying information, scores on the Secondary Entrance Assessment (SEA) (taken at age 11 at the end of 5th grade), and a ranked list of secondary schools the student wished to attend. We link these data (at the student level) to scores on low-stakes national exams taken three years later (at the end of 8th grade), 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 between 1990 and 2016, birth registry data between 2010 and 2016, and retirement contribution fund data as of May 2017. 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 (academic, social, etc.).

To estimate the effect of 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 (in this case, the secondary entrance exam score and the ranked school choices) the administratively assigned school is unrelated to the characteristics of students (Jackson, 2010). Exploiting this fact, we first estimate the intention-to-treat effect of being *assigned* to a school by comparing the average outcomes of students who made the same school choices and had similar incoming scores but were assigned to different schools due to discontinuities in the algorithm. In a second step, we use these intention-to-treat assigned school estimates to form treatment-on-the treated causal impacts of *attending* each school. Unlike existing work that relies on selection on observables assumptions (e.g. Abdulkadiroglu et al. 2017), we rely on quasi-random variation that is arguably more credibly exogenous. We show empirically that this matters when estimating preferences. Moreover, we can validate our estimated school impacts as being causal using a regression discontinuity design.

To estimate parent's preferences for schools, we make use of the fact that students must submit a ranked list of four secondary schools as part of the secondary school application process. The optimal strategy under deferred acceptance assignment algorithms has been studied by economists for some time. Chade and Smith (2006) demonstrate that the a students' optimal top choice school should be their preferred school among all possible choices. Also, Dubins and Freedman (1981) and Roth (1982) show that, among the set of schools listed, truthfully ranking schools is a weakly dominant strategy.[1] As such, similar to Abdulkadiroglu et al. (2017) and Hastings et al. (2005), we assume that the rankings follow this optimal strategy and estimate preferences using a discrete choice model (McFadden, 1973). Specifically, we implement a modified rank-ordered logit (or exploded multinomial logit) model to estimate parent preferences for schools based on observed

[1]More recently, it has been found that under the algorithm used to assign students to schools, students have incentives to truthfully reveal their preference rankings (Haeringer and Klijn 2009; Pathak and Sönmez 2013).

peer quality, average school characteristics, and estimated school impacts.

We find that schools have meaningful 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 live birth, performance on high-stakes school leaving exams, being arrested, and formal labor market participation. Being assigned to a school that improves these outcomes in other years (i.e. out of sample) is associated with improvements in that same outcome. Going from a school at the median of the impact distribution to one at the 85th percentile would increase low-stakes test scores by 0.27σ, increase high-stakes test scores by 0.29σ, reduce school dropout by 6.5 percentage points, reduce teen births by 1.4 percentage points, reduce teen arrests by 1.4 percentage points, and increase the likelihood of being employed in the formal labor market by 1.8 percentage points.

While we find school impacts on a range of outcomes, the correlations between school impacts on high-stakes tests and other outcomes is surprisingly low. For example, the correlation between school impacts on high-stakes academic tests and on reducing teen births is only 0.06. In fact, the correlation between a school's impact on high-stakes tests and being formally employed is actually *negative* 0.08. Generally, the patterns suggests 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 and higher-performing schools (Burgess et al. 2015; Abdulkadiroglu et al. 2017). 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. We also uncover several *new* patterns.

Adding to the literature, we distinguish preferences for average outcomes from preferences for better peers, and find that parents' choices are more responsive to average *incoming* test scores, than average secondary school test scores. This suggests that parental preferences for higher-achieving schools are driven, in large part, by a desire for higher-achieving peers (Epple and Romano, 1998). Another innovation of our work is the reliance on clean estimates of schools' causal impacts to disentangle parental preferences for high-achieving schools from those that are effective at improving outcomes. We find that parents have strong preferences for schools that have high causal effects on high-stakes exams. This pattern is robust to controls for peer quality and average outcomes.[2]

[2]Our finding that parents value school impacts conditional on peer quality stands in contrast to Abdulkadiroglu et al. (2017). We discuss possible explanations for such differences in Section V.4.

This suggests that parents can disentangle effective schools from schools with strong performance. Consistent with parents valuing school impacts on non-academic outcomes, parents prefer schools that improve formal labor market participation, and schools that reduce arrests. Models that allow for heterogeneity in preferences reveal some key patterns. Males choices are sensitive to the arrest impacts, such that males are more responsive to a school's impact on arrests as they are to a school's impact on high-stakes tests. Similarly, girls' school choices are responsive to teen pregnancy measures such that girls prefer schools with low pregnancy rates (but do not appear to be sensitive to schools impacts on pregnancy). Also, parents of girls prefer schools that improve low-stakes tests, reduce high-school dropout and increase labor market participation more than boys. These results paint a nuanced picture of parental preferences such that parents value school impacts on both academic and nonacademic outcomes and that there is heterogeneity in parental preferences for various dimensions of schools' output.

We contribute to existing literatures in many ways. In work on school impacts, researchers have documented that schools have causal impacts on crime (Deming, 2011), college going, and earnings (Sass et al., 2016). We build on this literature by also examining school impacts on low-stakes tests, dropout, teen motherhood and formal labor market participation. More importantly, we present the first analysis of the relationships between school impacts on all these different outcomes – directly exploring the multidimensionality of school quality.[3] We document that secondary schools that increase high-stakes test scores improve some non-academic and longer-run outcomes but not others. This finding is important because test-score value-added measures of school quality are part of school accountability report cards and are increasingly used for policy decisions.

In existing work on parental preferences, Hastings et al. (2006) estimate parental preferences for school average test scores, and Abdulkadiroglu et al. (2017) use school value-added estimates (on test scores and high-school graduation) to disentangle parental preferences for high value-added schools from that of schools with better average outcomes. We build on this work in two important ways. First, we move beyond academic outcomes and provide the first study that also estimates parental preferences for school impacts on fertility, crime, and labor market outcomes. Another key innovation of our study is that we rely on exogenous variation to identify school impacts. This is important because, as we show, even small biases in school impacts that are correlated with peer quality can lead to large biases in non-linear models that seek to disentangle preferences for school value added from that of peers. This paper is among the first to include causal school impacts in a school choice model, and the first to include causal impacts on a wide range of outcomes such as crime, teen motherhood, and formal labor market participation.[4] By using unbiased measures

[3] In a recent working paper Dobbie and Fryer (2016) examine the relationship between charter school impacts on test scores, high school graduation, and earnings.

[4] In closely related work Abdulkadiroglu et al. (2017) find that parents do not respond to school effectiveness conditional on peer quality. Our work finds a different result and suggests that parents are able to disentangle school

of school impacts on a wide set of outcomes, this study may shed new light on (a) the relationship between school effects on different kinds of outcomes and (b) parental preferences for schools' causal impacts on both academic and nonacademic outcomes.

The remainder of this paper is as follows; Section II describes the Trinidad and Tobago context and discuss the data used. Section III presents our empirical strategy for estimating school value-added. Section IV presents our value-added estimates and the validation of these estimates. Section V discusses our choice data, presents our choice models, and presents our estimates of parental preferences. Section VI concludes.

II The Trinidad and Tobago Context and Data

The Trinidad and Tobago education system evolved from the English system. At the end of primary school (after grade 5, typically between 10 and 11 years old), students register 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 then assigned 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.[5] There are two types of public secondary schools: Government schools (fully funded and operated by the government) and Government Assisted schools (managed by private bodies, usually a religious board, and all operating expenses funded by the government). There were 152 public secondary schools within 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. These are key outcomes in this study. The first secondary exam is the National Certificate of Secondary Education (NCSE) taken at the end of form 3 (grade 8) by all students (both in public and private schools) in eight subjects.[6] NCSE performance has no consequences in terms of school progression or future admission decisions into 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

quality from raw averages. We explore possible explanations for this different result in Section V.4. This paper also differs from Abdulkadiroglu et al. (2017) because we consider school effects on a broad range of outcomes while they consider only scholastic outcomes (which are closely related to test scores). This innovation is important.

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

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

examinations are given in 33 subjects. To be eligible for university admission, one must pass five or more subjects including the two core subjects of 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. Therefore, the CSEC is a high-stakes examination.

The CAPE is the equivalent of the British Advanced levels exam and was launched in 2005. The full CAPE program lasts for two years and includes three 2-unit subjects (each unit taken in different academic years) 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 Studies and Communication Studies. Finally, students 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). These are highly competitive scholarships and an average of only 165 scholarships per year (i.e. less than 1 percent of students) were awarded since its inception in 2005 until 2016. Given this, the CAPE is clearly a high-stakes examination. Next, we describe the data used for the analyses.

Secondary School Applications Data: The data include the full population of all students who applied to a public secondary school in Trinidad and Tobago between the years 1995 and 2012. We obtained the official administrative SEA data for each of these years. 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 the student wished to attend, and the administrative assignment by the MOE. The final SEA dataset contains information on about 320,000 students across the 18 SEA cohorts.

Examination Data: To track students' examination performance and educational attainment we collected data on the NCSE examinations (taken 3 years after secondary school entry, typically at age 14), the CSEC examinations (taken 5 years after secondary school entry, typically at age 16) and the CAPE examinations (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 students name, date of birth, gender and scores for the eight subjects assessed. The NCSE data were linked to the 2006 through 2012 SEA cohorts by full name (first, middle, and last), gender, and date of birth.[7] The CSEC data are available for all years between 1993 and 2016. These data include the students name, date of birth, gender and scores for each

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

subject examination taken. The CSEC data were linked to the 1995 through 2011 SEA cohorts by full name (first, middle, and last), gender, and date of birth.[8] The CAPE data are available for years 2005 through 2016, and are linked to the 1999 through 2009 SEA cohorts by name, gender, and date of birth.[9] As with the CSEC, these data contain scores for each examination 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 following information: 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 by full name (first, middle, and last), gender, and date of birth.

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 following information: the mother's full name, date of birth, gender, and date of the live birth. To explore teen live birth, these data were linked to the 2004 through 2010 SEA cohorts by full name of the mother (first, middle, and last) and her date of birth.

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 national social security system as of May 2017. For each affiliate, the data include the following information: 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 21 and 32 years old, these data were linked to the 1995 through 2006 SEA cohorts by full original name (first, middle, and last), gender, and date of birth.

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 243 students (column 1). Overall about 87.5 percent of students took the NCSE and 75.7 percent took at least one CSEC subject. The average student passed about 3.2 CSEC subjects and 34.4 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 0.24 standard deviations lower for males than for females (columns 2 and 3) and average scores of those assigned to the top ranked schools are 1.26 standard deviations higher than those assigned to the bottom ranked schools (columns 4 and 5). In top schools, the average cohort size is about 137 students while that for the least selective schools is about 346 students. There are relatively more females at selective schools. Given the differences in incoming

[8] 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.

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

scores, this is not surprising. Female students have lower dropout rates by age 14 since 89 percent took the NCSE compared to 85.8 percent among males. Likewise, girls score 0.39 standard deviations higher in the NCSE. Also, 41.0 percent of female students qualify for tertiary education while only 27.5 percent of males do. Students at the most selective schools score 0.84 standard deviations higher in the NCSE than the average student at less selective schools. They also pass about 5.3 CSEC subjects on average, and 65.2 percent qualify for tertiary education; while this is only accomplished by 16.3 percent of students at the least selective schools (column 5).

In terms of post-secondary education, about 19.8 percent of students took at least one CAPE unit and passed 1.4 units (column 1). About 14.8 percent of students earned an Associate's degree, and only 0.97 percent earned the prestigious CAPE scholarship. Female students passed 1.7 CAPE units, and 18.6 percent earned an Associate's degree. In comparison, males passed 1.1 units, and only 11.0 percent earned an Associate's degree. Among students at the most selective schools, 38.8 percent took at least one CAPE unit and 30.3 percent earned an Associate's degree. Among those at less selective schools, only 6.1 percent took at least one CAPE unit and 3.4 percent earned an Associate's degree.

Moving to nonacademic outcomes, 3.4 percent of the population had been arrested by age 18. However, arrests are concentrated among males of which 6 percent had been arrested by age 18. Arrests rates are low (1.5 percent) among students from more selective schools, and are much higher (4.3 percent) among students at the least selective schools. A similar pattern is observed for teen motherhood. While 5.9 percent of girls at the top schools had a live birth before age 19, as much as 14 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. In the next section, we describe our empirical approach to estimate secondary school causal impacts on these various outcomes of interest.

III Estimating School Impacts

Our first objective is to estimate the causal impacts of schools on a broad array of academic, nonacademic and social outcomes. One distinguishing feature of our approach to estimating school impacts is that we do not rely on a selection on observables assumption, but rather isolate exogenous variation in school attendance to identify the causal impacts of schools on outcomes. Following Jackson (2010), we exploit the fact that the Ministry of Education (MOE) uses a deferred acceptance mechanism to create an initial set of school assignments for students. Specifically, all students submit a rank-ordered list of secondary schools they wish to attend *before* they sit the Secondary Entrance Examinations. 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. This assignment mechanism generates test-score cutoffs for each secondary school above which applicants are admitted and below which they are not – this feature is central to our empirical design.[10] We refer to this purely rule-based initial assignment as the "tentative" assignment.

