NBER WORKING PAPER SERIES

INFORMATION, PREFERENCES, AND HOUSEHOLD DEMAND FOR SCHOOL VALUE ADDED

Robert Ainsworth
Rajeev Dehejia
Cristian Pop-Eleches
Miguel Urquiola

Working Paper 28267
http://www.nber.org/papers/w28267

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2020

For useful comments we thank Matt Backus, Peter Bergman, Michael Best, Sarah Cohodes, David Figlio, Dong Woo Hahm, Douglas Harris, Caroline Hoxby, Kirabo Jackson, Francisco Martorell, Karthik Muralidharan, Suresh Naidu, Christopher Neilson, Daniele Paserman, Jonah Rockoff, Richard Romano, Doug Staiger, and seminar participants at the NBER Summer Institute, APPAM, Princeton, the University of Florida, European University Institute, Bocconi, Nottingham, and UC Davis. We especially thank Miikka Rokkanen for early contributions to this project. For financial support, we are grateful to the Center for Development Economics and Policy (CDEP) and the Faculty Grants Program at SIPA (Columbia), to New York University, and to the KDI School of Public Policy and Management. This RCT was registered in the American Economic Association Registry for randomized control trials under trial number AEARCTR-0004496. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Information, Preferences, and Household Demand for School Value Added
Robert Ainsworth, Rajeev Dehejia, Cristian Pop-Eleches, and Miguel Urquiola
NBER Working Paper No. 28267
December 2020
JEL No. I20

ABSTRACT

This paper examines the roles that information and preferences play in determining whether households choose schools with high value added. We study Romanian school markets using administrative data, a survey, and an experiment. The administrative data show that, on average, households could select schools with 1 s.d. worth of additional value added. This may reflect that households have incorrect beliefs about schools' value added, or that their preferences lead them to prioritize other school traits. We elicit households' beliefs and find that they explain less than a fifth of the variation in value added. We then inform randomly selected households about the value added of the schools in their towns. This improves the accuracy of households' beliefs and leads low-achieving students to attend higher-value added schools. We next estimate households' preferences and predict their choices under the counterfactual of fully accurate beliefs. We find that beliefs account for 18 (11) percent of the value added that households with low- (high-) achieving children leave unexploited. Interestingly, for households with low-achieving children, the experiment seems to have affected both beliefs and preferences. This generates larger effects on choices than would be predicted via impacts on beliefs alone.

Robert Ainsworth
Economics Department
University of Florida
321 Matherly Hall
1405 W. University Avenue
Gainesville, FL 32611
robert.ainsworth@ufl.edu

Rajeev Dehejia
Robert F. Wagner Graduate School
of Public Service
New York University
295 Lafayette Street, 2nd floor
New York, NY 10012
and NBER
rajeev@dehejia.net

Cristian Pop-Eleches
The School of International and Public Affairs
Columbia University
1401A International Affairs Building, MC 3308
420 West 118th Street
New York, NY 10027
and NBER
cp2124@columbia.edu

Miguel Urquiola
Economics Department
Columbia University
420 W 118TH ST #1022
New York, NY 10027
and NBER
msu2101@columbia.edu

Introduction

Recent papers study whether households choose high-value added options across a series of domains. For instance, whether households choose residential neighborhoods that boost children's adult outcomes (Chetty et al. 2018), whether they select effective healthcare providers (Chandra et al. 2016), and whether they choose productive schools and colleges (Beuermann et al. 2019; Abdulkadiroglu et al. 2020; Chetty et al. 2020).

These questions matter for multiple reasons. First, they reveal whether households benefit from their choices, at least along the dimensions that researchers measure. In addition, they provide evidence on whether choice incentivizes providers to compete on value added—as opposed to other dimensions of service quality.[1]

Such work typically finds that households leave substantial value added "on the table." For instance, while better neighborhoods often have higher housing prices, there are "opportunity bargains" that produce good outcomes and are relatively affordable (Chetty and Hendren 2018). Similarly, while more selective colleges tend to be better at increasing earnings, some less selective colleges are also effective (Chetty et al. 2020). Finally, there appears to be little correlation between a school's value added and its popularity once one controls for the achievement level of the school's students.[2]

These findings beg the question of why households do not always favor their highest-value added options. What constraints or factors lead them to other options? This paper explores these questions in the context of school choice. We consider two reasons why households may not choose the schools that researchers deem most productive. First, households may simply lack *information*. Value added is the change in a student's outcomes due to attending a school. This is considerably more difficult to observe than other school attributes, such as the quality of the school's facilities or the achievement level of its students. Thus, it is possible that households do wish to attend high-value added schools, but do not know which those are. On the other hand, it may be that households' *preferences* lead them to prioritize other school traits. For example, a given school might not provide the largest gains in skill, but it may provide a safe environment or a short commute. In this case, households may willingly give up value added in exchange for alternative dimensions of quality.

Distinguishing between preferences and information has important policy implications. If information is the obstacle, then making it available would improve households' choices and spur providers to compete on value added. By contrast, if preferences are the constraint, then policy options to boost value added may be more limited. For instance, school choice may cause schools

1. If households choose providers based on value added, then providers will likely feel pressure to invest in this characteristic (Rothstein 2006).

2. In New York City, the correlation is zero (Abdulkadiroglu et al. 2020); in Trinidad and Tobago it is positive among households with high-achieving children (Beuermann et al. 2019).

to invest in other, possibly less desirable, quality dimensions.

We explore the distinction between preferences and information by studying high school admissions in Romania. We use administrative data, a survey, and an experiment. The administrative data reveal that, on average, households choose schools at the 68^{th} percentile of value added among their feasible options. This means that they could choose schools with about 1 s.d. worth of additional value added. To understand why households leave so much value added unexploited, we first assess households' knowledge of this school trait. We elicit beliefs about value added using a survey and find that these explain less than a fifth of the cross-school variation. We next inform randomly selected households about the value added of the schools in their towns. We find that the treatment improves the accuracy of households' beliefs. In addition, it induces certain students to attend schools with higher value added: namely, low-achieving students who failed to gain admission to their top choices.

To better interpret the experimental treatment effects, we estimate households' preferences for school traits. We find that households care about a number of traits (e.g, location, peer quality) in addition to value added. Given these preferences, we predict how choices would change if households' beliefs about value added were fully accurate. We find that for households with high- (low-) achieving children, fully correcting beliefs would eliminate only 11 (18) percent of the value added left unexploited, absent the experiment. Finally, we provide evidence that the treatment caused households with low-achieving children to care more about value added. As a result, it had a larger effect on their choices than predicted by its impact on beliefs.

To elaborate, our paper relies on two features of Romania's school system. First, high school is bookended by high-stakes standardized exams. Before entering high school, students take a national admissions test, the "transition exam." Before graduating, they take a national exit test, the "baccalaureate exam." Performing well on the baccalaureate exam is crucial for moving on to higher education. Passing is required for admission to any university, and a high score helps win merit scholarships and entry to selective programs. Thus, the two exams enable us to calculate school value added and to do so with regard to a central academic outcome. Using the administrative data, we obtain value added estimates for every high school in the country.

The second feature we rely on is Romania's student assignment mechanism: a serial dictatorship. This algorithm gives each student a score. It then considers applicants one at a time according to their scores, assigning each to his/her most-preferred school that has not yet reached capacity. A household can rank an unlimited number of options; as such, its dominant strategy is to rank truthfully according to its preferences (Chade and Smith 2006). The serial dictatorship thus allows us to observe the high schools that a student could attend, and to be confident that the one she enrolls in is her most preferred among the feasible options. In addition, the algorithm generates school-specific admissions cutoffs. These provide regression discontinuity (RD) estimates of the effect of access to each school. Following Angrist et al. (2017), we use the RDs to validate our value added estimates, which we find closely match causal effects.

In short, the administrative data allow us to calculate value added and to observe the *outcome* of household decision-making. In order to probe the *mechanics* of this decision-making, we visited middle schools to conduct surveys and run an experiment.[3] We collected a baseline survey at school-sponsored information sessions held to help households apply to high school. We interviewed parents to obtain the school preference rankings that they intended to submit for the high school allocation. We also asked them to evaluate the high schools in their town along multiple quality dimensions including value added. We implemented our experiment at the end of these sessions. At randomly selected treatment schools, we explained the concept of value added and distributed a ranking of the town's high schools based on this characteristic. After the assignment process was complete, we matched students with administrative data to obtain their official school assignments. We also ran an endline survey, interviewing parents by phone to gather the school preference rankings that they submitted, and to again elicit their beliefs about schools' value added.

This setup allows us to address four questions, which we use to organize the exposition and frame the contribution of the paper. The questions and a preview of our findings are as follows.

1. Do households choose schools with high value added?

Schools with higher value added face higher demand. The correlation between a school's value added and the selectivity of its admissions cutoff is 0.55.[4] In addition, households choose options that are above average by value added in their feasible choice sets. Nonetheless, they leave considerable value added on the table. Both low- and high-achieving students could gain, on average, about 1 s.d. worth of additional value added—or a 13 percentage point increase in the probability of passing the baccalaureate exam. By contrast, households come close to maximizing selectivity. For this school trait, they leave only 0.3 s.d. unexploited.

These results relate to work asking whether households favor productive schools (Beuermann et al. 2019; Abdulkadiroglu et al. 2020). Our contribution is to exploit a setting which reveals precisely the set of schools among which households choose, and induces households to truthfully reveal the one they most prefer.

2. How accurate are households' beliefs regarding a school's value added?

Households have limited knowledge of value added. When asked to score schools on this trait, their scores are off by an average of 1.1 within-town quintiles and explain only 17 percent of the variation. In contrast, they have more awareness of selectivity. Their scores for this school trait

3. We visited middle schools in May of 2019. Our sample included 194 middle schools in 48 moderately-sized towns. We aimed for towns to have enough options for households to have choice, but not to be so large as to make commuting distance a binding constraint.

4. This is a descriptive result—it could arise because households choose schools based on value added, but it could also arise if households seek a correlate of value added, or if schools that attract high-scoring students benefit from peer effects.

have a mean absolute error of 0.9 within-town quintiles and explain 33 percent of the variation. Finally, households with high-achieving children have more accurate beliefs than those with low-achieving ones. For the former, scores explain 20 percent of the variation in value added and 39 percent of that in selectivity; for the latter households, these fractions are 12 and 23 percent.

Our contribution is to provide, to our knowledge, the first comparison between researchers' and households' perceptions of school value added within entire markets.

3. Would households change their choices if given information on value added?

Providing information on value added improves the accuracy of households' beliefs and causes them to assign higher preference ranks to high-value added schools. Effects are larger for households with low-achieving children and for options that were initially less preferred. Specifically, our experimental treatment had no effect on beliefs or preference ranks for the two options that households ranked highest in the baseline survey. As a result, it had heterogeneous impacts on the real-world outcome of school assignments. For low-achieving students who were rejected by their two top choices, the treatment caused students to attend schools with 0.2 s.d. worth of additional value added. This was a 2.4 percentage point (9.6 percent) increase in the probability of passing the baccalaureate exam. For all other students, the treatment had no effect.

These results address work on whether information on school quality affects households' choices. Previous work finds positive effects from information on schools' *absolute* achievement (Hastings and Weinstein 2008; Andrabi, Das, and Khwaja 2017; Ajayi, Friedman, and Lucas 2017; Corcoran et al. 2018; Allende, Gallego, and Neilson 2019) but limited impacts from information related to value added (Imberman and Lovenheim 2016; Mizala and Urquiola 2013). Our contribution is to experimentally distribute information on value added, and to show that households absorbed the information, a frequent concern with such interventions.