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 assignments. As such, while the official MOE assignments are potentially subject to some manipulation by parents or principals, the tentative rule-based assignment is not. We use students' initial tentative assignments (solely due to the algorithm) to identify the selection-free causal impacts of schools on outcomes.

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 test 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. Accordingly, 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 test 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 (1) by Ordinary Least Squares (OLS).

$$Y_{i\tau ct} = I_{i,J=j} \cdot \theta_j^{ITT} + f(SEA_i) + \lambda_c + \mathbf{X_i'}\delta + \varepsilon_{ijct} \tag{1}$$

In (1), $Y_{i\tau ct}$ is the outcome of interest for student i who attended school τ, and belongs to choice group c in cohort t. $I_{i,J=j}$ is an indicator variable equal to 1 if student i was *tentatively assigned* to school j. **School j and 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. The choice group fixed effects, λ_c, denote groups of individuals who made the same school choices.[11] **Xi** is a vector

[10] 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. See Appendix A for a detailed description of the assignment process.

[11] Given that for SEA cohorts 2001-2006 the MOE allowed students to list up to 6 different school choices (instead

of individual-level baseline characteristics including sex, district of residence fixed effects, and religion fixed effects; while ε_{ijct} is an individual-level disturbance. Within this context, the estimated $\hat{\theta}_j^{ITT}$ from (1) identifies the average treatment effect of being tentatively assigned to school j so long as the contribution of the choice groups and that of test scores are additively separable. To assuage concerns that our estimated school impacts are not valid, we validate our effect estimates using a regression discontinuity design in Section IV. Because not all students who are tentatively assigned to school j actually attend school j, this is an Intention To Treat (ITT) estimate.

While $\hat{\theta}_j^{ITT}$ is a valid estimate of the effect of being tentatively assigned to school j, it is not an estimate of the effect of *attending* school j. To obtain the Treatment-On-the-Treated (TOT) estimate of the effect of *attending* school j, we scale each school's ITT estimate by the out-of-sample proportion of students tentatively assigned to school j who indeed ended up attending school j.[12] Specifically, where $\bar{p}_j$ is the proportion of students tentatively assigned to school j by the algorithm who attended school j, our treatment on the treated estimate of the effect of attending school j is as below.[13]

$$\hat{\theta}_j^{TOT} = \hat{\theta}_j^{ITT} / \bar{p}_j \tag{2}$$

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, crime outcomes, teen motherhood, and formal labor market participation. Using these estimates, in the next section we examine the extent to which school impacts on these different outcomes are correlated with each other.

III.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. To test for this empirically, we conduct factor analysis using both the principal and the principal-component factor methods to explore the extent to which school impacts on these different outcomes represent a single dimension of school quality, or if some schools improve some outcomes but not necessarily others. Both approaches yield very similar results. Statistically, this is essentially a test of whether the covariance in school impacts can be explained by a single underlying factor (i.e. there is a single dimension of school quality), or if the data are more con-

of the usual 4 choices), to avoid an excessive number of choice groups with few observations within each group, we grouped students within unique combinations of choice group-cohort for the first 3 choices and in separate fixed effects for the following 3 choices.

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

[13] After estimating our school impacts, we drop outlier school impacts that were more than 4sd away from the main impacts for that outcome. This resulted in our dropping 22 school-year observations out of 2725.

sistent with there being multiple factors (i.e. dimensions of school quality). The results from the exploratory factor analysis using the principal factor method are reported in Appendix Table B1.[14]

The first key result is that there is no single underlying factor that explains all of the covariance in school impacts.[15] That is, the data are not consistent with a model in which some schools are good at improving all outcomes, while others are bad at improving all outcomes. Instead, the data suggest that school output is multidimensional. Specifically, the data are consistent with there being 5 or 6 distinct dimensions.[16] As shown in Appendix Table B1, Factors 1 and 3 load on the two high-stakes exams (with the CAPE exams lading more heavily on Factor 1 and the CSEC loading more heavily on Factor 3). All of the arrest outcomes load heavily on Factor 2, while the low-stakes academic and non-academic outcomes load on to Factor 5. Being formally employed loads in factor 4, while having no live births by age 19 and not dropping out of school by age 14 do not follow any systematic pattern. In sum, school effects are multidimensional, and school impacts on similar outcomes tend to be related to each other.

Because we have several outcomes and many reflect the same basic dimension, we combine similar outcomes into indexes. Informed by the patterns in the exploratory factor analysis, we grouped variables that tended to move together and were conceptually similar to create six indexes from the seventeen individual outcomes. We created a "High-Stakes Exams" index by running a factor analysis (using the principal-component factor method) on all the CSEC and CAPE outcomes and then predicting the first unrotated factor. Using this same approach, we computed a "Low-Stakes Exams" index grouping both NCSE academic and non-academic performance. We also combined all the crime related measures into a "Crime" index. Table 2 shows the individual variables that comprise each index and the weights used to compute each index. Because they are conceptually distinct, no dropout by age 14, no live birth by age 19, and formal labor market participation each constitute their own single variable index. All indexes were standardized to have a mean of zero and have unit variance.

Having consolidated 17 individual outcomes into 6 indexes, we now estimate TOT school impacts on the six indexes directly (as opposed to combining the school impacts on the individual outcomes). Table 3 presents the correlations between estimated school impacts on the six indexes.

[14] Except for the number of arrests, all measures have been coded so that higher values reflect better outcomes.

[15] In the factor analysis using the principal factor method (Appendix Table B1), the first rotated factor explains about 38 percent of the variability in the data and has variance of 4.11, the second explains 37 percent and has a variance of 4, and the third explains about 10 percent and has a variance of 1.05. In the factor analysis using the principal-component factor method (not shown), the first rotated factor explains about 27 percent of the variability in the data and has a variance of 4.51, the second explains 26 percent and has a variance of 4.35, and the third explains about 10 percent and has a variance of 1.64. In sum, irrespective of the model chosen, the data are inconsistent with there being a single underlying dimension of school quality that explains covariance across school effects on all the outcomes.

[16] In the factor analysis using the principal-component factor method (not shown), the variance for the fourth factor is 1.30 and for the fifth and last factor 1.23. Consistent with this, in the factor analysis using the principal factor method (Appendix Table B1), the first 5 factors explain 98 percent of the variance in school impacts.

Note that each index was computed in a separate factor model so that the indexes can be correlated. To aid interpretation, all indexes are coded so that higher values reflect better outcomes. Table 3 reveals that the correlations between school impacts across dimensions are relatively small – suggesting that school output is multidimensional and that different schools are good at improving different student outcomes. The correlations between high-stakes exam impacts and other outcomes are remarkably low. The correlation between school impacts on the high-stakes exam index and teen motherhood is only 0.06, there is no correlation with the crime index, and the correlation with being formally employed is negative 0.08. This suggests that the schools that improve the high-stakes exams are associated with small improvements in teen motherhood, no improvement in crime outcomes, and a slight *decrease* in formal labor market participation. The correlation between performance on high-stakes and low-stakes exams is positive, but moderate ($r = 0.16$). This suggests that schools that improve low-stakes exams performance also tend to improve high-stakes performance, on average, but that only about 2.6 percent (i.e. 0.16 x 0.16 = 0.026) of the variation in school impacts on low-stakes exams 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 low-stakes exams has been documented in other settings (e.g. Mbiti et al. 2018).

Overall, the patterns in Table 3 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 statistic" for the quality of that school and that 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. However, the extent to which this matters depends on the size of school impacts on outcomes other than test scores. For example, if school impacts on dropout are small while impacts on test scores are large, then school impacts on test scores may capture much of the variation in school quality that matter in the long run. Conversely, if school impacts on dropout or crime are large, then the lack of any strong correlation with test score impacts would suggest that test score impacts are a very incomplete measures of schools' overall quality. We examine whether school impacts on the measured dimensions are economically meaningful in Section IV.

IV Magnitude of the School Effects

To assess the magnitude of the school impacts on each outcome, we first estimate the standard deviation of the intention to treat (ITT) impacts. Following Jackson (2013) we do this in two steps. In the first step we estimate equation (3) by OLS where θ_{jt}^{ITT} are *assigned* school-by-year fixed effects and all other terms are defined as in equation (1).

$$Y_{i\tau ct} = \theta_{jt}^{ITT} + f(SEA_i) + \lambda_c + \mathbf{X_i'}\delta + \varepsilon_{ijct} \tag{3}$$

Then we remove the contribution of incoming test scores, choices, and student-level covariates, resulting in the estimated school-by-year fixed effects plus the student-level residual (i.e. $\hat{\theta}_{jt}^{ITT} + \hat{\varepsilon}_{ijct}$). In the second step we decompose this combined residual into a permanent school effect (θ_j^{ITT}), a transitory school-by-year effect (μ_{jt}), and student-level random disturbances (ε_{ijct}). Under the assumption of joint normality of these components and the covariance structure in (4), one can uncover Maximum Likelihood estimates of the variance of the assigned school impacts ($\sigma^2_{\theta_j^{ITT}}$), the variance of the transitory assigned school-by-year impacts ($\sigma^2_{\mu_{jt}}$), and the variance of the student-level disturbances ($\sigma^2_{\varepsilon_i}$).

$$\begin{bmatrix} \theta_j^{ITT} \\ \mu_{jt} \\ \varepsilon_i \end{bmatrix} \sim N\left(0, \begin{pmatrix} \sigma^2_{\theta_j^{ITT}} I_J & 0 & 0 \\ 0 & \sigma^2_{\mu_{jt}} I_M & 0 \\ 0 & 0 & \sigma^2_{\varepsilon_i} I_N \end{pmatrix}\right) \tag{4}$$

Because not all students attend the school to which they are assigned, the maximum likelihood estimate of $\sigma_{\theta_j^{ITT}}$ will necessarily be smaller than the standard deviation of the attended school impacts $\sigma_{\theta_j^{TOT}}$. However, if all schools had the same compliance rate $\bar{p}$, then $\sigma_{\theta_j^{TOT}} = \sigma_{\theta_j^{ITT}}/\bar{p}$. Average compliance rates are about 50 percent (i.e. $pr(j = \tau) \approx 0.5$), so the standard deviation of the attended school impacts would be about twice as large as those estimated with the maximum likelihood approach. Table 4 reports the estimates of the standard deviation of the assigned school impacts ($\hat{\sigma}_{\theta_j^{ITT}}$) for each outcome along with their respective 95 percent confidence intervals.

As a complementary approach, and to estimate the the variance of the *attended* school impacts (i.e. the TOT effects) directly, we follow Kane and Staiger (2008) and Jackson (2018) in the teacher quality literature. Specifically, we estimate the variance of the persistent school effect on outcomes using the covariance of the school-year effects for the same school across time. That is, we compute the TOT estimate for each school in each year $\hat{\theta}_{jt}^{TOT}$ (using only data from year t). So long as the idiosyncratic school-year errors are not correlated over time and are uncorrelated with the school effects, then for any two years t and r, it follows that $cov(\hat{\theta}_{jt}^{TOT}, \hat{\theta}_{jr}^{TOT}) = var(\theta_j^{TOT})$. To estimate this, we pair each school-year estimate with all other school-year estimates for the same school and compute the sample covariance. To provide an estimate of the standard deviation of the school impacts for each outcome, we report the square root of the sample covariance for each outcome in Table 4. To allow for a direct comparison with the maximum likelihood approach, we also estimate the variance of the ITT impacts using the covariance approach. Reassuringly, the ITT covariance estimates are very similar to the estimates from the maximum likelihood approach, and the estimated TOT standard deviations (based on the covariance approach) are approximately twice as large as those of both ITT estimates.

Low-stakes test scores (age 14): The standard deviation of the ITT school effects on the low-

stakes index is 0.149 (p-value$<$0.05). That is, being assigned to 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 0.149 standard deviations. The covariance-based approach yields almost identical results with an ITT standard deviation of 0.144. The TOT school effect estimate from the covariance approach is 0.298 standard deviations (which is almost exactly twice as large as the ITT effect as hypothesized above based on the compliance rate). Looking at the TOT impacts, the estimates indicate that *attending* a school at the 85th percentile of the impact distribution compared to *attending* a school at the median would increase low-stakes test performance by approximately 0.298 standard deviations.

High-stakes exam outcomes (age 16 through 19): Recall that our high-stakes index combines two high-stakes exams. The assigned school effects for the high-stakes dimension have a standard deviation of 0.123 (p-value$<$0.05). This indicates that being assigned to a school at the 85th percentile of the impact distribution compared to being assigned to a school at the median would increase high-stakes test performance by approximately 0.123 standard deviations. As with the low stakes exams, the covariance-based approach yields almost identical results. Looking at the TOT impacts, the estimates indicate 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 0.27 standard deviations. These estimated school impact sizes are slightly larger than those found for school impacts on test scores in North Carolina (Jackson 2013; Deming 2014), smaller than the estimated impact of attending Boston charter schools (Angrist et al., 2013) and on the same order of magnitude as that of attending Promise Academy in the Harlem Children's Zone (Dobbie and Fryer, 2015). It is also interesting to note that the magnitude of the school impacts on high-stakes and low-stakes tests are very similar, but the two impacts are only weakly correlated with each other – indicating that those schools that raise high-stakes scores and those that improve low-stakes scores are often not the same set of schools.

Dropout: The first non-test-score outcome we examine is dropout. Because all students take the NCSE exams around age 14, our measure of dropout is not being registered for the NCSE exams. To aid interpretation, we present the standard deviation of school impacts on the binary outcomes directly in the lower panel (as opposed to the impacts on the standardized outcome in the top panel). The standard deviation of the ITT school impacts on dropout are 0.03 and 0.028 in the maximum likelihood and covariance estimates, respectively. The estimated standard deviation of the TOT impact is 0.065 – indicating that attending a school at the 85th percentile of the impact distribution compared to attending a school at the median would reduce high school dropout by approximately 6.5 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 within the range of what one might expect based on existing studies.[17]

Examining the impact on high school dropout is worthwhile because it is a strong predictor of overall adult well-being above and beyond test scores. Given the economically important impacts that schools have on dropout, the fact that those schools that raise high-stakes scores and those that reduce dropout are often not the same set of schools suggests that test scores may be an incomplete measure of school quality. This basic pattern exists for all the remaining non-test score outcomes – underscoring the fact that school impacts are multidimensional and a considerable amount of the potential value of attending a school may be unmeasured by impacts on high-stakes tests.

Teen motherhood: Next we examine school impacts on teen motherhood. The standard deviation of the school ITT effect on teen motherhood is 0.03 in the maximum likelihood model. The 95 percent confidence interval is between 0.023 and 0.039 so that this is statistically significantly different from zero. However, the covariance based estimates are somewhat smaller. These indicate that the standard deviation of the school ITT effect on teen motherhood is only 0.016. Moreover, the TOT estimates is only 0.014 (smaller than what the random effects estimates would suggest). The covariance based TOT effects indicate that going from a school at the median to one at the 85th percentile of the impact distribution would reduce teen live births by 1.4 percentage points. Given that the teen live birth rate is around 10 percent, these represent economically important relative impacts. However, the impacts may be modest and therefore difficult for parents to detect. Moreover, the fact that the covariance estimates and the maximum likelihood estimates are less consistent with each-other (though both are positive and on the same order of magnitude) may indicate that persistent school impacts on teen motherhood may be less robust. This is helpful to keep in mind as one interprets the parental preferences for schools in Section V.

Crime outcomes: To aid interpretation, we present estimated school impacts on having any arrest by the age 18 in the lower panel (as opposed to the full crime index in the top panel). We examine the standard deviation of school impacts on teen arrests. Echoing Deming (2011), we find that schools have meaningful and statistically significant impacts on arrests. The standard deviation of the ITT school effect is 0.006 which means that being assigned to a school at 85th percentile of the impact distribution as opposed to the median would reduce the likelihood of being arrested as a teenager by 0.6 percentage points. Reassuringly, the covariance approach yields very similar estimates. The covariance based estimated standard deviation of the TOT impact is 0.013 – indicating that attending a school at the 85th percentile of the impact distribution compared to attending a school at the median would reduce teen arrests by approximately 1.3 percentage points. Again, this is an economically important impact that is unrelated to school impacts on test scores

[17]The literatures on charter schools and choice school often find impact of high school completion between 10 and 15 percentage points. These impacts are larger. Our estimated standard deviation of the school impacts on dropout suggests that these choice schools may be as much as 2 standard deviations above the typical school in their impact on high school completion.

(remarkably, the correlation is less than 0.01).

Formal labor market participation: The final outcome we examine is participating in the formal labor market. That is 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 ITT school effect on this outcome is 0.013 using the maximum likelihood approach and 0.012 using the covariance approach. The covariance-based TOT impacts suggest that going from a school at the median of the impact distribution to one at the 85th percentile would increase the likelihood of being formally employed by about 1.8 percentage points. This effect is modest, but non-trivial. Whether an effect of this magnitude is large enough for parents to be aware of, and respond to, is an empirical question that we tackle in Section V.

IV.1 Validating school impacts using exogenous 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.[18] 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 simultaneously (1) showing out-of-sample school effects, and (2) exploring if our estimates are consistent with what one would obtain using quasi-random variation only.

Note that for the remainder of this paper we will use our estimated school impacts as predictors of student outcomes or student choices. To avoid mechanical endogeneity associated with using the outcomes of the same set students to form both the left and right-hand side variables, we create out-of-sample (or leave-year-out) estimated impacts for each school. That is, when analyzing the outcomes for students in cohort t, we use an out-of-sample prediction $\hat{\theta}^{TOT}_{j,t'}$, which is simply the TOT effect as in equation (2) but excluding all data from cohort t.

Under the algorithm used to create the tentative school assignments (discussed in Section III), 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

[18] In related work Abdulkadiroglu et al. (2017) 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.

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.[19]

Figure 1 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, one can clearly see a sharp discontinuity in preferred school attendance as a result of scoring above the school assignment cutoff. As we show in Appendix A, conditional on smooth functions of incoming SEA scores, scoring above the school assignment cutoff provides a credible source of exogenous variation to identify the causal effect of attending a preferred school. That is, scoring above the cutoff is not associated with an jump in density, or change in predicted outcomes, but *is* strongly associated with improvements in actual outcomes.

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.[20] 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.

$$Y_{i\tau} = Above_{ij} \cdot \gamma_j + f(SEA_i) + \mathbf{X_i'}\delta + \varepsilon_{ij} \tag{5}$$

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,

$$E[\gamma_j|X_i, SEA_i] = E(Y_{i\tau}|Above = 1) - E(Y_{iq}|Above = 0) \tag{6}$$

To simplify equation (6), 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

[19]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.

[20]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.

between observationally equivalent individuals in school τ 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.[21] As such, for compliers, $E[\gamma_j|X_i, SEA_i] = \theta_j^{TOT} - E[\theta_q^{TOT}]$, 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 $E[\gamma_j|X_i, SEA_i] = 0$. It follows that for the average applicant to school j, equation (6) can be written as equation (7) below.

$$E[\gamma_j|X_i, SEA_i] = \bar{p}_j \times (\theta_j^{TOT} - E[\theta_q^{TOT}]) \tag{7}$$

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 out-of-sample TOT impact of the school they attended, $\hat{\theta}_{\tau,t'}^{TOT}$ as below.

$$\hat{\theta}_{\tau,t'}^{TOT} = Above_{ij} \cdot \zeta_j + f(SEA_i) + \mathbf{X_i'}\delta + \varepsilon_{ij} \tag{8}$$

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 $E[\zeta_j|X_i, SEA_i] = E(\hat{\theta}_{\tau,t'}^{TOT}|Above = 1) - E(\hat{\theta}_{q,t'}^{TOT}|Above = 0)$. Using the same logic as above for compliers and non-compliers, it follows that

$$E[\zeta_j|X_i, SEA_i] = \bar{p}_j \times (\hat{\theta}_{j,t'}^{TOT} - E[\hat{\theta}_{q,t'}^{TOT}]) \tag{9}$$

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 (7) and (9) reveals that if our treatment on the treated estimated impacts for school τ and q are unbiased, then by the law of iterated expectations, $E[\zeta_j|X_i, SEA_i] = E[\gamma_j|X_i, SEA_i]$. In words, if our estimated school impacts from (2) 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 predicted 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.

[21] 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).

(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. We then regress the former coefficients on the latter while weighting each pair of coefficients by the number of observations used to estimate them. 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 2 shows scatterplots of the actual and predicted RD estimates for all of the outcomes combined. The left panel shows the actual and predicted RD estimates with all cutoffs weighted by the number of observations used to estimate them. To provide further evidence that our estimated school impacts are unbiased we examine this relationship only among those cutoffs that we know are not noisy (i.e. those that yield p-values below 0.2). In this sample, the slope is 0.89 and is statistically indistinguishable from 1. If one goes even further and focuses on those cutoffs with p-values smaller than 0.1, the slope is 1.18 and is also statistically indistinguishable from 1. In sum, the validation test indicates that our value-added estimates are unbiased on average. With this evidence, we are reasonably confident that the low correlations between the estimated impacts across outcomes are not driven by bias. Moreover, we are fairly confident that any parental preferences for estimated school value-added (or lack thereof) will not be the result of underlying biases in our estimated school impacts.

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 families' school rankings submitted during the application process to obtain information on individual's preferences for schools. We use a modified exploded multinomial logistic model (also known as rank-ordered multinomial logit) to estimate preferences for different school attributes (such as proximity, peer quality, estimated value added, etc.).

V.1 Simple model of school preferences

We derive the choice probability from the utility-maximizing behavior of households, indexed by $i \in N$. They face the decision of choosing a school among all schools in Trinidad and Tobago, where each school is indexed by $j \in J$. Suppose that the utility a household i derives from each school alternative j has the following general form:

$$U_{ij} = U(X_i, Z_j, \varepsilon_{ij}) = \delta(X_i, Z_j) + \varepsilon_{ij} \tag{10}$$

where $U(\cdot)$ is the function mapping school attributes and individual characteristics to utility values U_{ij}, X_i are observed household characteristics, Z_j are observed school-specific attributes, and ε_{ij} are unobserved household and school characteristics. The representative utility $\delta_{ij} = \delta(X_i, Z_j)$ is a function of the observed household characteristics and school attributes, and ε_{ij} encompasses all other unobserved (by the researcher) factors influencing schooling decisions.

We assume that the school choice set is the same for each individual (i.e., $J(i) = J \ \forall i$). Since students submit only one ranking of school preferences, there is only one choice situation per individual. Let $U_{ij}^{r_{is}}$ indicate the utility individual i gets from school j that she ranked in position s ($r_i = s$), so that $U_{ij}^{r_{i1}}$ is her utility for the school ranked first, $U_{ij}^{r_{i2}}$ is her utility for the school ranked second, and so on. Under the algorithm used to assign students to schools, among the ranked schools, students have incentives to truthfully reveal their preference rankings (Haeringer and Klijn 2009; Pathak and Sönmez 2013). Also, Chade and Smith (2006) demonstrate that the a students' optimal top choice school should be their preferred school among all possible choices. As such, if students make rational choices then:

1. $U_{ij}^{r_{i1}} > U_{ik} \ \forall k \neq j \in J$: Individual i prefers her first-ranked school over any other schools.
2. $U_{ij}^{r_{ia}} > U_{ik}^{r_{ib}} \ \forall k \neq j \in J$, $a < b$ and $b \neq \emptyset$: Individual i prefers her a-ranked school over any other school k ranked below.

Where R_i is the maximum number of alternatives ranked by individual i, assuming rational choices, the probability that an individual i submits a particular ranking on over all schools is

$$Pr\left[(U_{ij}^{r_{i1}} > U_{ik} \forall k \neq j \in J) \cap (U_{ij}^{r_{i2}} > U_{ik}^{r_{im}}, 2 < m, \forall m \in \{3, \ldots, R_i\}) \cap \ldots \cap (U_{ij}^{r_{iR_i-1}} > U_{ik}^{r_{iR_i}})\right]$$

V.2 Modified exploded multinomial logistic model

We assume that the choices are rational, and we parametrize δ_{ij} as a linear-in-parameters function of the characteristics of the schools:

$$U_{ij} = \beta' Z_j + \varepsilon_{ij} \tag{11}$$

where β is a vector of deterministic components of school preferences. We assume that ε_{ij} is distributed as an independent extreme value with Gumbel distribution for computational convenience: $F(\varepsilon_{ij} = Exp(-Exp(-\varepsilon_{ij}))$.

Following Train (2009) and Hastings et al. (2006), the probability that an individual i submits a particular ranking on over all schools is simply a product of standard logit formulas.[22] Accordingly,

[22] That is, with ε_{ij} being an independent extreme value over people, individuals and alternatives, conditional on β,

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

$$\log L(\beta) = \sum_{i=1}^{N} \log l_i(\beta) = \sum_{i=1}^{N} \log \left(Prob[r_{i1}, r_{i2}, \ldots, R_i] \right). \tag{12}$$

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). Note that this expression is equivalent to a generalization of a conventional multinomial logit model, where we allow for repeated choices for each household.

Our model is conceptually very similar to other papers in the literature. However, it includes an additional choice to the conventional exploded logit model formulation: In our first pseudo-observation, the individual chooses her first-ranked school over the set of all schools in Trinidad and Tobago. The rest of 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. This addition is empirically important: Suppose that parents really value having a short distance to school, and choose only the schools that are closest to their homes. Within each individual's school ranking, there will be little variation in the distance to school and, hence, the conventional exploded logit model would greatly understate the importance of this attribute. If proximity were correlated with other school attributes, it could lead to biased estimates for all the parameters in the choice model. Including the additional pseudo-observation, which specifies parents' first-ranked school choice over all schools in the sample, allows us to anchor each choice set (i.e., individual school ranking) to a common metric for all parents in the sample, making the choices and preferences comparable across individuals.[23]

V.3 Preference parameter estimates

In this section we examine whether parents value school impacts (estimated out of sample) on academic and non-academic dimensions, above and beyond easily observed school attributes. The estimates from the modified exploded multinomial logistic model are presented in Table 5. We compute three different specifications to investigate the robustness of the importance of each

the probability that the individual chooses the ranking $\{r_{i1}, r_{i2}, \ldots, R_i\}$ is:

$$Prob[r_{i1}, r_{i2}, \ldots, R_i] = \frac{\exp(\delta_{ij}^{r_{i1}})}{\sum_{k=1}^{J} \exp(\delta_{ik})} \cdot \frac{\exp(\delta_{ij}^{r_{i2}})}{\sum_{k=2}^{R_i} \exp(\delta_{ik}^{r_{ik}})} \cdots \frac{\exp(\delta_{ij}^{r_{iR-1}})}{\exp(\delta_{ij}^{r_{iR-1}}) + \exp(\delta_{ik}^{r_{iR_i}})}$$

[23] 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. Abdulkadiroglu et al. (2017) 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.

school attribute as more dimensions are added into the model. All specifications include control variables for whether the secondary school is in the same island, whether it is an all-girls school interacted with student gender and whether it is an all-boys school interacted with student gender. The full estimation sample includes 320,322 households making school choice decisions, 156,329 for male students, and 163,993 for female students. 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, and the stability across specifications. With the exception of log distance to school, all attributes have been standardized to be mean zero and unit variance.