4. How do households' preferences for other school traits affect their demand for value added?

We use households' school preference rankings and their elicited beliefs about school traits to (descriptively) study their preferences for these traits. We first estimate preferences using a discrete choice model. We then disentangle the roles of preferences and information in causing households to leave value added on the table. In particular, we compare predicted school assignments under accurate beliefs about value added with those under baseline beliefs. We predict that correcting households' beliefs would spur low- (high-) achieving students to attend schools with 0.13 (0.11) s.d. worth of additional value added, representing 18 (11) percent of the value added that these households would otherwise leave unexploited. Finally, we show that for households with low-achieving students, the experiment had larger-than-predicted effects on choices because it impacted both beliefs and preferences.

These results relate to papers on households' preferences for school traits (Hastings, Kane, and Staiger 2005; Burgess et al. 2015; Beuermann et al. 2019; Abdulkadiroglu et al. 2020). Our contribution is to calculate preferences using households' beliefs about these traits, rather than values of traits measured by researchers. More broadly, our paper relates to others assessing the roles of preferences and frictions in driving choices (Bergman et al. 2019; Hastings, Neilson, and Zimmerman 2018; Bergman, Chan, and Kapor 2020).

The paper proceeds as follows. Section 0 covers "preliminaries"—inputs needed to address our four guiding questions. Sections 1-4 analyze questions 1-4, respectively. Section 5 concludes.

0 Preliminaries

This section lays the groundwork for examining our four questions of interest. It describes: i) Romania's high school admissions, ii) the administrative data, iii) our value added measure, iv) the survey data, and v) the experiment.

0.1 High school admissions

In Romania, high schools cover grades 9-12 and are divided into *tracks*. These are self-contained units within schools that vary in their curricular focus.[5] The latter fall into three broad categories: a) humanities, b) math or science, and c) "technical" subjects with applied themes such as business or agriculture.[6]

Students are admitted to tracks based on their academic achievement in middle school (grades 5-8). In 8^{th} grade, a student takes a national high school entrance test, known as the transition exam.[7] The student's score on this exam is combined with her middle school GPA to generate an admissions score, called the transition score.[8]

Households have choice over the track their child attends.[9] After finding out its child's transition score, each household submits a ranked list, or preference ranking, of the tracks in its town.[10] The government then assigns students to tracks using a serial dictatorship. This algorithm considers track preference rankings in the order of students' transition scores. It first takes the student with the highest score and assigns him/her to his/her most-preferred track. It then proceeds down the score distribution, assigning each student to his/her most-preferred track that is not yet at capacity. Households' preference rankings can be as long as they wish.[11]

5. Students in a track take the same coursework and are often in the same physical class. High school classes are usually capped at 28 students, and some tracks offer multiple classes.

6. In addition, a small number of students participate in vocational programs that do not provide a path to university. We do not have data on these students.

7. The transition exam includes sections on math and Romanian language and, for students who attend schools for ethnic minorities, a native language exam.

8. For the 2019 admission, the GPA accounted for 20 percent of the score, and the national exam for 80 percent.

9. We assume that schooling decisions happen at the household level, as they involve both parents and students. Sometimes, as short-hand, we refer to a student's preferences or choices. When we do this, we mean the preferences or choices of the household that contains the student.

10. In principle, households can rank any track in the country. In practice, few households rank tracks outside small administrative units we call towns. We present data on this below.

11. More precisely, there is a cap of 287 choices. This number is much larger than the total number of options in most towns, or than most households would realistically consider.

As a result, the serial dictatorship is incentive compatible: the dominant strategy for households is to submit track preference rankings that truthfully reveal their preferences.

Formally, let $\mathcal{J}_i$ be the set of tracks in the town of household i, and let J_i be the number of tracks in this set. The household assigns a rank to each track $j \in \mathcal{J}_i$ based on the utility, U_{ij}, it expects to receive from it. Letting $r_{i,m}$ be the track to which i assigns a rank of m, the household's preference ranking is:

$$\begin{aligned}
r_{i,1} = j &\Leftrightarrow U_{ij} > U_{ik} && \forall\ k \in \mathcal{J}_i \setminus \{j\} \\
r_{i,2} = j' &\Leftrightarrow U_{ij} > U_{ij'} > U_{ik} && \forall\ k \in \mathcal{J}_i \setminus \{j, j'\} \\
&\dots \\
r_{i,J_i} = j'' &\Leftrightarrow U_{ik} > U_{ij''} && \forall\ k \in \mathcal{J}_i \setminus \{j''\},
\end{aligned}$$

where $m = 1$ corresponds to the household's top choice.

Next, let $\mathcal{J}_i^e \subseteq \mathcal{J}_i$ be the set of tracks that are still available when the serial dictatorship considers i's preference ranking. This is the set of tracks that i is eligible to attend, or i's *feasible choice set*. The student is assigned to track j_i^*, where:

$$U_{ij_i^*} > U_{ik} \quad \forall\ k \in \mathcal{J}_i^e \setminus \{j_i^*\}.$$

In words, j_i^* is the household's most-preferred track within the feasible set.

0.2 Administrative data

We use administrative data to calculate value added for each high school track and to examine whether households choose tracks with high value added. These data cover the universe of students admitted from 2004 to 2019. They provide information on students' demographics, academic achievement, and school assignment. Specifically, we observe a student's middle school, middle school GPA, performance on the transition exam, and assigned high school track.[12] We add information on gender by classifying male and female first names. For the 2004 to 2014 admissions cohorts, we also merge data on students' performance on the national high school exit baccalaureate exam.[13] Since our focus is track choice, we restrict the sample in each year to towns that have at least two tracks. For the 2004-2019 admissions cohorts there are, on average, 395 towns that meet this criterion. In these markets, each cohort of about 142,000 students

12. In contrast with several recent papers, we do not observe students' submitted track preference rankings. However, as we discuss below, our survey collects these for a sample of students.

13. We matched the baccalaureate exam data with the admissions data based on a student's name, gender, and high school track. We first conducted an exact merge and then used a fuzzy matching procedure to account for minor changes in the spelling of names. The results are not sensitive to changes in this procedure.

chooses among close to 1,200 high schools and 3,800 tracks.[14]

Table 1: Summary statistics for the administrative data

Covariate	Mean	Std. dev.	Years	N
Female	0.527	0.499	2004-2017, 2019	2,162,736
High school admissions:				
Transition score	7.70	1.35	2004-2017, 2019	2,162,736
Middle school GPA	8.65	0.97	2004-2017, 2019	2,162,736
Transition exam score	7.05	1.69	2004-2017, 2019	2,162,736
Middle school:				
Number of students	50.6	43.7	2004-2017, 2019	2,162,736
Avg. transition score	7.67	0.73	2004-2017, 2019	2,162,736
Avg. middle school GPA	8.63	0.43	2004-2017, 2019	2,162,736
Avg. transition exam score	7.02	0.96	2004-2017, 2019	2,162,736
High school track:				
Number of students	61.2	46.6	2004-2019	2,272,837
Minimum transition score (MTS)	6.92	1.61	2004-2019	2,272,837
Baccalaureate exam:				
Took the exam	0.686	0.464	2004-2014	1,710,030
Passed the exam	0.533	0.499	2004-2014	1,710,030
Perfect score	0.001	0.025	2004-2014	1,710,030

The table describes the administrative data and lists the admissions cohorts for which they are available—due to a reporting issue, we have limited data on the 2018 cohort. "High school admissions" includes variables related to students' transition scores. "Middle school" variables describe students' middle schools. All middle school variables except the number of students are calculated excluding the student of interest; that is, they characterize the student's peers. Variables under "High school track" are characteristics of the student's track, and those under "Baccalaureate exam" relate to the student's performance on the baccalaureate exam.

Table 1 (page 8) describes key variables in the administrative data. One of the covariates merits special comment. The *minimum transition score* (MTS) in a high school track is the transition score of the track's lowest-scoring student. It is the track's admissions cutoff: students with higher scores are eligible to attend the track, while those with lower scores are not. In other words, the MTS measures a track's selectivity and is a proxy for the demand the track experiences—tracks that are more popular reach capacity earlier in the allocation process and thus have higher cutoffs. When the government announces the set of tracks that will be accepting students in a given year, it provides the tracks' MTS from the previous admissions round. Anecdotal evidence suggests that households pay attention to this information in determining their track preference rankings.

0.3 Value added

We calculate value added for each track in each year. Tracks may add value in a number of dimensions. We focus on their impact on a student's performance on the end-of-high-school baccalaureate exam, as this test is salient and has high stakes. Students who pass the exam receive a baccalaureate diploma. Even for students who do not pursue higher education, a

14. Table A1 (page 54) shows the sample size by year. There is variation over time in the number of tracks, reflecting factors such as student enrollment, instructor availability, and the emergence/demise of technical fields.

diploma can be a strong labor market signal. A diploma is necessary for admission to any university, public or private. At less selective universities, it is the only requirement. This is even the case at some selective public schools, provided a student is willing to pay tuition, in contrast to his higher-scoring peers. Aside from tuition-free admission, high scores help students gain access to programs at prestigious universities (Borcan, Lindahl, and Mitrut 2017).[15]

An important feature of the baccalaureate exam is that students choose whether to take it, and whether to continue with other subjects if they fail the first.[16] This means that value added calculated on a student's score could be biased by sample selection. We deal with this by calculating value added on two alternative outcomes. In the main analysis, we focus on value added on the probability that a student *passes* the exam. We show that results are robust to using valued added on the *percentile rank* of a student's exam performance. This latter outcome is defined as the percent of students in an admissions cohort who perform worse on the exam than the given student, with all students who do not pass being assigned a value of 0.[17] We consider these measures to be complementary. Value added on passing the exam relates to a readily interpretable outcome and can be relatively easily explained to parents. Value added on the percentile rank of performance allows more precise results for selective tracks in which large shares of students pass.

We estimate value added using a conventional selection-on-observables model (Rothstein 2010; Angrist et al. 2017). For each student i, let p_i be the outcome of interest. For value added on passing the exam, p_i is an indicator equal to 1 if i passes:

$$p_i = \mathbb{1}\{i \text{ passes the bacc.}\},$$

with $p_i = 0$ if i either fails or does not attempt the test. For value added on the percentile rank of performance, p_i is defined as discussed above.

Let d_{ij} be an indicator equal to 1 if i attends track j, and let X_i be a vector of i's covariates,

15. Rules for admission vary across time and programs. Importantly, they depend on a student's total score on the baccalaureate exam, rather than on exam sub-components. For a limited number of programs at elite universities (such as law or medicine) the baccalaureate plays a limited role, as these use additional admissions tests. In the earlier years in our sample period, additional admissions tests were more common—but baccalaureate performance was still crucial. For instance, Borcan et al. (2017) report that for one elite university, the baccalaureate score represented 50 percent of the admissions score in 2009 and 67 percent in 2010 and 2011, with the rest based on grades and an additional test. Beginning in 2012, the baccalaureate score became the main criterion in most programs.

16. For the 2004-2014 admissions cohorts, 69 percent of high school students attempted the test, 53 percent passed it, and 0.1 percent achieved a perfect score (Table 1). Figure A1 (page 49) shows how these values vary with a student's transition score.

17. The benefits of taking the baccalaureate exam are dependent on passing it. Consequently, students who take the exam but fail to pass obtain the same result as those who do not take it. Since all of these students achieve better than no other students, we assign them a percentile rank of 0. Next, students who receive the minimum passing score are assigned a value equal to the percent of students in the cohort who do not pass. Finally, students with higher scores receive values equal to the percent of students who perform worse.

such as gender and the components of the transition score. We estimate value added by regressing p_i on the set of track attendance dummies, d_{ij}, and on flexible controls for covariates, $f(X_i)$.[18] We allow both value added and the effects of controls to vary by year. Thus, for each cohort, we estimate:

$$p_i = {\gamma_t}' \cdot f(X_i) + \sum_j \mathrm{V}_{jt} \cdot d_{ij} + e_i, \quad i \in \mathcal{I}_t. \tag{1}$$

Here, $\mathcal{I}_t$ is the set of students in cohort t, and V_{jt} is the value added of track j for cohort t. With finite data, we obtain value added estimates $\hat{\mathrm{V}}_{jt}$.

Equation (1) assumes that tracks exert a common effect on all students. However, a track's value added might vary across student types. To allow for this possibility, we calculate value added measures that let a track's effect differ by whether a student is male or female or by whether the student scores more highly in math or language.[19] Specifically, let g index the group that a student falls into, either by gender or relative academic strength. We then fit the model:

$$p_i = {\gamma_{gt}}' \cdot f(X_i) + \sum_j \mathrm{V}_{jgt} \cdot d_{ij} + e_i, \quad i \in \mathcal{I}_{gt}. \tag{2}$$

Here $\mathcal{I}_{gt} \subset \mathcal{I}_t$ is the set of students in cohort t who are in group g, and V_{jgt} is track j's value added for these students.[20]

It turns out that our value added estimates are similar across all measures. In Table A2 (page 54), we correlate estimates for our main measure (track-year effects on passing) with those for the alternative measures (track-year effects on the percentile rank of performance, and passing by student gender or relative academic strength). In all cases, the correlation exceeds 0.9.

Next, Table 2 (page 11), presents information on the magnitude of value added. The results are for our main measure—track-year effects on passing the baccalaureate exam. Specifically, the column labeled $\hat{\mathrm{V}}_{jt}$ presents year-specific standard deviations for estimated value added. The column labeled V_{jt} displays standard deviations for the "true effects," also known as "sig-

18. We specify $f(X_i)$ to include an indicator for female; cubics in the student's: i) middle school GPA, ii) score on the math section of the transition exam, iii) score on the language section, and iv) middle school's enrollment; interactions between female and i)-iv); and levels of variables about other individuals in the student's middle school: a) the standard deviation of transition score, b) the average GPA, c) the average score on the math section of the transition exam, and d) the average score on the language section.

19. For the second partition, we standardize students' scores on the math and language components of the transition exam and identify the one on which the student did better.

20. There is an additional complication when calculating value added on passing the exam. For this measure, p_i is a binary variable, and (1) and (2) are linear probability models. These assume that a track exerts a constant effect (either by year or by year/group) on a student's probability of passing, regardless of her baseline achievement. This may be reasonable in some cases but is a poor assumption for students whose baseline achievement gives them a high chance of passing or failing. To test the sensitivity of our results to this assumption, we have fit versions of (1) and (2) using a logit. This assumes that a track exerts a constant effect on the index function for the probability of passing the exam, rather than on the raw probability itself. The results are hardly changed.

nal" standard deviations (Chetty and Hendren 2018). These values are calculated by adjusting the standard deviations of $\hat{V}_{jt}$ for measurement error.[21] The column titled p_{jt} lists standard deviations for track "pass rates"—the fraction of students in the track-year who pass the exam.

Table 2: Summary statistics for value added on passing the baccalaureate exam

Years	**Standard deviation**				**R-sq.**	**Towns**	**Tracks**	**Students**
	p_{jt}	$\hat{V}_{jt}$	V_{jt}	V^P_{jt}				
2004	0.318	0.212	0.206	-	-	426	3,691	185,383
2005	0.256	0.173	0.163	-	-	405	3,500	146,712
2006	0.289	0.202	0.194	-	-	386	3,284	136,671
2007	0.350	0.220	0.214	-	-	383	3,259	134,692
2008	0.365	0.194	0.187	0.166	0.822	476	4,851	172,174
2009	0.369	0.161	0.153	0.132	0.795	438	4,470	170,087
2010	0.365	0.146	0.137	0.116	0.756	417	4,018	164,146
2011	0.364	0.139	0.130	0.113	0.765	437	4,506	187,442
2012	0.374	0.134	0.123	0.110	0.807	410	4,234	146,114
2013	0.372	0.126	0.114	0.104	0.791	420	4,269	141,934
2014	0.356	0.137	0.125	0.110	0.797	378	3,784	124,675
2015	-	-	0.122	0.109	-	368	3,649	121,880
2016	-	-	0.117	0.104	-	362	3,541	115,902
2017	-	-	0.117	0.104	-	351	3,427	109,694
2019	-	-	0.118	0.105	-	312	3,038	105,230
2004-2007	0.314	0.203	0.196	-	-	1,600	13,734	603,458
2008-2014	0.371	0.151	0.142	0.125	0.793	2,976	30,132	1,106,572
2015-2019	-	-	0.119	0.106	-	1,393	13,655	452,706

The table presents summary statistics for a track's value added on passing the baccalaureate exam. p_{jt} is the pass rate in track j in year t, $\hat{V}_{jt}$ is the value added estimate from equation (1), V_{jt} is the track's true value added, and V^P_{jt} is the forecast based on a local linear forest (Athey et al. 2019). See sections 0.3 and 0.3.2 for details. Standard deviations and R-squared values are weighted by student. R-sq is the fraction of the variation in V_{jt} that is predicted by V^P_{jt}. The values for the standard deviation of V_{jt} and for R-squared are adjusted for measurement error; see Appendix A1. Due to a data reporting issue, we do not calculate values for 2018.

Table 2 reveals that tracks vary widely in both pass rates and value added. We focus on the 2004-2014 cohorts, the years for which we observe baccalaureate outcomes.[22] Within this set we also sometimes distinguish between the 2004-2007 and 2008-2014 cohorts.[23] First, the results for p_{jt} show that pass rates vary widely across tracks. For the 2008-2014 cohorts, a 1 standard deviation increase in a track's pass rate is equal to a 37 percentage point increase in the probability of passing the exam. For the 2004-2007 cohorts, the standard deviation is somewhat smaller and more variable. The results for V_{jt} show that tracks also vary considerably in value added. For the 2008-2014 cohorts, a one standard deviation increase in value added is equivalent to a 14 percentage point increase in the probability of passing. Year-specific values range from

21. Appendix A1 (page 69) describes the procedure.

22. Section 0.3.2 explains why there are values in the column labeled V_{jt} for the 2015-2019 cohorts.

23. In the first of these there were frequent instances of cheating. Beginning with the 2008 cohort, the government cracked down on this by installing video surveillance in exam centers and by drastically increasing punishments. These measures greatly reduced cheating for the later cohorts (Borcan et al. 2017). We find that dropping the 2004-2007 cohorts does not affect our main results. Consequently, we include them, with the caveat that a track's value added in this period could reflect both effects on learning and opportunities for cheating.

11 to 19 percentage points. Thus, in these years, value added explains between 9 and 26 percent of the variation in pass rates. For the earlier period, a 1 standard deviation increase in value added amounts to a 20 percentage point increase in the probability of passing. This can vary from 16 to 21 percentage points. Thus, in this period, value added explains between 37 and 45 percent of the variation in pass rates. Finally, the results for $\hat{\text{V}}_{jt}$ reveal that measurement error only slightly inflates the standard deviation of estimated value added.

0.3.1 Validating value added

Value added calculated relying on the selection-on-observables assumption may suffer from bias. In particular, it will fail to capture the causal effect of attending a track if students' track choices are correlated with the unpredictable component of their baccalaureate performance (Rothstein 2010; Angrist et al. 2017). Prior work has found that selection-on-observables value added measures nevertheless often closely approximate causal effects (Rothstein 2010, 2017; Chetty, Friedman, and Rockoff 2014; Deming 2014; Angrist et al. 2017). However, whether this holds in any particular setting is an empirical question.

Fortunately, Romania offers a natural experiment to test the validity of our value added measure. As stated, the serial dictatorship creates an admissions cutoff for each track. We can thus estimate the causal effect of being eligible to attend a track using a regression discontinuity (RD) design that compares outcomes for students who score just above the cutoff with those of students who score just below.

Appendix A2 (page 71) explains how we assess the quality of our value added measure using the structure of the RD effect. Intuitively, the RD effect for a particular track c is a weighted sum of the local average treatment effects of attending the track versus each of the less-selective tracks in the town. If there is no selection bias and if we appropriately capture treatment effect heterogeneity, then the local average treatment effect of attending track c versus fallback track f is equal to the difference in value added between the two tracks. In order to obtain a quantity that is comparable with the RD treatment effect, one has to appropriately weight these value added differences. We do this by running the RD on the value added of a student's track. Thus, for each track, we calculate two RDs: the traditional one, on a student's own outcome, and a non-traditional one, on the value added of the student's track. If the value added measure is valid, these RDs are weighted sums of the same treatment effects and are calculated using the same weights. Thus, they should be equal, at least up to measurement error.

We test this equality in two ways. First, we calculate the fraction of the variation in the RDs on students' baccalaureate outcomes that is explained by the RDs on the value added of students' tracks. Second, we adapt an IV procedure developed by Angrist et al. (2017), which allows us to test for bias using all tracks at once. The results (Appendix A2) suggest that our value added measures closely match a track's causal effect. In addition, our main measure of a single track-year effect performs as well as the measures that allow for treatment effect heterogeneity by gender or by relative academic strength.[24]

0.3.2 Forecasting value added at the time of track choice

Our experiment aims to inform a household about what a track's value added will be for the admissions cohort of its child. This is a non-trivial task because a track's value added for a given cohort is not known when households make their choices—it cannot be observed until students take the baccalaureate exam. To deal with this, we forecast value added using the information available at the time of track choice. We obtain the forecasts using a local linear forest algorithm (Athey et al. 2019).[25] Our model incorporates current and lagged values of a large number of track covariates, including past value added. Tables A3-A5 (pages 55-56) list the covariates and lags we use: the first lists the covariates that relate to the track itself, the second displays the covariates that relate to the track's high school, and the third lists covariates of the track's town.

We make predictions, V^P_{jt}, for the 2008-2019 admissions cohorts. These are the years for which we have sufficient prior data to compute lagged values of covariates. The predictions are "out-of-bag" in that they use only the trees in the forest that do not include the track-year being predicted.[26] The column labeled V^P_{jt} in Table 2 (page 11) lists the standard deviations of the predictions by year. For the 2008-2014 cohorts, these are similar to, but slightly smaller than, those of the true effects. Next, for the 2015-2019 cohorts, we use the standard deviations of the predictions to compute standard deviations for the true effects.[27] Reassuringly, these values are similar to those for the cohorts that immediately precede 2015 (e.g., 2012-2014). Finally, for the 2008-2014 cohorts, we can formally compare predicted and true value added. We do this by calculating R-squared in predicting true value added using our predictions, as described in Appendix A1.4 (page 70). The results, in the column labeled R-sq. in Table 2, show that our model has substantial predictive power. Overall, our predictions account for 79 percent of the variation in tracks' true value added.

24. One might wonder why we do not focus our analysis on RD effects rather than value added. There are two reasons. First, the RD effects are much noisier. In particular, the RD treatment effect of attending track c is a local average treatment effect calculated by dividing the reduced-form RD treatment effect of being eligible to attend track c by the first-stage RD treatment effect on the probability of attending the track. This value is only for students with transition scores at the cutoff. For a single track-year, these quantities can be very noisy. Second, as alluded to above, RD treatment effects have a complex interpretation: the RD treatment effect of attending track c is a weighted average of pairwise treatment effects between track c and each of the less-selective tracks in a town. It depends on both tracks' causal effects and on the probabilities that students "fall back" to each of the less selective tracks if not admitted to c. If we had data on track preference rankings, we could disentangle these factors, but we do not.

25. This algorithm combines a random forest with a local linear regression. Athey et al. (2019) find that it improves over a random forest when there is a smooth relationship between outcomes and covariates.

26. We account for missing values of covariates by substituting the mean.

27. Appendix A1.3 (page 70) elaborates on the procedure.

0.4 Baseline survey

The administrative data allow us to measure value added and to observe households' track choices, but provide little insight into households' preferences or beliefs. We therefore conducted a baseline survey and an experiment. In the survey, we interviewed parents of 8^{th} graders to collect: i) their beliefs about the attributes of tracks in their towns, and ii) their intended track preference rankings. To do this, we visited information sessions organized by middle schools to inform parents about the high school application process. These occur in May, about a month before households submit their track preference rankings.