The baseline specification (column 1) in Table 5 includes the log of the distance to the school, and the potential peers' academic quality (the average SEA score of the incoming cohort). Both attributes are easily observed by parents and commonly used in school choice models (e.g., Hastings et al. 2005; Hastings et al. 2006; Hastings and Weinstein 2008).[24] Perhaps not surprisingly, the most important attribute when choosing a potential school is the average peer academic quality, as reflected by the relatively large magnitude and high significance of the coefficient. Parents also prefer schools that are closer to their primary school, and the magnitude of this coefficient is stable across models, suggesting that the distance to school is a relevant feature and relatively independent to other school attributes. The importance of these variables in shaping the schooling decision is consistent with previous literature on the determinants of school choice (Hastings et al. 2005; Abdulkadiroglu et al. 2017). A comparison of the coefficients on peer quality and proximity implies that increasing peer quality by 0.4 standard deviations (about the difference between a student's top choice school and the second 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 twice as far to attend a secondary school with 0.4 standard deviations higher incoming peer scores. To better understand parental preferences for peer quality, we examine the coefficient on peer quality as one adds other attributes to the model. Including schools' causal impacts on outcomes (column 2) reduces the coefficient on peer quality by about 14 percent - suggesting that the strong preferences for incoming peer quality was correlated with, but not largely driven by preferences for higher value-added schools. However, in models that also include average school-level outcomes (column 5), the coefficient on peer quality falls by almost 40 percent. This suggests that peer quality may be highly correlated with other average school characteristics that are also valued by parents. An implication of this is that without explicitly including these other school attributes into the model, one may overestimate the importance of average peer quality on the schooling decision.

We examine parental preferences for some of these other school attributes below. Because

[24] We consider the average SEA score of the incoming cohort "observable" since the school average SEA score is made public and appears in the newspapers.

parental preferences for schools that may improve high-stakes exam scores have been the focus of existing research (e.g. Hastings et al. 2009; Abdulkadiroglu et al. 2017; and others), we start by discussing parental preferences for school impacts on high-stakes exams and then discuss preferences for schools that may improve other outcomes.

V.3.1 High-stakes exams

The second specification (columns 2 to 4) is the Value-Added-only model, which adds the variables measuring the school's causal impacts on academic outcomes (high-stakes exam index and low-stakes exam index), on non-dropout by age 14, and on non-academic longer-run outcomes (arrests index, non motherhood by age 19, and formal labor market participation) to the baseline specification. Interestingly, conditional on proximity and peer quality, parents do value schools with larger casual impacts. We discuss this in turn.

The coefficient on the high-stakes index value-added in column 2 shows that parents prefer schools that *causally* increase students' performance on high-stakes exams. Comparing the coefficients on proximity and high-stakes value added, the point estimates in column 2 suggest that (conditional on average peer scores) parents would be willing to travel about 50 percent farther to attend a secondary school with 1 standard deviation greater high-stakes value-added. Put differently, (conditional on average peer scores) parents would be willing to travel about twice as far to attend a secondary school at the 85th percentile of the high-stakes test scores effectiveness distribution than one at the 15th percentile. While this average impact is nontrivial, there is considerable heterogeneity in this preference that we document below.

To investigate whether parents can disentangle school's causal impacts on outcomes from school-level averages, and whether they only value the overall average, the value-added, or both, we construct the full model by adding the school attribute averages (also created out of sample) to the second specification (columns 5 to 7).[25] Also, because average effects may mask considerable heterogeneity, we estimate the VA-only model and the full model separately for each (SEA score ventile)×(gender) cell to examine whether preferences vary based on the student's gender and the incoming ability, measured by the individual SEA entrance exam score. Figures 3 through 8 plot the estimated coefficients from these regressions by sub-group separately for each outcome.

Parents of high-achieving students have stronger preferences for schools with larger causal impacts on the high-stakes exam performance than lower-achieving students. This is illustrated by the positive and significant relationship between the individual's score percentile and the coefficient magnitude in the left panel of Figure 3. The figure reveals that parents of children with low incoming scores *do not* prefer schools that raise high-stake tests. Indeed, among those in the bottom half

[25] MacLeod and Urquiola (2018) present a model in which it may be rational for parents to prefer schools with high absolute achievement rather than high value added.

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 contrast, parents of high-achieving children value school effectiveness a lot. For those in the top decile of incoming test scores, parents (conditional on average peer scores) would be willing to travel more than four times as far to attend a secondary school at the 85th percentile of the high-stakes test scores effectiveness distribution than one at the 15th percentile. The median distance to a school in the choice set is about 6 kilometers, so that the typical parent of a very high-achieving child would be willing to travel about 20 kilometers to attend a school at the 85th percentile than one at the 15th percentile. This is about half the width of the mainland Trinidad.

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 falls by about 20 percent, but remains large and statistical significant (Figure 3, middle), especially for females and the upper half of the individual score percentile distribution. While parents do prefer schools with higher-achieving peers and they do value schools with better average outcomes (MacLeod and Urquiola, 2018), we find that parents of high-achieving children prefer schools that raise high-stakes tests even conditional on average incoming test scores and average high-stakes outcomes at the school. This key result stands in contrast to Abdulkadiroglu et al. (2017) who find that conditional on average outcomes parents do not value school value-added. We discuss possible reasons for these differences in Section V.4. Parents of high-achieving students also value the average high-stakes performance of the school (Figure 3, right), suggesting that they independently care about both school attributes: They prefer schools that have better average performance in high-stakes exams (keeping the school's causal impact on the exam's performance constant) and, between schools that have the same high-stakes exam performance, parents will tend to choose schools that improve students' performance on the high-stakes exam. Importantly, this is all driven by parents of children in the top half of the incoming achievement distribution. Parents of less-able students do not place particular importance on outcomes related to the school's high-stakes performance, and the right panel of Figure 3 suggests that they may avoid schools with high average high-stakes exams performance, perhaps favoring schools that focus on other dimensions. In sum, although parents may not have explicit information through reports or news outlets on the schools' casual impacts on high-stakes exams, they appear to have a sense on the school's average performance *and* on the school's ability to improve student's achievement, and they take each of these attributes into account when making the schooling decision.

V.3.2 Other outcomes

In contrast to high-stakes exams, the school's causal impact on the low-stakes exam does not influence the schooling decision of males and weakly influences female decisions.[26] The value-added coefficients for females are significant but about a fourth of the size of the high-stakes coefficient (Table 5, columns 4 and 7). Looking at average outcomes, parents of males and of low-achieving females prefer schools with better average performance on the low-stakes exam (Figure 4, right). In sum, parents prefer schools with better low-stakes exam outcomes, but only parents of girls weakly prefer schools that are most effective at increasing low-stakes exam scores. In Trinidad and Tobago, average school outcomes on high-stakes exams are made public, while average school outcomes on low-stakes exams are not. The results are consistent with parents of boys being unable to discern school impacts on low-stakes exams. They are, of course, also consistent with such parents not caring about school impacts on low-stakes tests because they are low stakes.

Regarding dropout, the insignificant coefficients associated with the school's ability to improve dropout (Figure 5, left and middle panels) suggest that the average parent does not value differences in school's value-added on dropout. However, this masks some heterogeneity by gender and incoming ability. Specifically, parents of high-achieving girls in the top ventile of the incoming ability distribution have strong preferences for schools that reduce dropout. The point estimates indicate that parents of these high-achieving girls would be willing to travel 4 times as far to send their child to a school that was at the 85th percentile of the dropout value-added distribution versus one at the 15th percentile. Looking at average dropout rates, the patterns suggests that low-achieving male and female students prefer schools with low average dropout rates, as shown by the positive and significant estimated coefficients at the bottom of the ability distribution in the right panel. However, the coefficient on the (non) dropout rate is negative for parents of high-achieving children. Taken at face value, this would suggest that such parents prefer schools with higher dropout rates. While this result may seem counterintuitive, it can be explained by the fact that many of the schools that are best at improving high-stakes exams may be those that are least concerned with low-achievers on the margin of dropping out.

Next we examine whether parents take into account school impacts on non-academic outcomes when making schooling decisions. Between two schools with the same average academic performance and value-added on academic performance, parents prefer schools with lower average teen pregnancy rates and lower average juvenile arrests. The estimated coefficients on both average outcomes are statistically significant (Table 5, column 5), and their magnitudes are similar to the

[26] Note that 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.

average low-stakes exam performance coefficient, indicating that parents place similar value on the school's average academic performance and the school's average prevalence of high-risk adolescent behavior. 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 in column 5 of Table 5, the average parent would be willing to increase their distance by about 35 percent to send their child to a school that was at the 85th percentile of the crime index distribution versus one at the 15th percentile. Similarly, the average parent would be willing to increase their distance by about 30 percent to send their child to a school that was at the 85th percentile of the (non) teen motherhood distribution versus one at the 15th percentile. However, as one might expect, on average parents of females are most sensitive to pregnancy rates than those of males (Table 5, column 7), with particular intensity among high-achieving girls (Figure 6, right). Parents of females in the top decile of the achievement distribution would be willing to more than triple their distance to send their child to a school that was at the 85th percentile of the (non) teen motherhood distribution versus one at the 15th percentile.

The results show that parents clearly prefer schools with a lower prevalence of risky behaviors (conditional on peer achievement and average test score outcomes). We now examine if parents value schools that actually reduce these behaviors. Overall, neither parents of boys nor parents of girls prefer schools that reduce teen motherhood (Figure 6, left and middle). Given that this was one of the least persistent school impacts over time, we suspect that school impacts on teen motherhood may be simply difficult for parents to observe.[27] Unlike teen motherhood, we do find that parents value schools that reduce crime. Specifically, parents of high-achieving males have very strong preferences for schools that are effective in reducing the likelihood of juvenile arrests (Figure 7, left and middle). The responsiveness of males to crime value-added is very strong. Parents of males in the top decile of the achievement distribution would be willing to increase their travel distance by a factor of 10 to send their child to a school that was at the 85th percentile of the crime value-added distribution versus one at the 15th percentile. In fact, parents of high-achieving males are even more responsive to crime value added that they are to incoming peer quality –suggesting that crime and safety are very salient and important elements of the schooling decision for boys.

The last outcome we examine is formal labor market participation. All parents value schools' impacts on employment of its graduates (Table 5, column 2), but parents of high-achieving students have stronger preferences for this attribute (Figure 8, left). As we control for school averages, the effect vanishes (Table 5, column 5), but we still see significant preferences for females

[27] Parents of high-achieving girls, however, appear to place negative value on schools that reduce teen motherhood (driving the negative coefficients on teen motherhood value-added in columns (4) and (7) of Table 5). This likely follows from the fact that schools with higher averages of teen motherhood are also likely the most effective in reducing teen motherhood. Therefore, preferences for schools with lower averages teen motherhood rates among parents of high-achieving girls may result in an avoidance of schools that are effective at reducing teen motherhood.

(males) at the top half (decile) of the ability distribution (Figure 8, middle). This shift is driven by the fact that overall parents value schools with higher employment rates among their graduates (Table 5, columns 5 to 7). The pattern of results suggest that parents of low-achieving children value schools with high average formal employment rates (Figure 8, right), while parents of high-achieving children value schools that increase formal labor market participation (Figure 8, left and middle). One interpretation is that parents of higher achieving children are more likely to be more highly educated, and therefore better able to distinguish between school averages and causal impacts. Importantly, among this population (who is most likely to be sophisticated), parents value schools' causal impacts on formal labor market participation (conditional on the impact on high-stakes tests). Specifically, parents of children in the top decile of the achievement distribution would be willing to almost double their travel distance to send their child to a school that was at the 85th percentile of the formal labor-market participation value-added distribution versus one at the 15th percentile. This is a meaningful effect.

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. One consistent pattern is that parents of high-achieving children are the most responsive to schools causal impacts, while parent of low-achieving children are relatively unresponsive to causal impacts (but are responsive to average outcomes). It is important to note that valuing schools' average outcomes is not necessarily irrational, because schools with higher average outcomes may confer benefits to students that are not measured in our data.[28] The results also suggest that parents appear to know what school attributes may affect their own children the most, and choose schools accordingly: parents of male students seem to favor schools with a good record of lowering the arrests probabilities among their students, whereas parents of females prefer schools with lower teenage pregnancy rates. Overall, parents appear to be relatively sophisticated in their understanding of school quality.