0.4.1 Sample selection

To select our sample, we had to choose towns and then middle schools within towns. We chose towns using two criteria. First, we considered only moderately-sized towns, defined as those that had between 7 and 28 tracks in 2018. We excluded towns with fewer than 7 tracks because we wanted households to have a significant number of options; we dropped towns with more than 28 tracks because larger towns are more likely to contain unobserved sub-markets based on commuting distance. Next, among moderately-sized towns, we chose those in which value added was most predictable. Specifically, for each town we calculated R-squared from predicting value added, V_{jt}, using predicted value added, V^P_{jt}, over the 2008-2014 admissions cohorts. We then selected the towns with the highest R-squared values. Finally, we chose middle schools which had at least 15 students and in which it was logistically feasible for our surveyors to visit the information sessions.

Our target sample consisted of 228 middle schools in 49 towns. Some schools did not grant us permission to conduct the survey, and thus our final sample covered 194 middle schools in 48 towns.[28] The towns had an average R-squared of 0.77. For the 2019 cohort—the year in which we conducted the survey—they had an average of 13 tracks and 412 students. We interviewed the households of 3,898 students, with an average of 81 students per town.[29]

0.4.2 Survey questions

The baseline survey asked parents about their intended track preference rankings, and attempted to recover their beliefs about the tracks in their towns.[30] To capture beliefs, we asked parents to score the tracks—on a scale of 1 to 5—on the dimensions listed in Table 3 (page 15). We first

28. In order to minimize spillovers and general equilibrium effects, we visited only a fraction of middle schools in each town. On average, we visited 11 percent of middle schools, and in no town did we visit more than a third.

29. The information sessions are held separately for groups of about 30 students known as classrooms. In most middle schools we visited the session for a single classroom. However, to increase the sample size, in 44 middle schools we visited sessions for two classrooms. Table A6 (page 57) describes the towns included in the survey.

30. A labor dispute caused the government to delay announcing the list of tracks that would be available in 2019. As a result, we lacked time to gather information on tracks that were newly created in 2019. Instead, we asked households only about tracks that existed in both 2018 and 2019. This is not a major issue, as only 43 out of 614 tracks in the survey towns existed in 2019 and not in 2018.

asked about two school attributes that research consistently indicates households care about: location and peer quality. We also asked about our definition of value added ("this track will help my child pass the baccalaureate exam"), as well as alternative dimensions of value added related to college and labor market success. Finally, we asked parents to score tracks on teacher quality, on whether their curriculum is a good fit for their child, and on whether the track is attractive because it is also used by their child's siblings or friends.

Table 3: Track characteristics covered in the survey

Characteristic	**Definition**
Location	This track has a convenient physical location (close to my home or preferred means of transport)
Peer quality	This track attracts academically gifted students
VA: pass the bacc.	This track will help my child pass the baccalaureate exam
VA: college	This track will help my child go to the college that I would like for him or her
VA: wages	This track will raise my child's earnings at age 30
Teacher quality	This track has good teachers
Curriculum	My child will enjoy this track's curriculum
Siblings & friends	My child's siblings and friends also attend this track (or this track's school)

The table displays the definitions of the track characteristics covered in the baseline survey. Survey respondents were asked to score tracks on these characteristics on a scale of 1 to 5.

Table A7 (page 58) presents summary statistics for parents' scores. It shows that means and standard deviations are similar for the various quality dimensions. Table A8 (page 58) presents an across-dimension correlation matrix. Its values lend credibility to the scores. For instance, the largest correlations are those among the three value added dimensions; the lowest are those that include scores for a track's location or for whether the child's siblings/friends attend it.

Finally, Table 4 (page 16) describes other variables in the survey data. It reveals a few notable facts. First, households do not rank all tracks; they focus on only a subset of options. Namely, on average households assigned ranks to 43 percent of the tracks in their town, and they assigned scores to 35 percent. Appendix A4 (page 77) provides additional details on households' behavior in this regard. Table 4 also shows that we were able to match 83 percent of students in the survey sample with administrative data on the 2019 admissions cohort.[31] For these students, we observe official transition scores and track assignments.

0.5 Experiment

We conducted an experiment to explore whether households' choices can be influenced by providing information on value added. This took place at the end of the middle school information sessions at which we implemented the baseline survey. We randomly split the middle schools into

31. We matched students by name and middle school using a fuzzy matching procedure allowing for slight misspellings of names. Students do not appear in the administrative data if they do not get assigned to a track. This occurs when a student does not submit a track preference ranking or when the student does not rank any tracks that he/she is eligible to attend. A small number of the unassigned students participate in a secondary allocation that occurs at the end of the summer. These students get assigned to tracks that did not reach capacity in the initial allocation. The remaining students either drop out of school or attend vocational schools.

Table 4: Summary statistics for the survey data

	Mean	Std. dev.	Min	Max	N
High school application process:					
Num. of tracks in the town	13.0	4.62	5	23	3,898
Share of tracks scored on passing the bacc.	0.35	0.42	0	1	3,898
Share of tracks scored on peer quality	0.37	0.42	0	1	3,898
Share of tracks ranked	0.43	0.33	0	1	3,898
Very certain of preference ranking	0.43	0.49	0	1	3,560
Somewhat certain of preference ranking	0.50	0.50	0	1	3,560
Care more about tracks than schools	0.63	0.48	0	1	3,341
Intend to apply in the town	0.93	0.25	0	1	3,898
Parent's prediction for student's transition score:					
Predicted middle school GPA	9.02	0.91	5	10	3,538
Predicted transition exam score: math	7.51	1.64	1	10	3,773
Predicted transition exam score: language	8.19	1.43	2	10	3,775
Student demographics:					
Female	0.52	0.50	0	1	3,898
Mother's years of schooling	12.0	2.17	0	14	3,759
Parents not married	0.16	0.36	0	1	3,609
Administrative data:					
Matched with the administrative data	0.83	0.38	0	1	3,898
Transition score	7.83	1.36	2.8	10	3,218

The table describes the survey data. The sample consists of 3,898 students in 194 middle schools in 48 towns. Of these, 3,218 were matched with administrative data on the 2019 admissions cohort. The variables under "Parent's prediction for student's transition score" are predictions that parents made during the information session.

treatment and control groups.[32] We concluded the survey by distributing a flyer with information on the high school application process.

In the control middle schools, the flyer provided links to government websites, including one listing the prior-year minimum transition score for each track. In the treatment middle schools, the flyer also explained the concept of value added ("which tracks most effectively improve students' chances of passing the baccalaureate exam relative to their 9^{th} grade starting points") and included a ranking of the tracks in the town by our prediction for value added. An example treatment flyer is in Figures A2-A3 (pages 50-50). Control flyers were identical to Figure A2, save for excluding the fourth bullet point. Respondents were allowed to keep the flyers.

After the high school allocation occurred, we obtained students' track assignments from the Ministry of Education. In addition, we conducted a follow-up (or "endline") phone survey. In this survey, we collected the final track preference rankings that households submitted, asked households to score tracks on a scale of 1 to 5 in terms of value added, and inquired about households' experience with the admissions process. The data on track assignments allow us to observe whether the information affected the tracks that students attend. The follow-up survey allows us to probe the mechanics by which the information influenced choices.[33]

32. We used a clustered randomization process. We first matched pairs of middle schools within towns based on school characteristics. We then randomized within these matched pairs. Appendix A3 (page 77) provides details.

33. In the time between our creation of the matched pairs and our baseline survey, some of the middle schools in our sample withdrew their permission for our study. For every school at which this occurred, we still conducted the baseline survey in the other middle school in its matched pair. However, we removed these other schools from

Table 5: Summary statistics and balance tests for the experiment

Covariate	Summary statistics		Balance tests			
	Mean	Std. dev.	Coef.	Std. error	Clusters	N
Matched with the administrative data	0.859	0.348	0.022	0.019	78	3,186
Assigned to a track	0.846	0.361	0.023	0.020	78	3,186
In the follow-up survey	0.569	0.495	-0.014	0.026	78	2,692
Student demographics:						
Female	0.530	0.499	0.015	0.022	78	2,692
Mother's years of schooling	12.3	1.9	0.111	0.102	78	2,625
Parents not married	0.136	0.343	-0.013	0.015	78	2,516
Parent's prediction for student's transition score:						
Predicted middle school GPA	9.17	0.75	0.027	0.057	78	2,515
Predicted transition exam score: language	8.47	1.16	0.055	0.084	78	2,653
Predicted transition exam score: math	7.79	1.46	0.156	0.104	78	2,653
High school application process:						
Num. of tracks in the town	13.1	4.6	0.260	0.324	78	2,692
Share of tracks ranked	0.478	0.312	-0.011	0.029	78	2,692
Share of tracks scored on peer quality	0.422	0.424	-0.005	0.032	78	2,692
Share of tracks scored on passing the bacc.	0.411	0.421	-0.014	0.032	78	2,692
Very certain of preference ranking	0.443	0.497	0.038	0.027	78	2,642
Somewhat certain of preference ranking	0.498	0.500	-0.022	0.022	78	2,642
Care more about tracks than schools	0.605	0.489	0.026	0.032	78	2,496
Administrative data:						
Transition score	7.87	1.31	0.126	0.093	78	2,692
Middle school GPA	9.20	0.68	0.041	0.051	78	2,692
Transition exam score: language	8.13	1.48	0.176*	0.101	78	2,692
Transition exam score: math	6.91	1.80	0.125	0.126	78	2,692

The table presents summary statistics and balance tests for the experimental sample. The sample consists of 3,186 students in 170 middle schools in 45 towns. "Matched with the administrative data" is an indicator for whether the student was found in the administrative data. "Assigned to a track" is an indicator for whether the student was found in the administrative data and assigned to a track in the high school allocation. The sample for the remaining variables is limited to the students for whom "Assigned to a track" is equal to 1. "Coef." is the coefficient in a regression of the listed variable on the treatment indicator. It measures the difference in the mean of the variable between treatment and control groups. Standard errors are clustered by the middle school treatment-control pairs within which we conducted the randomization. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5 (page 17) presents summary statistics and balance tests for the experiment. The first row displays the share of students in the experimental sample that we were able to match with the administrative data. Students do not appear in the administrative data if they do not participate in the high school allocation.[34] In addition, it is possible that a small number of students were not matched due to data issues, such as misspelled names. In total, 86 percent of students were matched, with an insignificant difference of 2.2 percentage points between students in the treatment and control groups. The second row shows the share of students for whom we observe track assignments. These are students who appear in the administrative data and who listed at least one track in their preference rankings that they were eligible to attend. Overall, we observe track assignments for 85 percent of students, with an insignificant difference of 2.3

the experimental sample. In addition, we removed students who said that they did not intend to apply to high school in the town of their middle school. Thus, while the survey sample includes 3,898 students in 194 middle schools in 48 towns, the experimental sample includes only 3,186 students in 170 middle schools in 45 towns.

34. Recall that students can instead attend vocational school, which involves a different application process.

percentage points between treatment and control groups.

The values in the remaining rows of Table 5 are calculated for the sample of students for whom we observe track assignments. Comparing these values with those in Table 4 (page 16) reveals that students in the experiment are representative of those in the baseline survey. Moreover, the balance tests suggest that the randomization was successful: the differences between treatment and control groups are small relative to the variables' standard deviations. Only one of these differences is statistically significant (at the 10 percent level).

We now use the above inputs to address the four questions that motivate this paper.

1 Do households choose tracks with high value added?

The first question we study is whether households choose tracks with high value added. This question matters because it reveals whether households gain academic benefits from their choices. It also reveals whether there is scope to increase their benefits by influencing their choices. We note that our analysis of this question is descriptive. In particular, the results only illuminate whether the tracks that households choose happen to have high value added. They do not reveal whether households make choices based on value added.[35]

We address this using the administrative data. We first examine whether a track's value added is correlated with the demand it faces, as measured by the selectivity of its admissions cutoff. We then exploit our knowledge of households' feasible choice sets to compare the tracks that households choose with their available options. Notably, we calculate the amount of value added that a household could gain by having its child switch to its best available option.

1.1 The relationship between value added and selectivity

We start by inspecting the relationship between a track's value added and its selectivity as measured by its cutoff or MTS. This relationship will be positive if tracks with high value added are popular and reach capacity early in the assignment process.[36]

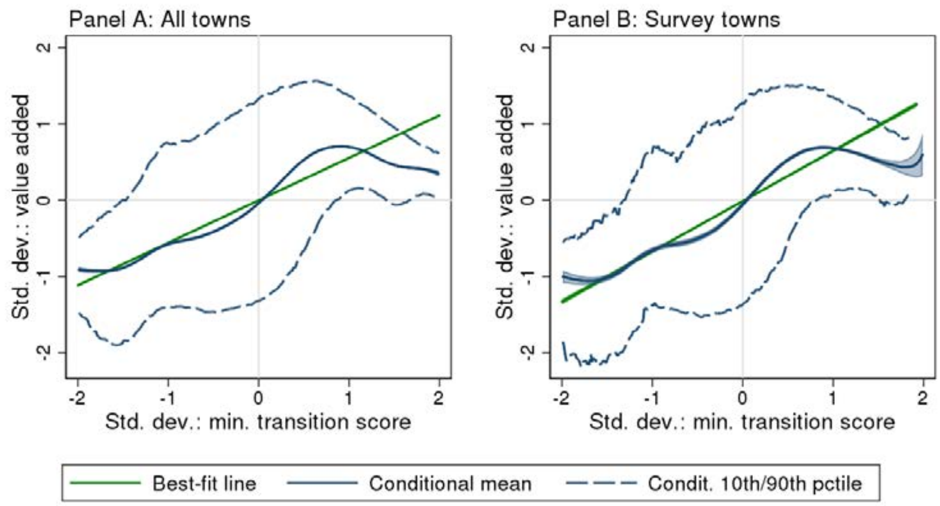
Figure 1 (page 19) displays this relationship for our main value added measure, a track-year effect on the probability of passing the baccalaureate exam. It summarizes how a track's estimated value added, $\hat{\mathrm{V}}_{jt}$, varies with its minimum transition score, MTS_{jt}. Specifically, the figure plots the conditional mean and the conditional 10^{th} and 90^{th} percentiles of $\hat{\mathrm{V}}_{jt}$ for given values of MTS_{jt}. It also includes a best-fit line from a linear regression. Since both variables are standardized, the slope of this line equals the variables' correlation coefficient.[37] The figure uses

35. One way to understand this distinction is to imagine what would happen if a track were to invest in improving its value added. If households make choices based on value added, then the track would become more popular. However, if households choose based on other characteristics, then its popularity would be unchanged.

36. We reiterate that a positive relationship need not imply that households choose tracks based on value added. For example, more selective tracks have higher-achieving students and could thus have higher value added due to peer effects or teacher sorting.

37. Technically, this is true only for the full sample. For best-fit lines calculated on subsets of the sample, the

Figure 1: The relationship between value added, $\hat{V}_{jt}$, and selectivity, MTS_{jt}



The figure summarizes the relationship between value added and selectivity. The best-fit line is from a linear regression of standardized values of value added estimates, $\hat{V}_{jt}$, on standardized values of minimum transition score, MTS_{jt}. "Conditional mean" plots predictions from a local linear regression, and "conditional 10^{th} and 90^{th} percentiles" from local quantile regressions. The value added measure is a track-year effect on the probability of passing the baccalaureate exam. Variables are standardized by year, and regressions weighted by student. The results are for the 2004-2017 and 2019 cohorts. For cohorts after 2014, they use predicted value added, V^P_{jt}. Panel *A* presents results for the full set of towns, and Panel *B* for towns in the baseline survey.

all cohorts and provides results separately for the full set of towns (Panel A) and for the towns in the baseline survey (Panel B). In each, there is an overall positive relationship between value added and selectivity, as captured by the positively sloped best-fit lines.[38] However, this overall relationship obscures substantial nonlinearity. The relationship is strongly positive for less- and moderately selective tracks but is flat or even negative for highly selective tracks.

Table 6 (page 20) presents statistics that quantify the results in Figure 1: coefficients from regressions of standardized value added estimates on standardized minimum transition score. The values in the rows labeled "All tracks" match the slopes of the best-fit lines in Figure 1. The remaining values are meant to capture the non-linearity in the figure—they are coefficients from regressions that split the sample by tercile of selectivity. The table confirms that the overall correlation between value added and selectivity is substantial. For the country as a whole, it is 0.55; for survey towns, it is 0.64. However, this is driven entirely by tracks in the bottom two-thirds of the selectivity distribution. For the most-selective third, an increase in selectivity is associated with a *decline* in value added. Among all towns, this decline is 0.26 standard deviations for a one standard deviation increase in selectivity; for survey towns, it is 0.17.

These results are robust; they persist irrespective of the time period or value added measure used. Table A9 (page 58) presents year-specific correlations, showing that these values are

slope can be interpreted as the increase in (full-sample) standard deviations of value added for one (full-sample) standard deviation increase in selectivity.

38. This parallels findings in Abdulkadiroglu et al. (2020) and Beuermann et al. (2019).

Table 6: Regressions of standardized value added estimates on standardized selectivity

Sample	Coefficient	Std. error	Town-years	Track-years	Students
Panel A: All towns					
All tracks	0.545	0.005	5,969	57,521	2,162,736
By tercile of selectivity:					
Least selective	0.387	0.017	5,710	24,934	723,446
Moderately selective	1.049	0.027	4,325	17,207	723,023
Most selective	-0.264	0.024	2,420	15,380	716,267
Panel B: Survey towns					
All tracks	0.644	0.010	720	11,253	424,508
By tercile of selectivity:					
Least selective	0.400	0.041	717	4,319	135,007
Moderately selective	1.169	0.054	718	3,898	162,887
Most selective	-0.174	0.041	676	3,036	126,614

The table quantifies the results in Figure 1. It presents coefficients from regressions of standardized values of value added estimates, $\hat{V}_{jt}$, on standardized values of minimum transition score, MTS_{jt}. The coefficients from the rows labeled "All tracks" match the slopes of the best-fit lines in Figure 1. The value of this coefficient in Panel *A* has an interpretation as a correlation coefficient. "Tercile of selectivity" indicates whether the track is in the lowest, middle, or highest third of MTS_{jt} by year. Regressions are weighted by student, and standard errors are clustered by town-year. See Figure 1 for additional details.

stable over time. Figure A4 (page 51) replicates Figure 1 for alternative value added measures, displaying similar results.[39]

Finally, we ask whether the results are affected by compositional changes in the share of tracks with a given curriculum. In particular, it is possible that the relationship between value added and selectivity is uniformly positive for tracks with the same curriculum, but that less- and more-selective tracks tend to differ in curriculum. To examine this, we run the local linear regressions in Figure 1 separately by whether a track focuses on humanities, math and science, or technical subjects. The results are in Figure A5 (page 51). The relationship between value added and selectivity within these track categories is broadly similar to that in the full sample.

1.2 Comparing households' choices with the available options

In order to further characterize choice behavior with regard to value added, we next take advantage of information on a household's feasible choice set. This is the set of tracks that a student is eligible to attend, given the admissions cutoffs and the student's transition score. We compare the value added of the track the household chooses with the value added of the other

39. This latter finding mitigates one potential concern regarding Figure 1. In particular, it is conceivable that the negative relationship between value added and selectivity for highly selective tracks reflects a mechanical constraint on value added for these tracks. If students in highly selective tracks are certain to pass the baccalaureate exam regardless of the track they attend, then there will be a cap on the tracks' estimated value added. However, this cap is less likely to be binding for value added on the percentile rank of a student's exam performance. Figure A4 shows that results are unchanged for this alternative measure. In addition, we have explored the relationship between selectivity and a track's value added on a student's exam score. As discussed, we are generally hesitant to use value added on exam score due to selection into exam-taking. However, selection is of limited concern for highly selective tracks. Further, there is unlikely to be a ceiling on this value added measure, since only 0.1% of students achieve a perfect score (Table 1). Reassuringly, we again find a negative relationship.

options in this set. We also compute the amount by which a household could increase the value added its child receives by switching to its highest-value-added option. As a comparison, we present a parallel analysis for selectivity, asking whether households choose the tracks with the highest-achieving students.

To elaborate, for each household we calculate two quantities. First, the *percentile rank of the student's track among feasible tracks* is the rank of the student's track (by either value added, $\hat{V}_{jt}$, or selectivity, MTS_{jt}) within the feasible set divided by the number of tracks in the set. It represents the share of feasible tracks with values that are less than or equal to that of the track the student attends.[40] Second, the *potential increase among feasible tracks* is the difference between the maximum value (of value added or selectivity) within the feasible set and the value for the student's track. It captures how much of an improvement a household could obtain by optimally switching its child's track.

Table 7: Summary statistics on households' track choices

	All towns			**Survey towns**		
	All students	Low-achieving	High-achieving	All students	Low-achieving	High-achieving
Panel A: Percent of students with only one feasible track	2.4	4.8	0.0	1.3	2.6	0.0
Panel B: Mean percentile rank of student's track among feasible tracks						
Value added, $\hat{V}_{jt}$	67.9	63.3	72.3	67.8	61.7	73.7
Selectivity, MTS_{jt}	81.0	74.9	86.9	79.7	74.6	84.8
Panel C: Mean potential increase (std. dev.) among feasible tracks						
Value added, $\hat{V}_{jt}$	1.07	1.06	1.09	0.93	0.94	0.92
Selectivity, MTS_{jt}	0.32	0.34	0.30	0.34	0.34	0.35
Number of students	2,162,736	1,081,075	1,081,661	424,508	211,917	212,591

The table presents summary statistics on households' track choices. Panel *A* displays the percent of students who are eligible for only one track. Panels *B* and *C* show means for the "percentile rank of the student's track among feasible tracks" and the "potential increase among feasible tracks". See Section 1.2 for definitions. The values in Panels *B* and *C* are calculated for students with multiple feasible options. The sample includes the 2004-2017 and 2019 cohorts. Variables are standardized by year. A student is defined as low- (high-) achieving if his/her transition score is in the bottom (top) half of the within-year distribution. The value added measure is a single track-year effect on the probability of passing the baccalaureate exam.