V.4 Discussion of parental preference results

One of our key findings is that parents value school effectiveness above and beyond peer quality and average outcomes. However, this has not been found in all settings. We argue that a key difference between our work and others is that our estimated school impacts are not biased while others may be. To demonstrate the importance of this using our data, we introduce small biases of known form and magnitude to our school impacts and then observe how this changes results in

[28] 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.

a discrete-choice model. This is detailed in Appendix C. We were able to introduce bias to our value-added estimates that were either positively or negatively correlated with average achievement. In all cases, the biased school impacts and the real school impacts had correlation greater than 0.9. Using the unbiased estimates, parents value school effectiveness both conditional on average outcomes and unconditionally. However, in all scenarios with biased estimates, parents appear to value school effectiveness when average outcomes are not controlled for and do not appear to value school effectiveness when average outcomes are controlled for. This underscores the difficulty of disentangling parental preferences for school effectiveness from that of peer quality (or average school outcomes), and highlights the importance of using unbiased school effect impacts when estimating parental preferences for schools.[29] The extent to which existing studies employed school value-added estimates that are biased is unknown. However, we can demonstrate that bias is a plausible explanation for the differences between what we find and what others have found.

As pointed out in Beuermann and Jackson (2018) there is a lack of robust estimated achievement effects of attending schools that parents prefer. This stands in contrast to our findings here. One possible explanation for the differences between what we found and what others have found is that school value-added may be easier to infer in the studied context. 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 relatively easy for a 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 parents value school effectiveness (conditional on average outcomes) while others do not.

The ability to infer value-added can also explain some differences across outcomes. While school rankings are produced for high-stakes exams performance, there are no such rankings for the low-stakes exams. As such, our results do not necessarily imply that parents value low-stakes exams less, but may reflect the fact that parents are unable to infer school value-added on low-stakes tests. In essence, this is true for most outcomes other than high-stakes tests. However, the fact that parents of high-achieving students do appear to value schools that raise employment and reduce arrests (even conditional on school averages) suggests that in some instances parents can discern value-added even when information is imperfect (perhaps through reputation effects).

[29]Note that Abdulkadiroglu et al. (2017) use average *predicted* outcomes as a proxy for peer quality rather than average actual outcomes as we use in our illustration. Irrespective of the measure of peer quality used, the basic result stands– even slightly biased value-added estimates that are correlated with peer quality may be difficult to disentangle from the impact of peer quality itself.

VI Conclusions and Policy Implications

We estimate schools causal impacts on an broad array of outcomes. 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).

While we find that parents prefer high value-added schools, we do find heterogeneity in this effect. Parents value effectiveness on high-stakes exams, dropout, reduced crime, and formal employment. However, in all cases, it is parents of high-achieving children who are responsive to schools' causal impacts and not parents of low-achieving children. 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 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). 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.

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 (%)	51.20			54.20	51.37
	(49.99)			(49.82)	(49.98)
Admitted cohort size	243.03	248.78	237.89	136.61	346.26
	(188.73)	(188.50)	(188.79)	(93.96)	(199.90)
Standardized SEA score	0.00	-0.12	0.12	0.85	-0.41
	(1.00)	(1.04)	(0.94)	(0.56)	(0.77)
Individuals	320,322	156,329	163,993	130,820	134,880
Panel B: NCSE data (linked to SEA cohorts: 2006 - 2012)					
Took NCSE (%)	87.50	85.83	88.98	93.28	82.21
	(33.08)	(34.87)	(31.32)	(25.03)	(38.25)
Standardized NCSE score	0.00	-0.21	0.18	0.37	-0.47
	(1.00)	(1.00)	(0.97)	(0.87)	(0.94)
Individuals	102,224	48,096	54,128	55,154	33,151
Panel C: CSEC data (linked to SEA cohorts: 1995 - 2011)					
Took at least 1 subject (%)	75.70	69.72	81.44	90.39	71.96
	(42.89)	(45.95)	(38.88)	(29.47)	(44.92)
Number of subjects passed	3.15	2.53	3.74	5.27	2.00
	(3.10)	(2.96)	(3.12)	(2.85)	(2.46)
Qualified for tertiary (%) *	34.38	27.48	41.01	65.24	16.27
	(47.50)	(44.64)	(49.19)	(47.62)	(36.90)
Individuals	306,183	150,026	156,157	123,151	131,349
Panel D: CAPE data (linked to SEA cohorts: 1999 - 2009)					
Took at least 1 unit (%)	19.79	15.36	24.10	38.80	6.08
	(39.84)	(36.05)	(42.77)	(48.73)	(23.90)
Number of units passed	1.40	1.06	1.72	2.80	0.38
	(2.94)	(2.62)	(3.18)	(3.68)	(1.58)
Earned Associate's Degree (%)	14.80	10.95	18.55	30.31	3.42
	(35.51)	(31.23)	(38.87)	(45.96)	(18.17)
Earned scholarship (%)	0.97	0.68	1.25	2.20	0.02
	(9.78)	(8.20)	(11.10)	(14.68)	(1.29)
Individuals	203,472	100,396	103,076	88,519	90,146
Panel E: Criminal records (linked to SEA cohorts: 1995 - 2010) - in percent					
Arrested by 18	3.38	6.02	0.82	1.46	4.31
	(18.07)	(23.79)	(9.00)	(11.98)	(20.30)
Individuals	292,195	143,736	148,459	115,383	127,893
Panel F: Birth records (linked to SEA cohorts: 2004 - 2010) - in percent					
Live birth by 19			10.02	5.94	14.00
			(30.02)	(23.63)	(34.70)
Individuals			48,091	25,890	17,775
Panel G: Labor market data (linked to SEA cohorts: 1995 - 2006) - in percent					
Formally employed	75.73	78.12	73.39	77.85	75.90
	42.87	41.35	44.19	41.53	42.77
Individuals	234,835	116,432	118,403	83,722	109,089

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 (i.e. students who did not score high enough to gain an school assignment within their choice sets are not included).

Table 2: 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

Crime Index	Weight
Never arrested	0.223
Not arrested by 17	0.248
Not arrested by 18	0.268
Not arrested by 19	0.267
Number of times arrested ever	-0.191

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 3: Correlation Between School Impacts

	High-Stakes Index	No Dropout by 14	Low-Stakes Index	Formally Employed	No live birth by 19	Crime Index
High-Stakes Index	1.00					
No Dropout by 14	0.15	1.00				
Low-Stakes Index	0.16	-0.02	1.00			
Formally employed	-0.08	0.04	0.02	1.00		
No live birth by 19	0.06	0.05	-0.04	0.11	1.00	
Crime Index	0.00	0.01	0.01	-0.04	-0.07	1.00

Notes: This table reports correlation coefficients of the estimated school impacts, $\hat{\theta}_j^{TOT}$, across the different dimensions measured.

Table 4: Standard Deviation of School Impacts

	Maximum Likelihood approach			Covariance approach	
Outcomes	SD school	[95% CI]		ITT	TOT
Low-Stakes Index	0.149	0.125	0.177	0.144	0.298
High-Stakes Index	0.123	0.106	0.141	0.134	0.270
No Dropout by 14	0.087	0.067	0.112	0.080	0.186
No live birth by 19	0.043	0.013	0.138	0.054	0.047
Crime Index	0.045	0.035	0.058	0.039	0.078
Formally employed	0.031	0.020	0.048	0.027	0.043
Binary outcomes					
No Dropout by 14	0.030	0.023	0.039	0.028	0.065
No live birth by 19	0.030	0.023	0.039	0.016	0.014
Not arrested by 18	0.006	0.005	0.008	0.005	0.013
Formally employed	0.013	0.009	0.021	0.012	0.018

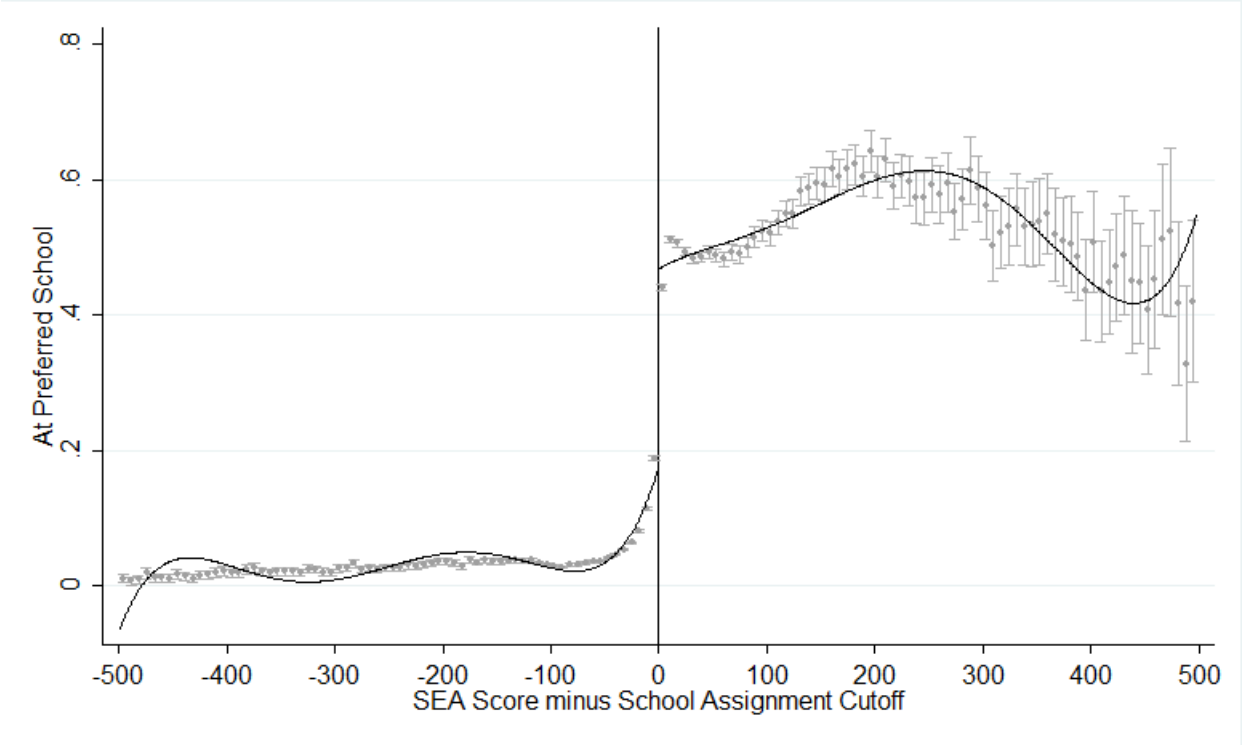
Notes: The left panel of the table reports estimates of the standard deviation of the assigned school impacts ($\hat{\sigma}_{\theta_j^{ITT}}$) along with their respective 95 percent confidence intervals obtained from the maximum likelihood approach. The right panel reports the implied standard deviation of the assigned school impacts ($\hat{\sigma}_{\theta_j^{ITT}}$) and the implied standard deviation of the attended school impacts ($\hat{\sigma}_{\theta_j^{TOT}}$) obtained from the covariance approach.

Table 5: Choice Model Estimates

		VA only model			Full model		
	Observables	All	Males	Females	All	Males	Females
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log distance to school	-0.561***	-0.561***	-0.608***	-0.529***	-0.563***	-0.611***	-0.529***
	(0.0130)	(0.0122)	(0.0128)	(0.0167)	(0.0118)	(0.0126)	(0.0161)
Average peer quality	1.424***	1.234***	1.126***	1.268***	0.764***	0.691***	0.830***
	(0.0453)	(0.0349)	(0.0488)	(0.0416)	(0.0323)	(0.0440)	(0.0418)
VA high-stakes index		0.285***	0.247***	0.420***	0.153***	0.112***	0.299***
		(0.0255)	(0.0320)	(0.0454)	(0.0279)	(0.0370)	(0.0383)
VA low-stakes index		0.0606***	-0.0045	0.0946***	0.0343	-0.266	0.0819**
		(0.0146)	(0.0209)	(0.0169)	(0.0219)	(0.0245)	(0.0324)
VA no dropout by 14		0.0523	-0.0339	0.0923*	0.0419	-0.0618*	0.115**
		(0.0355)	(0.0399)	(0.0542)	(0.0341)	(0.0354)	(0.0534)
VA crime index		-0.109	0.439***	-0.562***	-0.0258	0.467***	-0.546***
		(0.104)	(0.136)	(0.127)	(0.101)	(0.130)	(0.129)
VA no live birth by 19		-0.112***	-0.0787***	-0.129***	-0.0518*	0.0065	-0.0748*
		(0.0263)	(0.0303)	(0.0428)	(0.0289)	(0.0313)	(0.0443)
VA formally employed		0.112***	0.0093	0.182***	-0.0356	-0.143***	0.0444
		(0.0248)	(0.0334)	(0.0333)	(0.0255)	(0.0335)	(0.0348)
Average high-stakes index					0.248***	0.230***	0.200***
					(0.0215)	(0.0389)	(0.0200)
Average low-stakes index					0.116***	0.153***	0.0504
					(0.0256)	(0.0239)	(0.0408)
Average no dropout by 14					-0.0646***	-0.0479	-0.0398
					(0.0218)	(0.0303)	(0.0296)
Average crime index					0.100***	0.118***	0.127***
					(0.0212)	(0.0270)	(0.0304)
Average no live birth by 19					0.0930***	-0.0083	0.143***
					(0.0297)	(0.0300)	(0.0433)
Average formally employed					0.215***	0.227***	0.231***
					(0.0283)	(0.0430)	(0.0377)
Observations	320,322	320,322	156,329	163,993	320,322	156,329	163,993

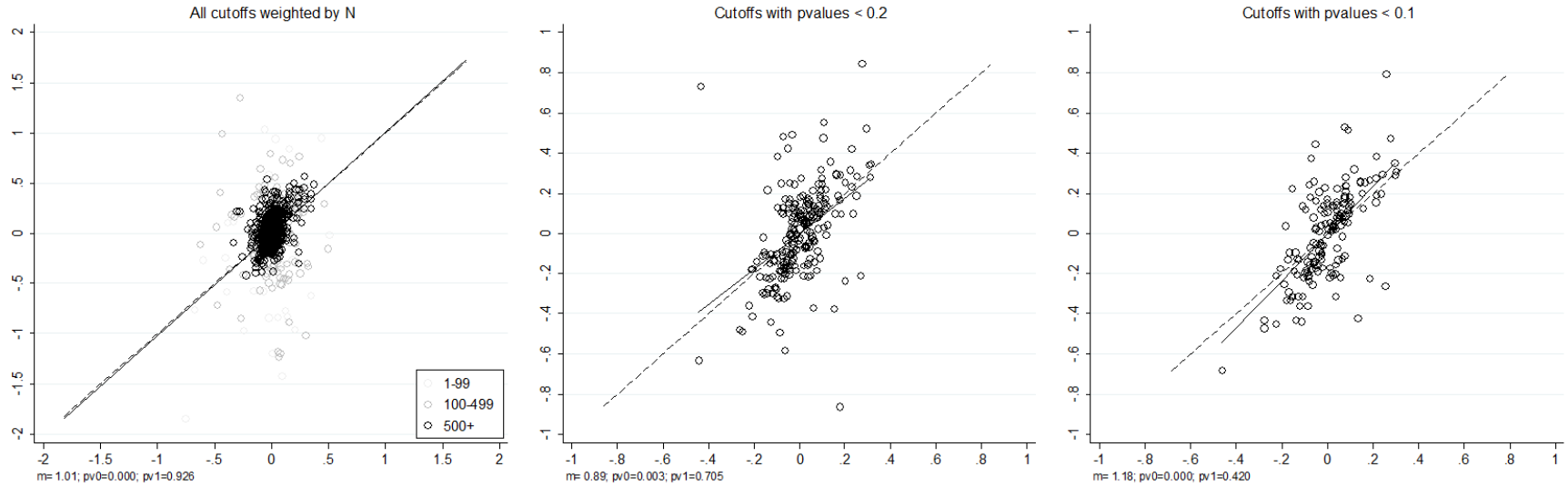
Estimates of a exploded logit model with fixed coefficients, with an additional choice of the first school compared to all possible schools in Trinidad and Tobago. Value-addeds adjusted by the compliance rates. Estimated standard errors clustered at the (SEA score ventile × gender × school district level). Model includes as control variables whether the school is in the same island, whether the school is only-males or only-females, and interaction terms between only-males and only-females binary variables and gender. *** p<0.01, ** p<0.05, * p<0.1

Figure 1. Discontinuity in Preferred School Attendance Through Assignment Cutoffs



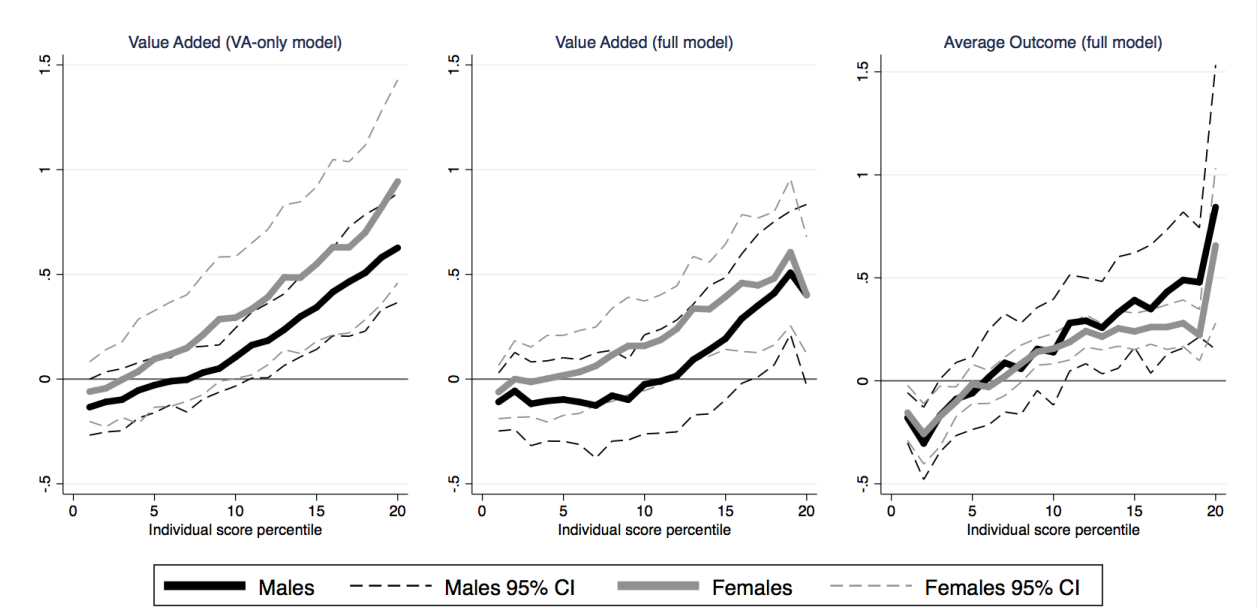
Notes: The X-axis is the SEA score relative to the assignment cutoff. The circles are means corresponding to 7-point bins of the relative score. The solid lines are the fitted school attendance rates generated by fitting a fifth degree polynomial of the relative score fully interacted with an indicator for scoring above the school assignment cutoff. The gray vertical bars depict the 90 percent confidence intervals for each bin average.

Figure 2. Predicted Cutoff Effects versus Actual Cutoff Effects



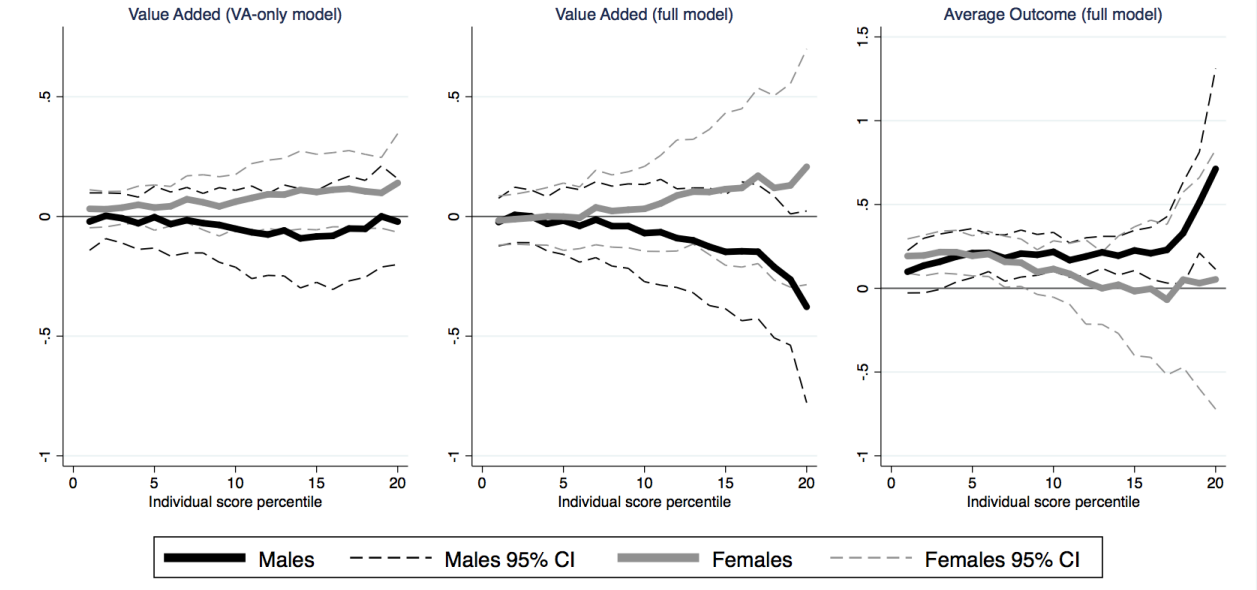
Notes: The X-axis represents the estimated coefficients on the 'Above' indicator resulting from model (8); estimated for each school j and for each outcome (estimated school impacts enter as dependent variables). The Y-axis represents the estimated coefficients on the 'Above' indicator resulting from model (5); estimated for each school j and for each outcome (individual level outcomes enter as dependent variables). The connected lines represent lineal fits; while the 45° lines are dashed in each panel. Estimated slopes and p-values resulting from testing for whether these slopes differ from both 0 and 1 are shown below each panel. The left panel includes all estimated effects weighted by the number of observations in each estimation. The center and right panels include only cutoffs that produced estimated coefficients with p-values below 0.2 and 0.1 respectively. Cutoffs with higher number of observations are shown in darker tones. In the middle panel, one outlier was removed (from the crime dimension) as it was located more than 4 standard deviations away from the mean and was estimated with less than 30 observations.

Figure 3: High-Stakes Index



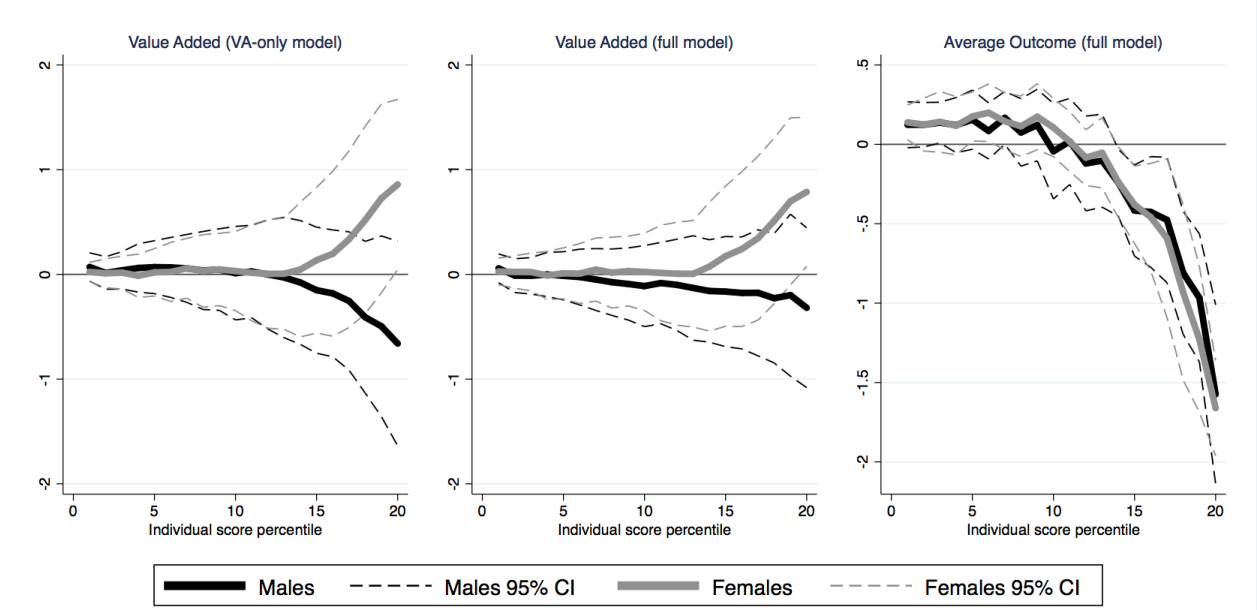
Notes: The connected lines represent the estimated coefficients for the outcome, computed separately for each (SEA score ventile)×(gender) cell for two different models: VA-only model (left panel), and full model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure 4: Low-Stakes Index