Table 7 (page 21) presents results for our main value added measure. It uses all cohorts of data and provides estimates both for all towns (columns 1-3) and for towns in the baseline survey (columns 4-6). Panel A lists the share of students who have only one track in their feasible set and hence no choice over tracks. In the full sample, 2 percent of students are in this situation. The remaining panels concern the students with choice. Panel B reveals that on average these students attend tracks at the 68$^{\text{th}}$ percentile of value added among their feasible sets. Panel C shows that they thus leave substantial value added on the table. If they were to switch to their value added-maximizing options, they would gain, on average, 1 standard deviation of value

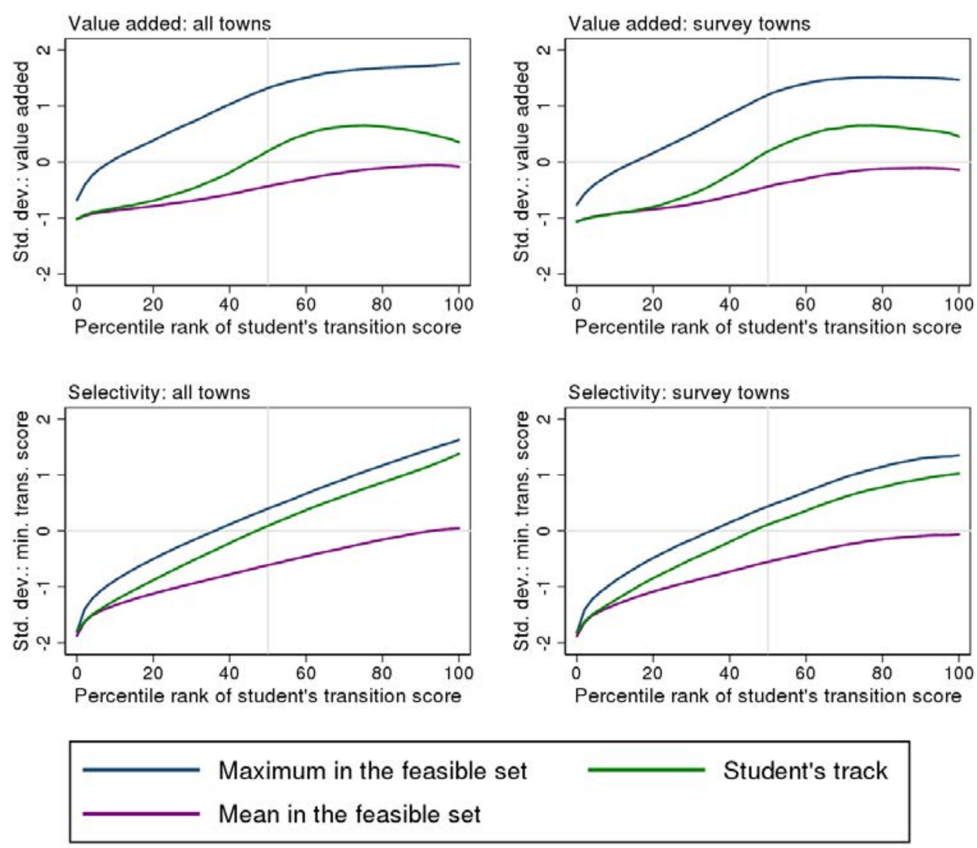
40. A value of 100 indicates that the household chooses the best option by value added or selectivity within its feasible set. A value of $100/J_i^e$ implies that the household chooses the worst option.

added. Based on 2019 data, this is equal to a 13 percentage point increase in the probability of passing the baccalaureate exam. Results for survey towns (column 4) are similar.

The results for selectivity show that students are much closer to "maxing out" on this track characteristic. Over the full sample, students on average attend tracks with selectivity at the 81st percentile among their feasible tracks. This amounts to an average potential increase in selectivity of only 0.32 standard deviations. The results for survey towns are again similar.

The remaining columns of Table 7 explore whether there is heterogeneity in choice patterns based on a student's academic achievement. They present results separately for students with transition scores in the bottom half ("low-achieving") and top half ("high-achieving") of the within-year distribution. These columns suggest that there is limited heterogeneity.

Figure 2: Choice patterns by transition score



The figure shows how choice patterns vary with a student's transition score. Specifically, it plots the relationship between the percentile rank of the student's transition score and three choice-related variables. The blue line is the maximum value of value added, $\hat{V}_{jt}$, or selectivity, MTS_{jt} in the student's feasible set. The purple line is the mean value in the feasible set, and the green line is the value in the track the student attends. The lines are calculated using local linear regressions. The difference between the blue and green lines is the mean potential increase for the given percentile rank. See Table 7 for additional details.

This is visible in Figure 2 (page 22), which shows how the potential increase in value added or selectivity varies with the percentile rank of a student's transition score. The figure plots the relationship between the student's percentile rank and three variables: i) the maximum value (in standard deviations of value added or selectivity) in the student's feasible set, ii) the mean value in the set, and iii) the value for the track the student attends. The difference between the lines for the maximum and for the value of the student's track is equal to the mean potential increase in value added or selectivity for students with a given transition score. For both outcomes, these potential increases are relatively constant across the transition score distribution (they are smaller for the lowest-achieving students, who have limited choice). We also find that the results in this section are robust to a range of alternative specifications.[41]

Finally, we investigate whether results are influenced by the fact that tracks differ in curriculum. If households have strong curricular preferences, they may choose the best available track within their preferred curriculum. In this story, households leave value added on the table because they willingly exchange it for this other track characteristic. We find that this holds only partially. Table A10 (page 59) replicates Table 7 while restricting a household's choice set to the subset of feasible tracks whose curricula fall into the same category as that of its child's track. It shows that for all towns the average student attends a track with value added (selectivity) at the 65^{th} (80^{th}) percentile among this restricted choice set (Panel B). On average, these students could obtain increases in value added (selectivity) of 0.6 (0.3) standard deviations (Panel C). In survey towns, the corresponding values are 0.5 (0.3) standard deviations. Thus, even within curricular categories, households leave significant value added unexploited.

2 Do households have accurate beliefs regarding value added?

The previous section examined the *consequences* of household decision-making, asking whether students end up at one of the tracks with higher value added among those in their feasible sets. This section investigates a key aspect of the *mechanics* of decision-making. Namely, the accuracy of households' beliefs regarding track value added.

To do this, we use the baseline survey, in which we asked households to score tracks on a variety of dimensions on a scale of 1 to 5 (Table 3, page 15). We analyze the accuracy of these scores by comparing them with the values of track characteristics that we observe as researchers (the "true values"). In particular, we compare households' scores for a track's value added with our predictions for this characteristic, V^{P}_{jt}. As a benchmark, we also compare their scores for a track's peer quality with the track's prior-year selectivity, MTS_{jt-1}. This benchmark is noteworthy because selectivity is salient in Romania and households can view each tracks' prior-year selectivity on the official admissions website. Thus, this benchmark represents the accuracy of households' quality scores under a scenario of potentially full access to information.[42]

41. Figure A6 (page 52) replicates Figure 2 using alternative value added measures, with similar results

42. Arguably, households' peer quality scores should reflect a track's *current-year* selectivity, rather than its prior-year selectivity. In determining peer quality scores, a rational household might combine the data on prior-year selectivity with its knowledge of changes in tracks' characteristics from the prior year to the current year. To explore this, we have calculated results using current-year selectivity and find that they are similar.

We begin by simply regressing the true values of track characteristics on households' quality scores. Since the scores are on a scale of 1 to 5, we transform the true values into corresponding units by calculating their within-town quintile. Specifically, we estimate:

$$\begin{aligned}\text{quint}(\text{V}^P_{jt}) &= \alpha_{0,V} + \alpha_{1,V} \cdot s^V_{ij} + \alpha_{ij,V} \\ \text{quint}(\text{MTS}_{jt-1}) &= \alpha_{0,PQ} + \alpha_{1,PQ} \cdot s^{PQ}_{ij} + \alpha_{ij,PQ},\end{aligned} \tag{3}$$

where quint$(\cdot)$ is a within-town quintile. Here $\alpha_{1,V}$ is the association between a one point increase in a household's quality score for value added, s^V_{ij}, and within-town quintiles of our prediction for value added, quint(V^P_{jt}). Similarly, $\alpha_{1,PQ}$ is the association between a one point increase in a household's score for peer quality, s^{PQ}_{ij}, and within-town quintiles of prior-year selectivity, quint(MTS_{jt-1}). Thus, if households' scores were fully accurate, then the slope coefficients and R-squared would both equal 1.

Table 8: Explaining within-town quintiles of track attributes using households' quality scores:

	All students		**Low-achieving**		**High-achieving**	
	quint(V^P_{jt})	quint(MTS_{jt-1})	quint(V^P_{jt})	quint(MTS_{jt-1})	quint(V^P_{jt})	quint(MTS_{jt-1})
Score: VA-pass, s^V_{ij}	0.416***		0.380***		0.435***	
	(0.019)		(0.032)		(0.018)	
Score: Peers, s^{PQ}_{ij}		0.572***		0.507***		0.611***
		(0.016)		(0.032)		(0.012)
R-sq.	0.17	0.33	0.12	0.23	0.20	0.39
Clusters	188	188	171	171	177	177
Students	2,370	2,370	883	883	1,487	1,487
Student-tracks	17,460	17,460	6,433	6,433	11,027	11,027

The table presents regression results from equation (3). quint$(\cdot)$ is the within-town quintile of the given variable. The sample drops student-track observations if the survey respondent did not score the track on both value added and peer quality. It also excludes observations for 43 tracks that were newly created in 2019. Standard errors are clustered by middle school.

The results from regression (3) are in Table 8 (page 24). The first two columns provide results for all students; the remaining columns provide them separately for low- and high-achieving students. The first column is for the regression with respect to value added; it shows that households' scores are significant but imperfect predictors for this track characteristic. A one point increase in a household's score for value added is associated with a 0.44 quintile increase in our prediction for value added. The R-squared reveals that households' scores explain 17 percent of the variation in quintiles of predicted value added. The second column presents the selectivity regression. It shows that households' scores are substantially more accurate in this case: households' peer quality scores explain 33 percent of the variation in quintiles of prior-year selectivity. The columns that distinguish by student achievement reveal that households with high-achieving children have more accurate scores than those with low-achieving children. Nonetheless, both types are better at evaluating selectivity than value added. For instance, for households with high-achieving children, peer quality scores explain almost 40 percent of the variation in quintiles of prior-year selectivity; meanwhile, their value added scores explain only 20 percent of the variation in quintiles of predicted value added.

These results are highly robust. Tables A11-A13 (pages 59-60) present alternative versions of regression (3) that address various potential concerns. First, it is possible that the results in Table 8 are impacted by the fact that most households score only a subset of the tracks in their towns. Table A11 (page 59) restricts the sample to the 21 percent of households with no missing scores, with similar results. Second, it may be that households gather information only on tracks that their child is likely to be eligible for. In this case, the results in Table 8 would average over accurate scores for tracks that are plausibly feasible and inaccurate ones for tracks that are out of reach. Table A12 restricts the sample to tracks that a student would have been eligible to attend in the prior year, again producing similar results. Finally, it may be that households had not yet studied their options when the baseline survey took place. To investigate this, we restrict the sample to the 43 percent of households that reported already being "very certain" of their preference rankings during the baseline survey. The results, in Table A13, are still similar.[43]

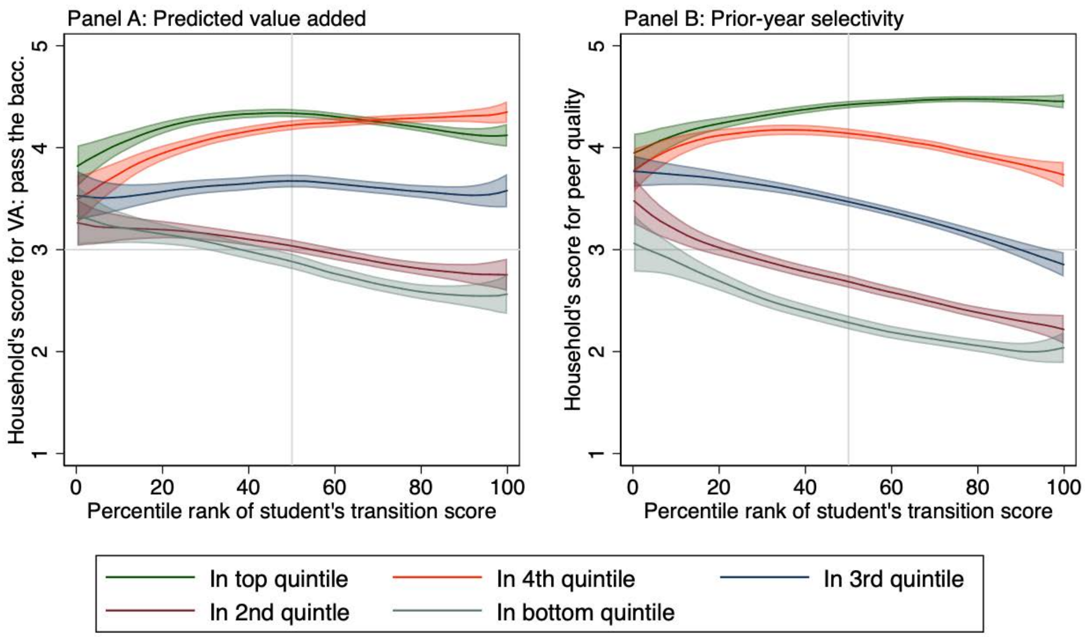
We next explore whether households are better informed about certain types of tracks. First, we study how the accuracy of beliefs depends on a track's position within the town's distribution of value added or selectivity. Figure 3 (page 26) shows how households' quality scores vary with a track's within-town quintile, and how these relationships depend on the achievement level of a household's child. Specifically, for tracks in each within-town quintile of value added or selectivity, we plot separate local linear regressions of households' quality scores (for value added and peer quality, respectively) on the percentile rank of a student's transition score. If households' scores were fully accurate, each of the curves in Figure 3 would be a horizontal line at the value of the given quintile; if scores were random, each of the curves would be a horizontal line at a value of 3.