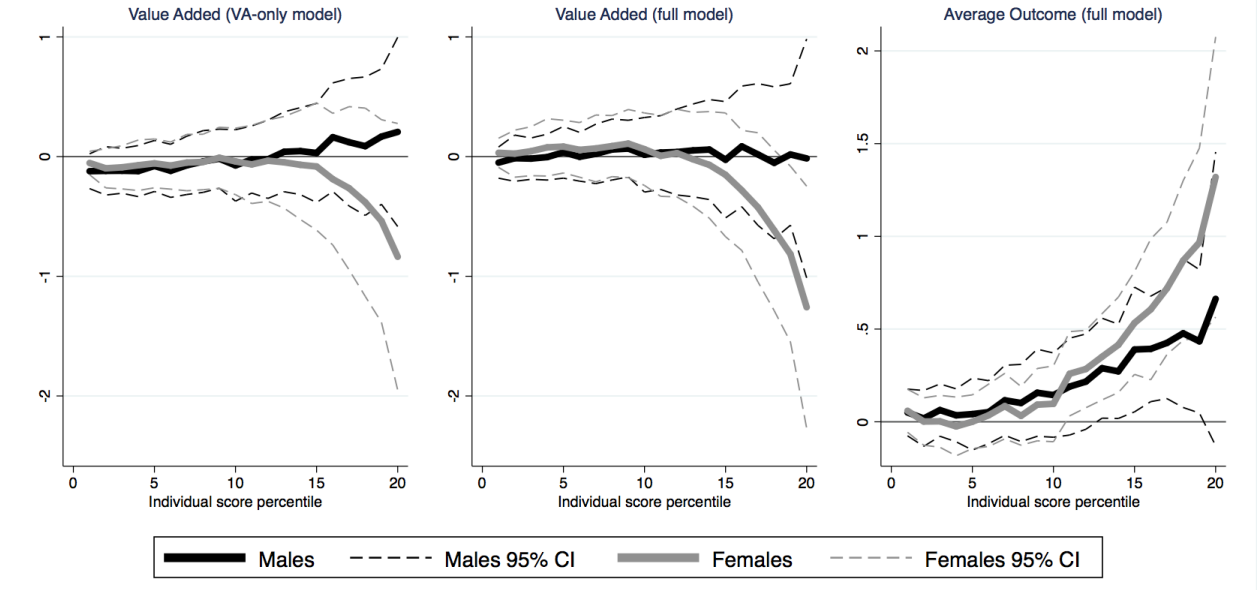
Notes: The connected lines represent the estimated coefficients for the outcome, computed separately for each (SEA score ventile)×(gender) cell for two different models: VA-only model (left panel), and full model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure 5: No Dropout by Age 14



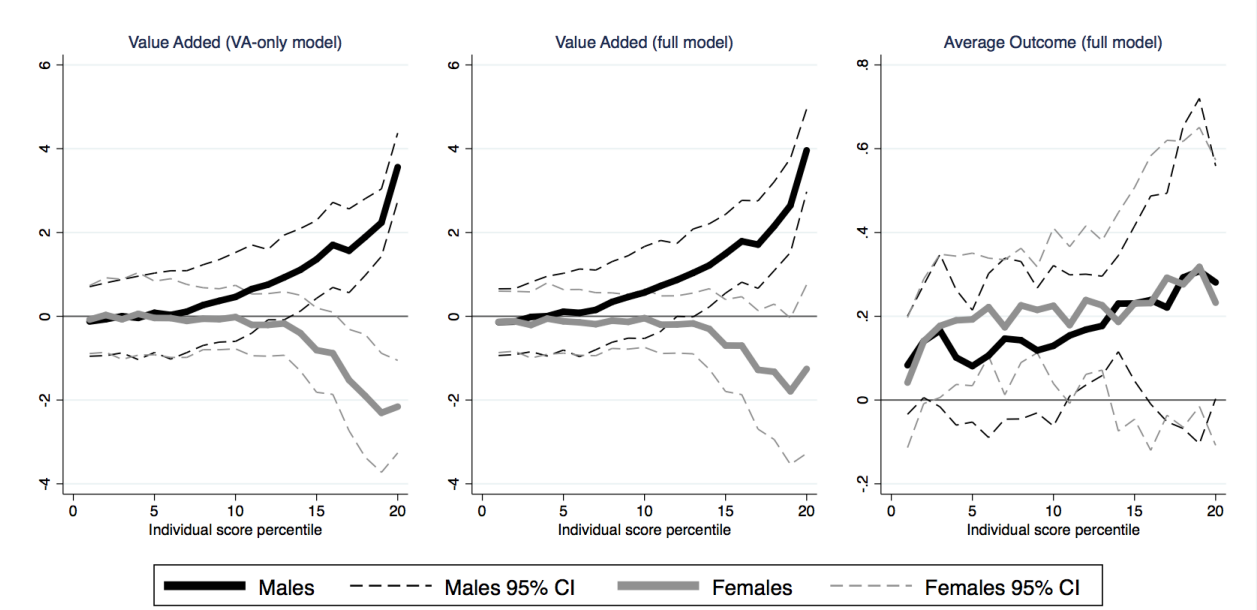
Notes: The connected lines represent the estimated coefficients for the outcome, computed separately for each (SEA score ventile)×(gender) cell for two different models: VA-only model (left panel), and full model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure 6: No Live Birth by Age 19



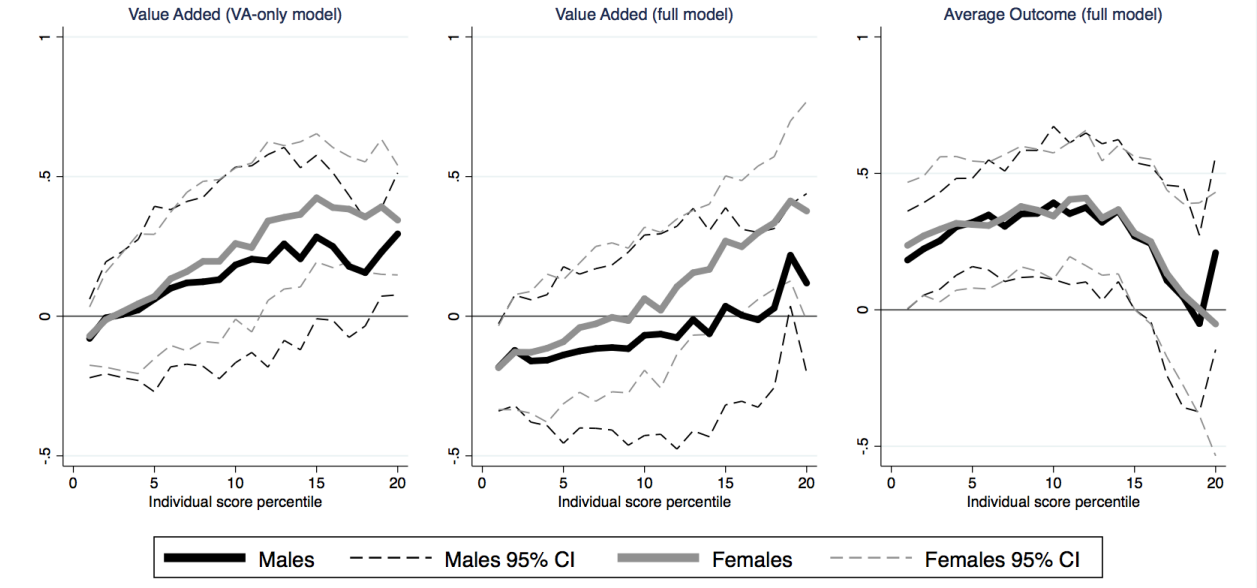
Notes: The connected lines represent the estimated coefficients for the outcome, computed separately for each (SEA score ventile)×(gender) cell for two different models: VA-only model (left panel), and full model (middle and right panels). The dashed lines represent the associated 95% confidence intervals.

Figure 7: Crime 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: Formal Employment



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, 1 2014.

Atila Abdulkadiroglu, Parag Pathak, Jonathan Schellenberg, and Christopher Walters. Do Parents Value School Effectiveness? Technical report, National Bureau of Economic Research, Cambridge, MA, 10 2017.

Joshua D Angrist, Parag A Pathak, and Christopher R Walters. Explaining Charter School Effectiveness. *American Economic Journal: Applied Economics*, 5(4):1–27, 10 2013.

Diether Beuermann and C. Kirabo Jackson. Do Parents Know Best? The Short and Long-Run Effects of Attending The Schools that Parents Prefer. Technical report, National Bureau of Economic Research, Cambridge, MA, 8 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, 4 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, 9 2015.

Hector Chade and Lones Smith. Simultaneous Search. *Econometrica*, 74(5):1293–1307, 9 2006.

David J. Deming. Better Schools, Less Crime? *. *The Quarterly Journal of Economics*, 126(4): 2063–2115, 11 2011.

David J. Deming. Using School Choice Lotteries to Test Measures of School Effectiveness. *American Economic Review*, 104(5):406–411, 5 2014. ISSN 0002-8282. doi: 10.1257/aer.104.5.406.

David J. Deming, Justine S. Hastings, Thomas J. Kane, and Douglas O. Staiger. School Choice, School Quality, and Postsecondary Attainment, 2014.

Will Dobbie and Roland Fryer. Charter Schools and Labor Market Outcomes. Technical report, National Bureau of Economic Research, Cambridge, MA, 8 2016.

Will Dobbie and Roland G. Fryer. The Medium-Term Impacts of High-Achieving Charter Schools. *Journal of Political Economy*, 123(5):985–1037, 10 2015.

L. E. Dubins and D. A. Freedman. Machiavelli and the Gale-Shapley Algorithm. *The American Mathematical Monthly*, 88(7):485, 8 1981. ISSN 00029890. doi: 10.2307/2321753.

Dennis Epple and Richard E. Romano. Competition between Private and Public Schools, Vouchers, and Peer-Group Effects, 1998.

Milton Friedman. The Role of Government in Education (1955), 1955.

D. Gale and LS Shapley. College Admissions and the Stabilty of Marriage. *The American Mathematical Monthly*, 69(1):9–15, 1 1962. ISSN 00029890. doi: 10.2307/2312726.

John I. Goodlad. *A place called school : prospects for the future*. McGraw-Hill Book Co, 1984. ISBN 0070236267.

Guillaume Haeringer and Flip Klijn. Constrained school choice. *Journal of Economic Theory*, 144 (5):1921–1947, 9 2009. ISSN 0022-0531. doi: 10.1016/J.JET.2009.05.002.

Eric A. Hanushek. Teacher characteristics and gains in student achievement: Estimation using micro data. *American Economic Review*, 61(2):280–288, 1971. ISSN 0002-8282. doi: 10.2307/1817003.

Hart Research Associates. Public School Parents On The Value Of Public Education. Technical report, 2017.

Justine Hastings, Thomas Kane, and Douglas Staiger. Parental Preferences and School Competition: Evidence from a Public School Choice Program. Technical report, National Bureau of Economic Research, Cambridge, MA, 11 2005.

Justine Hastings, Thomas Kane, and Douglas Staiger. Preferences and Heterogeneous Treatment Effects in a Public School Choice Lottery. 4 2006. doi: 10.3386/w12145.

Justine Hastings, Christopher Neilson, and Seth Zimmerman. The Effects of Earnings Disclosure on College Enrollment Decisions. Technical report, National Bureau of Economic Research, Cambridge, MA, 6 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, 11 2008. ISSN 0033-5533. doi: 10.1162/qjec.2008.123.4.1373.

Justine S Hastings, Thomas J Kane, and Douglas O Staiger. Heterogeneous Preferences and the Efficacy of Public School Choice. *Working Paper*, 2009.

Amanda Hay and Myra Hodgkinson. Rethinking leadership: a way forward for teaching leadership? *Leadership & Organization Development Journal*, 27(2):144–158, 2 2006. ISSN 0143-7739. doi: 10.1108/01437730610646642.

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, 7 2006. ISSN 0734-306X. doi: 10.1086/504455.

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, 12 2010. ISSN 00130133. doi: 10.1111/j.1468-0297.2010.02371.x.

C. Kirabo Jackson. Match Quality, Worker Productivity, and Worker Mobility: Direct Evidence from Teachers. *Review of Economics and Statistics*, 95(4):1096–1116, 10 2013.

C. Kirabo Jackson. What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes. *Journal of Political Economy*, 5 2018. doi: 10.3386/w22226.

Thomas Kane and Douglas Staiger. Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation. 12 2008.

Tim Kautz, James J. Heckman, Ron Diris, Bas ter Weel, and Lex Borghans. Fostering and Measuring Skills. *National Bureau of Economic Research,*, No. w19656, 11 2014.

Adrienne M. Lucas and Isaac M. Mbiti. Effects of School Quality on Student Achievement: Discontinuity Evidence from Kenya. *American Economic Journal: Applied Economics*, 6(3):234–263, 7 2014. ISSN 1945-7782. doi: 10.1257/app.6.3.234.

W. Bentley MacLeod and Miguel Urquiola. Is Education Consumption or Investment? Implications for the Effect of School Competition. Technical report, National Bureau of Economic Research, Cambridge, MA, 10 2018.

Isaac Mbiti, Karthik Muralidharan, Mauricio Romero, Youdi Schipper, Constantine Manda, and Rakesh Rajani. Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania. 7 2018. doi: 10.3386/w24876.

Justin McCrary. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2):698–714, 2 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, 2 2013. ISSN 0002-8282. doi: 10.1257/aer.103.1.80.

Cristian Pop-Eleches and Miguel Urquiola. Going to a Better School: Effects and Behavioral Responses. *American Economic Review*, 103(4):1289–1324, 6 2013.

Neil. Postman. The End of Education: Redefining the Value of School. *New York*, page 94, 1996. ISSN 03802361. doi: 10.2307/1495104.

Alvin E. Roth. The Economics of Matching: Stability and Incentives. *Mathematics of Operations Research*, 7(4):617–628, 11 1982. ISSN 0364-765X. doi: 10.1287/moor.7.4.617.

Donald B Rubin. FORMAL MODES OF STATISTICAL INFERENCE FOR CAUSAL EFFECTS. Technical report, 1990.

Tim R. Sass, Ron W. Zimmer, Brian P. Gill, and T. Kevin Booker. Charter High Schools' Effects on Long-Term Attainment and Earnings. *Journal of Policy Analysis and Management*, 35(3): 683–706, 6 2016.

Kenneth E. Train. *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge, 2009. ISBN 9780511805271. doi: 10.1017/CBO9780511805271.

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 admitted 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 admitted 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 assignment 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 assign 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.[30] Scoring above zero means scoring above the cutoff for a preferred school. Figure 1 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.[31] 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. 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.[32] 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 of the location of the cutoffs and are forced to make school choices before they take the SEA

[30] 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.

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

[32] We implement these tests using the *rddensity* command in Stata.

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 the NCSE English language score, the NCSE math score, an indicator for earning the CSEC tertiary qualification, and an indicator for earning the CAPE associate degree on the number of SEA attempts (repeater status in 5th grade), the student's sex, the student's religion, indicators for the primary school, 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 these outcomes such that, as shown in Appendix Table A2, column (1) they yield adjusted R-squares ranging from 0.17 to 0.38.

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:

$$Y^{p}_{i\tau t} = \pi \cdot Above_{ijt} + f(SEA_{it}) + C_{jt} + \varepsilon_{ijt} \tag{13}$$

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 a 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' outcomes 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 (13) 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

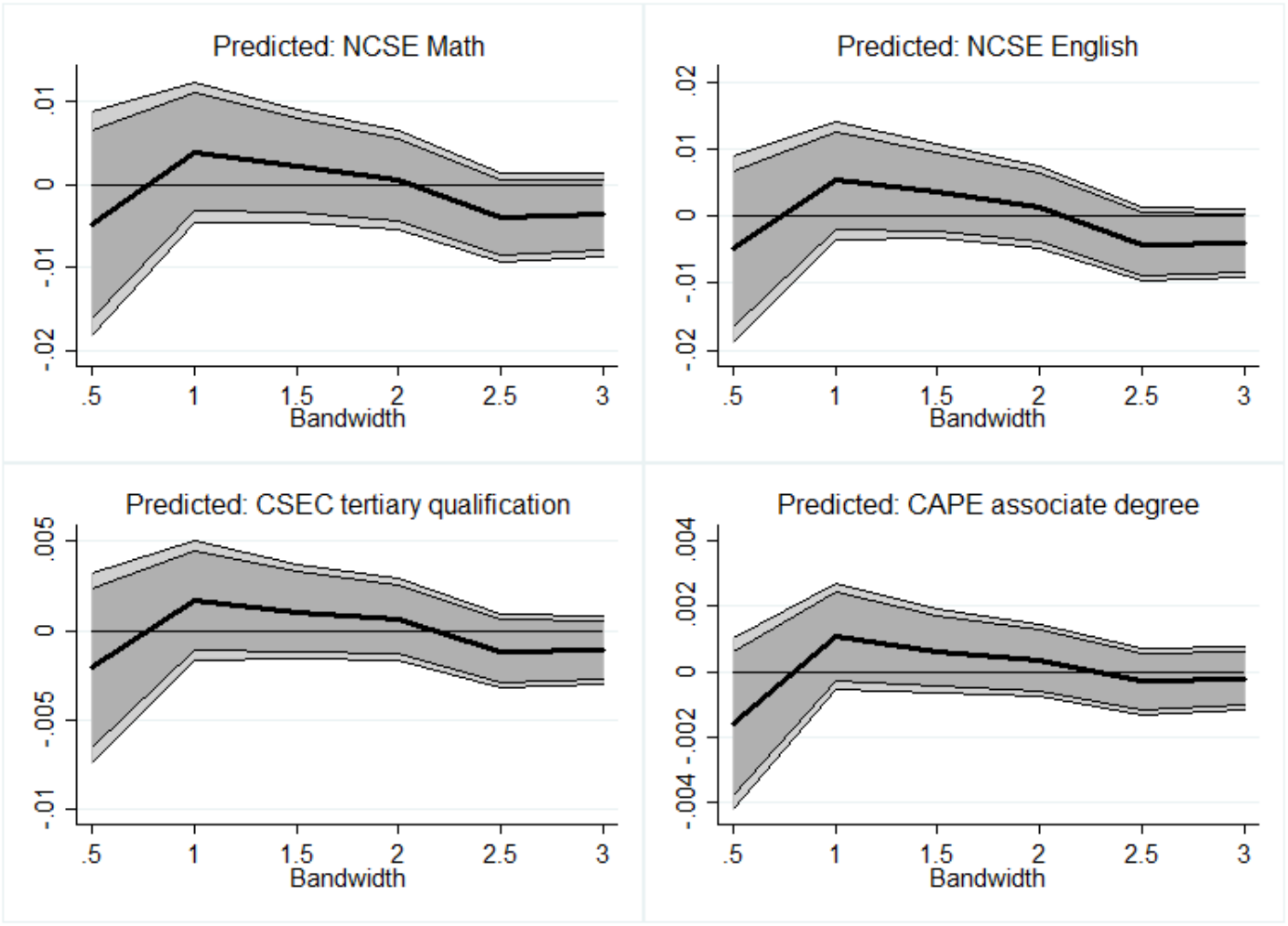
Notes: This table reports *p-values* of differential density tests across school assignment cutoffs for each SEA cohort included in the study.

Table A2. Reduced-form effects on predicted and actual outcomes

	Predicted Outcomes			Actual Outcomes			
	Prediction R2	Effect	*p-value*	Effect	*p-value*	Effect	*p-value*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
NCSE math	0.375	-0.004	0.123	0.032	0.001	0.035	<0.001
NCSE English	0.365	-0.004	0.151	0.032	0.001	0.033	<0.001
CSEC tertiary qualification	0.299	-0.001	0.270	0.009	0.003	0.009	0.002
CAPE Associate degree	0.174	<0.001	0.665	0.019	<0.001	0.019	<0.001
Cutoff fixed effects		Yes		Yes		Yes	
Sociodemographics		No		No		Yes	

Notes: This table reports estimated coefficients on 'Above' resulting from equation (13). Models were estimated using all available observations within a bandwidth of +/-3 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 (13). The estimated coefficients are reported for each bandwidth between +/-0.5sd an +/-3sd from the school assignment cutoff. The 90 (95) percent confidence intervals of the estimated coefficients are presented in dark (light) gray.

Appendix B: Exploratory Factor Analysis

Table B1. Exploratory Factor Analysis Among Estimated Impacts

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
No live birth by 19	0.22	-0.40	0.01	-0.04	-0.04	-0.02
No Dropout by 14	0.10	0.13	0.12	0.17	-0.04	-0.03
NCSE Total Academic	0.28	-0.01	0.12	0.00	0.49	-0.11
NCSE Total Non academic	-0.21	0.03	0.04	0.21	0.51	0.01
Number of CSEC subjects passed	0.63	0.05	0.64	0.19	0.05	0.02
CSEC tertiary qualification	0.49	0.08	0.70	-0.06	0.03	0.00
CSEC tertiary qualification attempt	-0.04	0.23	0.15	0.64	0.14	-0.04
CAPE scholarship	0.53	-0.05	-0.21	-0.13	0.00	0.06
CAPE scholarship attempt	0.97	-0.15	0.14	-0.06	-0.02	0.00
Number of CAPE units passed	0.97	-0.14	0.10	-0.03	-0.01	0.03
CAPE Associate degree	0.98	-0.13	0.09	-0.04	0.02	0.03
Never arrested	-0.16	0.73	0.02	-0.04	0.08	-0.44
Not arrested by 17	-0.04	0.85	0.08	0.17	-0.04	0.23
Not arrested by 18	-0.11	0.96	0.07	0.00	0.03	0.02
Not arrested by 19	-0.08	0.96	-0.02	-0.03	-0.03	-0.04
Number of times arrested ever	0.28	-0.73	0.07	0.09	-0.03	0.43
Formally employed	-0.22	-0.27	-0.08	0.59	-0.04	0.13
Variance	4.11	4.00	1.05	0.94	0.54	0.47
Proportion of variance explained	0.38	0.37	0.10	0.09	0.05	0.04

Notes: Exploratory factor analysis (using the principal factor method) was conducted among the estimated school impacts on all individual outcomes in order to determine how to group outcomes within indexes. Scores reported here are only used to identify which outcomes tend to move together and are not used to compute the indexes. The weights used to construct the multi-outcome indexes are shown in Table 2. NCSE academic subjects include mathematics, English, Spanish, sciences, and social studies. NCSE non academic subjects include arts, physical education, and technical studies. 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 degree is awarded when passing seven CAPE units (including Caribbean and Communication studies).

Appendix C: Can Bias Explain Differences with Existing Work?

To demonstrate the importance of using credibly causal school value-added in the choice models, using our data, we introduce bias of known form and magnitude to our school estimates are observe how this changes results in a discrete-choice model. Because we aim to illustrate how discrete choice models may behave with biased school impacts (as opposed to uncovering true preference parameters) we use the more commonly used rank-ordered logit model for this exercise. We will show that under plausible and realistic scenarios of bias, models that condition on peer quality can lead to wildly misleading conclusion regarding the value parents place of school effectiveness condition on peer quality.

Example Case 1: In most discussions of self-selection bias, one would expect that individual school effects are more biased among those schools that are most selective. To approximate this, we simply added a constant to the school impacts for all schools with peer quality in the top third of schools.[33] We present the estimated preference parameters with this bias introduced to the value-added in Table C1. The new biased estimate has correlation 0.97 with the true estimate. Using the real effect (columns 1 and 2), the coefficient in the model without peer quality is 0.886 (p-value=0.000), and that with peer quality controls is 0.14 (p-value$<$0.000). Using the biased effect (columns 3 and 4), the coefficient in the model without peer quality is 0.971 (p-value=0.000), and that with peer quality controls is 0.039 (p-value=0.287). In this scenario, the estimated impact is biased upward without controls for peer quality, but biased down when peer quality (in this case average outcomes) is controlled for. To be more realistic, we allow the extent of the bias to increase smoothly with selectivity in a nonlinear manner.[34] The new biased estimate has correlation 0.87 with the true estimate. Using the biased effect (columns 3 and 4), the coefficient in the model without peer quality is 0.94 (p-value=0.000), and that with peer quality controls is 0.046 (p-value=0.19).

Example Case 2: Because the model is nonlinear, biases can influence the estimates is unexpected ways. For example, one would expect that positive bias at the top of the distribution would lead to opposite impacts as positive bias at the bottom of the distribution. However, this need not be the case. To show this, we created smooth bias at bottom of distribution.[35] The new biased estimate has correlation 0.92 with the true estimate. Unlike the previous scenarios in which the bias is positively correlated with peer quality, here the bias is negatively correlated with peer quality (in this case proxied with average outcomes). However, using the biased effect (columns 7 and 8), the coefficient in the model without peer quality is 0.35 (p-value=0.000), and that with peer quality controls is 0.014 (p-value=0.6).

In essence, the exercise above demonstrates that if school impacts are biased, and the bias is correlated with peer quality, choice models that condition on peer quality are unlikely to yield the influence of school effectiveness conditional on peer quality, and may be impossible to interpret in any meaningful way. We have no way of knowing the extent to which existing studies employed school value-added estimates that are biased. However, we can demonstrate that bias is a plausible explanation for the differences between what we find and what other have found. Specifically, by introducing relatively small amounts of bias that is correlated with measures of peer quality, we

[33] We added a constant of 0.7 to schools with average standardized peer quality greater than 0.5.

[34] Here the bias is an increasing quadratic function of peer quality, but only for schools with peer quality above 0.5.

[35] Here the bias is an decreasing quadratic function of peer quality (so that is is positive when peer quality is below zero), but only for schools with peer quality below 0.

are able to replicate results akin to Abdulkadiroglu et al. (2017) where parents appear to value schools that are effectiveness on average, but do not value such schools conditional on measures of peer quality. This underscores the difficulty of disentangling parental preferences for school effectiveness from that of peer quality, and highlights the importance of using unbiased school effect impacts when estimating parental preferences for schools.

Table C1. Estimated Preferences for Value-Added With and Without Bias

	Dependent Variable: Rank							
	1	2	3	4	5	6	7	8
Average High Stakes		0.993**		1.042**		1.042**		1.064**
		(0.643)		(0.675)		(0.675)		(0.689)
Value-Added Real	0.886**	0.140**						
	(0.492)	(0.078)						
Value-Added Biased			0.972**	0.0395	0.941**	0.046	0.347**	0.0144
			(0.540)	(0.022)	(0.523)	(0.026)	(0.193)	(0.008)
corr(bias,real)			0.971		0.876		0.925	
corr(bias,Ave High-Stakes)			0.862		0.733		-0.728	
Observations				1,344,027				
Number of students				296,462				

Notes: Robust normalized beta coefficients in parentheses.
** p<0.01, * p<0.05, + p<0.1
Value-addeds adjusted by the compliance rates. Estimated standard errors clustered at the (SEA score ventile × gender × school district level). Model includes as control variables whether the school is in the same island, whether the school is only-males or only-females, and interaction terms between only-males and only-females binary variables and gender.