Figure 3 confirms key takeaways from Table 8. Namely, households' peer quality scores are more accurate in explaining selectivity (Panel B) than are their value added scores in explaining value added (Panel A). Similarly, households with high-achieving children are better informed than those with low-achieving children; however, the difference is particularly pronounced for selectivity.[44] The figure also provides insight into the source of the inaccuracies in households'

43. We consider two additional potential concerns. First, households may assign scores based on the national rather than within-town distribution of tracks. In results not shown, we estimate (3) using quintiles of true values calculated using the national distribution. We find that R-squared values are slightly lower than those in Table 8. Second, the R-squared statistics we provide may be misleading. These are R-squared in terms of explaining predicted value added, V_{jt}^{P}. However, we are ultimately interested in R-squared in terms of explaining true value added, V_{jt}. To investigate this distinction, we run a version of regression (3) that uses values of track characteristics in standard deviation units, rather than within-town quintiles. For this alternative parameterization, we can calculate R-squared in terms of explaining V_{jt}. We do this by adjusting the R-squared for V_{jt}^{P} for forecast error (Appendix A1.4 describes the procedure). The results are in Table A14 (page 60). They show that R-squared for true value added, V_{jt}, is similar to but slightly lower than that for predicted value added, V_{jt}^{P}.

44. For instance, for households with children at the top percentile of the transition score distribution, peer

Figure 3: Mean quality scores by quintile of the track's true values



The figure displays mean household scores for tracks in different within-town quintiles of predicted value added, V^P_{jt}, or prior-year selectivity, MTS_{jt-1}. Specifically, for tracks in each quintile of value added or selectivity, it presents separate local linear regressions of households' quality scores on the percentile rank of a student's transition score. "In top quintile" refers to the tracks in the highest within-town quintile, and "In bottom quintile" indicates the tracks in the lowest within-town quintile. The sample drops student-track observations if the survey respondent did not score the track on both value added and peer quality. It also excludes observations for 43 tracks that were newly created in 2019.

beliefs. For value added, households on average are able to distinguish three groups of tracks: those in the top two-fifths of the town's distribution, those in the middle fifth, and those in the bottom two-fifths. However, they cannot differentiate any further.[45] By contrast, for selectivity, households' mean scores increase almost monotonically with a track's quintile.

Finally, we explore whether beliefs are more accurate for the tracks that households rank highly. If households spend more time researching these tracks, they may score them more accurately. However, they may also assign inappropriately high scores to these tracks in order to rationalize their preference for them. Table 9 (page 27) presents the results. Its first three columns refer to all tracks, and the rest to the two tracks the household ranked the highest in the baseline survey.[46]

Panel A explores accuracy, displaying the mean absolute difference between a household's score and the within-town quintile of a track's true value. It reveals that households' scores for their most-preferred tracks are only slightly more accurate than their scores for all tracks. Among all students and over all tracks, households' value added scores have a mean absolute error of

quality scores are nearly accurate for all tracks except those in the bottom quintile of selectivity.

45. In fact, households with the highest-achieving children mistakenly believe that tracks in the fourth quintile have higher value added than those in the top quintile.

46. As noted in Appendix A4, over 95 percent of households believe their child will attend one of these tracks.

Table 9: The accuracy of households' quality scores

	All tracks			**Top two most-preferred tracks**		
	All students	Low-achieving	High-achieving	All students	Low-achieving	High-achieving
Panel A: Mean absolute difference						
Value added: s_{ij}^{V} v. quint(V_{jt}^{P})	1.13	1.19	1.09	0.99	1.06	0.95
Selectivity: s_{ij}^{PQ} v. quint(MTS_{jt-1})	0.90	1.01	0.84	0.78	1.06	0.62
Panel B: Mean difference						
Value added: s_{ij}^{V} v. quint(V_{jt}^{P})	0.35	0.45	0.29	0.63	0.64	0.63
Selectivity: s_{ij}^{PQ} v. quint(MTS_{jt-1})	0.17	0.36	0.06	0.35	0.65	0.17
Students	2,370	883	1,487	2,283	837	1,446
Student-tracks	17,460	6,433	11,027	3,900	1,420	2,480

The table presents statistics on the accuracy of households' quality scores. Panel *A* reveals the mean absolute difference between a household's quality score and the within-town quintile of the associated track characteristic. Panel *B* exhibits the mean difference between these quantities. "Low-achieving" ("High-achieving") students are those with transition scores in the bottom (top) half of the national distribution. "'Top two most-preferred tracks" are the two tracks that the household ranked the highest in the baseline preference ranking. The sample drops student-track observations if the survey respondent did not score the track on both value added and peer quality. It also excludes observations for 43 tracks that were newly created in 2019.

1.1 quintiles; for their top two most-preferred tracks, households' scores are off by an average of 1 quintile. For selectivity, the corresponding values are 0.9 and 0.8 quintiles. Broadly speaking, these patterns do not depend on whether a household has a low- or high-achieving child.[47]

Panel B explores bias, displaying the mean difference between a household's score and the within-town quintile of a track's true value. It indicates that households overestimate the quality of their most-preferred tracks. Among all tracks, households' value added scores are biased upwards by an average of 0.4 quintiles. By contrast, for their most-preferred tracks, households' value added scores overestimate by 0.6. This again does not depend on their student's achievement level. On the other hand, results for peer quality do vary by student achievement. For households with low-achieving students, peer quality scores are too high by 0.4 quintiles among all tracks and 0.7 quintiles among the top two most-preferred tracks. For households with high-achieving students, these values are only 0.1 and 0.2 quintiles.

In short, households are only partially informed about track value added. Interestingly, many households are also imperfectly aware of track selectivity, despite access to information on this trait. Finally, households are only slightly more knowledgeable about their most-preferred tracks than about the others, and they seem to over-estimate the quality of these tracks.

3 Does providing information change households' choices?

The previous sections suggest that households may lack information on track value added. In particular, households leave value added on the table when choosing tracks, and have beliefs on this attribute that are only partially accurate. In this section, we use our experiment to test whether informing households about value added can influence their track choices. This could

47. The only exception is that for households with low-achieving students, peer quality scores are actually less accurate for their most-preferred tracks.

occur if information causes households to update their beliefs (e.g., the mean or precision), or if it alters their preferences over track characteristics (e.g., by making value added more salient).

We begin by presenting our main reduced-form treatment effect, revealing whether providing information caused students to attend tracks with higher value added. We then make use of the follow-up survey and examine effects on beliefs and track preference rankings. We use these results to explain the heterogeneity we find. Finally, we investigate tradeoffs, asking whether students were induced to give up other track attributes in exchange for value added.

3.1 Effects on the value added of students' tracks

Our main outcome of interest is the value added of the track that a student attends. This outcome affects students' learning during high school and their chances of progressing to higher education. In order to calculate the treatment effect on this outcome, we estimate:

$$\mathrm{sd}(\mathrm{V}^{P}_{j_i^* t}) = \eta_0 + \eta_1 \cdot T_i + {\eta_X}' \cdot X_i + \eta_i. \tag{4}$$

Here, $\mathrm{sd}(\mathrm{V}^{P}_{j_i^* t})$ is the predicted value added of the track of student i in standard deviation units, T_i is an indicator for whether i is in the treatment group, and X_i is a vector of i's covariates.[48] The coefficient of interest is η_1. It captures the average treatment effect of providing information on the value added of the track a student attends.

Table 10: Average treatment effects on the value added of students' tracks

	All students	**Low-achieving**	**High-achieving**
Treated	0.048*	0.121**	-0.002
	(0.025)	(0.049)	(0.023)
Effect in percentage points	0.56	1.43	-0.02
Predicted pass rate	62.9	29.2	83.2
Clusters	78	78	77
Students	2,692	1,012	1,680

The table presents results from regression (4). "Effect in percentage points" is the magnitude of the treatment effect in terms of the probability that a student passes the baccalaureate exam. This is calculated by multiplying the effect in standard deviation units by the 2019 standard deviation of true value added, V_{jt}. The "Predicted pass rate" is the share of students in the regression sample who are predicted to pass. This is calculated in two steps. First, we predict the probability of passing for each student by calculating the share of students with the same transition score percentile rank who passed in the 2004-2014 admission cohorts. Second, we average these values over the students in the regression sample. Low-achieving (high-achieving) students are those with transition scores in the bottom (top) half of the national distribution. Standard errors are clustered by the middle school treatment-control pairs within which we conducted the randomization.

Table 10 (page 28) presents the results. The first column is for all students and shows that, over the full sample, providing information had only a modest effect. It caused students to attend

48. Due to the randomization, covariates are not necessary for identification but may increase precision. In our main specification, X_i includes (i) an indicator for whether the student ranked a feasible track in the baseline survey and (ii) the value added of the track to which the student would have been assigned based on their baseline preference ranking. This latter covariate is calculated as the value added of the feasible track that the student ranked highest in the baseline survey. It is set to zero if the student did not rank any feasible tracks in this survey.

tracks with value added that was higher by an average of 0.05 standard deviations (significant at a 10 percent confidence level). This amounts to an increase in the probability of passing the baccalaureate exam of 0.56 percentage points, which is small relative to the 63 percent predicted pass rate. The second and third columns differentiate by student achievement and reveal that the main effect is driven entirely by low-achieving students—those with transition scores in the bottom half of the national distribution. For these students, the treatment effect is 0.12 standard deviations (significant at a 5 percent level). This is a 1.4 percentage point increase in the probability of passing, as compared to a 29 percent predicted pass rate. By contrast, for high-achieving students, the treatment effect is virtually zero and statistically insignificant.

The results in Table 10 are robust. Table A15 (page 61) presents results from alternative versions of regression (4), including some that control for different sets of covariates and others that employ a difference-in-difference design. The treatment effects are not sensitive to these perturbations. We also assess whether the treatment effects are confounded by spillovers. Namely, it is possible that treated households shared information with others in the town, including, some in the control group. If so, then the treatment effects would be biased toward zero. Appendix A5 (page 82) tests for spillovers by examining whether treatment effects are larger in towns in which we visited a smaller fraction of middle schools. We find no evidence that they are.

We next investigate whether treatment effects vary based on whether a student was eligible for the tracks they preferred in the baseline survey. As noted in Appendix A4 (page 77), over 95 percent of households expected their child to be admitted to at least one of the two tracks that they ranked highest in the baseline. Thus, it is possible that households were more willing to change their choices over the other tracks, given that they did not expect those choices to be relevant for their track assignments. Under this scenario, treatment effects would be larger for students who did not end up being eligible for their two top baseline choices and smaller for those who did. Importantly, almost a quarter of students fall into the former group.

Table 11: Effects on value added by eligibility for the tracks preferred in the baseline

	Eligible for x^{th} most-preferred track in the baseline					
	Most-preferred	2nd-most-preferred	$\geq$ 3rd-most-preferred	$\geq$ 4th-most-preferred	$\geq$ 5th-most-preferred	$\geq$ 6th-most-preferred
Treated	0.019	-0.072	0.184***	0.173**	0.171**	0.190**
	(0.018)	(0.102)	(0.065)	(0.066)	(0.075)	(0.084)
Effect in percentage points	0.22	-0.85	2.17	2.04	2.02	2.25
Predicted pass rate	75.6	51.7	32.8	30.5	29.0	28.3
Clusters	77	72	76	75	73	71
Students	1,766	288	638	507	427	375

The table presents results from regression (4) for subsets of students by eligibility for the tracks that students ranked highly in the baseline survey. "Most-preferred" is the set of students who were eligible for their most-preferred baseline track. "2nd-most-preferred" is the set who were eligible for the track that they ranked second highest in the baseline, but not for the track they ranked highest. "$\geq$ 3rd-most-preferred" is the set who were not eligible for either of their two most-preferred tracks, and so on. See the notes to Table 10 for additional details.

Table 11 (page 29) presents the results, showing treatment effects for subsets of students who differ in eligibility for their top baseline choices. The first column refers to students who scored above the admissions cutoff for their most-preferred baseline choice. The second column is for students who scored above the cutoff for their second-most-preferred track in the baseline, but not for their top choice. The remaining columns are for students who were eligible for only their third-most-preferred baseline choice or worse, their fourth-most-preferred baseline choice or worse, etc. The results corroborate the story in the previous paragraph. The first two columns show that treatment effects are statistically insignificant and close to zero for students who were eligible for one of their two top baseline choices. By contrast, the remaining columns reveal that treatment effects are large and statistically significant for the rest. Depending on the sample, the effects for these students range from 0.17 to 0.19 standard deviations. These effects translate into increases in the probability of passing the exam of 2-2.25 percentage points, which are substantial relative to predicted pass rates of 28-33 percent.

Table 12: Treatment effects by students' characteristics and eligibility for top baseline choices

	All	**Achievement**		**Gender**		**Mother's schooling**	
		Low	High	Female	Male	$\leq$ 12 years	> 12 years
Panel A: Eligible for at least one of two top baseline choices							
Treated	0.007	0.035	0.000	0.001	0.013	0.021	-0.004
	(0.024)	(0.058)	(0.022)	(0.024)	(0.038)	(0.032)	(0.028)
Effect in percentage points	0.09	0.41	0.00	0.02	0.15	0.25	-0.05
Predicted pass rate	72.3	33.7	84.0	75.3	68.6	63.4	80.3
Clusters	78	72	77	78	77	78	77
Students	2,054	479	1,575	1,120	934	981	1,073
Panel B: Ineligible for two top baseline choices							
Treated	0.184***	0.204***	-0.023	0.193**	0.180*	0.154*	0.221*
	(0.065)	(0.069)	(0.123)	(0.096)	(0.105)	(0.082)	(0.127)
Effect in percentage points	2.17	2.40	-0.27	2.28	2.12	1.81	2.61
Predicted pass rate	32.8	25.1	72.1	32.7	32.9	28.0	42.7
Clusters	76	76	28	71	67	75	64
Students	638	533	105	306	332	430	208

The table presents results from regression (4) for subsets of students. The subsets represent the interaction between student characteristics (achievement, gender, or mother's schooling) and whether the student was eligible for at least one of the two tracks s/he listed as most preferred in the baseline survey. Low-achieving (high-achieving) students are those with transition scores in the bottom (top) half of the national distribution. See the notes to Table 10 for additional details on the regressions.

Finally, we examine how this heterogeneity interacts with student characteristics. We estimate regression (4) for different types of students, always distinguishing between those who were eligible for their two top baseline choices and those who were not. The results are in Table 12 (page 30). Panel A refers to students who did gain admission to their top baseline choices. For these individuals, treatment effects are small and statistically insignificant regardless of achievement, gender, or mother's schooling. Panel B concerns the other group—those rejected by their top baseline choices. Even within this group, treatment effects are driven by low-achieving students. For them, the effect is 0.2 standard deviations (significant at a 1 percent level), an increase in the probability of passing the exam of 2.4 percentage points over a predicted pass rate of 25 percent. Meanwhile, treatment effects for high-achieving students (Panel B) are close to zero and statistically insignificant. The remaining columns in the panel indicate that treatment effects do not vary by a student's gender or by his/her mother's schooling.

In short, this subsection shows that the treatment had heterogeneous effects. It had little impact for high-achieving students or for low-achieving students admitted to one of their top baseline choices. By contrast, for low-achieving students who were rejected by these choices, the treatment had large effects. It caused these students to attend tracks with value added that was higher by 0.2 standard deviations. In the next two subsections, we probe these patterns further by using the follow-up survey to investigate effects on beliefs and track preference rankings.

3.2 Effects on beliefs regarding value added

In this subsection, we explore whether the treatment affected households' beliefs regarding track value added. In particular, we assess whether it caused an increase in the accuracy of households' quality scores for this track characteristic.

Before turning to results, we note that there are two ways this analysis of quality scores may understate treatment effects on beliefs. First, it is possible for information to influence the precision of households' beliefs without changing their quality scores. Second is an issue of timing. The follow-up survey took place a few weeks after households submitted their track preference rankings. It is possible that households had turned their attention to other topics by this point and had forgotten some of what they knew when they were deciding their track preferences. This lag may have introduced measurement error in the follow-up quality scores. While acknowledging these caveats, we believe that treatment effects on quality scores are a useful proxy—and a possible lower bound—for those on beliefs.

To conduct the analysis, we estimate:

$$|\text{quint}(\text{V}^P_{jt}) - s^V_{ij,\text{fs}}| = \eta_0 + \eta_1 \cdot T_i + {\eta_X}' \cdot X_{ij} + \eta_{ij}. \tag{5}$$

Here, $|\text{quint}(\text{V}^P_{jt}) - s^V_{ij,\text{fs}}|$ is the absolute difference between: i) the predicted value added of track j in units of within-town quintiles, $\text{quint}(\text{V}^P_{jt})$, and ii) household i's quality score for the track's value added from the follow-up survey, $s^V_{ij,\text{fs}}$. As in regression (4), the coefficient of interest is η_1. It represents the average treatment effect on the absolute error of households' quality scores. If the treatment caused households' scores to become more accurate, then η_1 will be negative.

Table 13 (page 32) presents the results. The first column provides results for all tracks, while the others distinguish by a track's position in a household's baseline preference ranking. Specifically, the second and third columns are for the tracks that households listed, respectively, as most- and second-most-preferred. The fourth through seventh columns are for the tracks that households listed as third-most-preferred or worse, fourth-most-preferred or worse, etc.

Table 13: Effects on the accuracy of households' value added scores

	x^{th} most-preferred track in the baseline						
	All tracks	Most-preferred	2nd-most-preferred	≥ 3rd-most-preferred	≥ 4th-most-preferred	≥ 5th-most-preferred	≥ 6th-most-preferred
Treated	-0.055	0.032	-0.033	-0.101**	-0.124**	-0.156**	-0.181***
	(0.034)	(0.041)	(0.053)	(0.045)	(0.053)	(0.060)	(0.063)
Mean abs. difference: baseline	1.02	0.93	1.07	1.06	1.11	1.16	1.18
Mean abs. difference: follow-up	1.00	0.86	1.03	1.06	1.12	1.14	1.15
Clusters	76	76	75	76	76	76	76
Students	1,525	1,263	962	1,352	1,134	967	868
Student-tracks	4,970	1,263	962	2,745	2,100	1,727	1,487

The table presents results from regression (5). The columns provide results for different sets of tracks. "Most-preferred" refers to the track that a household ranked highest in the baseline survey. "2nd-most-preferred" is the track that a household ranked second highest in this survey. "≥ 3rd-most-preferred" are all tracks other than the two that the household ranked highest. The remaining columns are defined analogously. The regressions include indicators for the value of the absolute difference between: i) the within-town quintile of a track's predicted value added, quint(V^P_{jt}), and ii) the household's baseline score for the track on this dimension, s^V_{ij}. In calculating these indicators, we create a separate group for the tracks for which the household did not provide scores in the baseline survey. "Mean abs. difference: baseline" is the mean absolute difference between quint(V^P_{jt}) and s^V_{ij} for the sample. Similarly, "Mean abs. difference: follow-up" is the mean absolute difference between quint(V^P_{jt}) and $s^V_{ij,\text{fs}}$. Standard errors are clustered by the middle school treatment-control pairs within which we conducted the randomization.

The results indicate that the treatment increased the accuracy of households' scores, but only for their less-preferred tracks. For the full set of tracks, the treatment led to a statistically insignificant improvement in accuracy of 0.06 quintiles. This is a small effect relative to the mean inaccuracy of about one quintile.[49] For the tracks that households' ranked highest and second-highest in the baseline survey, treatment effects are even smaller. By contrast, for the remaining tracks, the treatment induced sizable improvements in accuracy. These effects (columns 4-7), are all significant at either the 1 or 5 percent confidence level. The improvements appear to be larger for tracks farther down a household's baseline preference ranking. For instance, for tracks other than a household's two top baseline choices, the improvement is 0.1 quintiles; for tracks other than a household's five top baseline choices, the improvement rises to 0.2 quintiles.

Tables A16 and A17 (page 61) replicate Table 13, examining whether effects differ for households with low- and high-achieving children. The results are similar for both types of households. Providing information improved the accuracy of quality scores only for tracks that households did not prefer in the baseline survey. For these tracks, treatment effects are twice as large for households with low-achieving children as they are for households with high-achieving children. For households with high-achieving children, effects are always of expected sign but small in magnitude. Only one of them is significant at the 10 percent confidence level.

To summarize, this subsection shows that the treatment was only partially successful at influencing beliefs. It impacted households' quality scores for tracks that they did not initially

49. Specifically, the mean absolute difference between households' scores and quint(V^P_{jt}) was 1.02 quintiles in the baseline survey and 1.00 quintiles in the follow-up survey.

prefer. Even for these tracks, it had small effects for households with high-achieving children.

3.3 Effects on track preference rankings

We next analyze effects on households' track preference rankings. We examine whether information caused households to assign higher preference ranks to tracks with higher value added. Specifically, we calculate the treatment effect on the association between the within-town percentile rank of a track's value added and the percentile rank of the track in a household's preference ranking. We estimate:

$$\text{ppr}_{ij,\text{fs}} = (\delta_1 + \delta_2 \cdot T_i) \cdot \text{pr}(\text{V}^P_{jt}) + (\delta_{X,1} + \delta_{X,2} \cdot T_i)' \cdot X_{ij} + \delta_{ij}. \tag{6}$$

In this regression, $\text{ppr}_{ij,\text{fs}}$ is household i's percentile preference rank for track j, as reported in the follow-up survey.[50] It is calculated by dividing a track's rank in the household's preference ranking by the number of tracks in the town. The variable is ordered such that a value of 1 indicates a household's most-preferred track.[51] Next, $\text{pr}(\text{V}^P_{jt})$ is track j's within-town percentile rank of value added. It is calculated by dividing the track's within-town value added rank by the number of tracks in the town. To be consistent with $\text{ppr}_{ij,\text{fs}}$, it is ordered such that a value of 1 indicates the town's best track by value added. X_{ij} is a set of indicators for track j's position in household i's baseline preference ranking. The coefficient of interest is δ_2; it measures the effect of the treatment on the association between value added and preference ranks.[52]

Table 14 (page 34) presents the results from regression (6). These exhibit a pattern that is similar to that for effects on the accuracy of quality scores from Section 3.2. The first column provides results for the full set of tracks. It shows that, overall, providing information caused a modest increase in the association between value added and preference ranks. A one percentile increase in value added is associated with an increase in preference ranks that is higher by 0.05 percentiles for the treatment group than for the control. This difference is significant at a 10 percent confidence level. The remaining columns show that treatment effects exist only for tracks that households did not initially prefer. Namely, the second column reveals that for the two tracks that a household ranked highest in the baseline survey, the treatment effect is insignificant and has the wrong sign. By contrast, the third column shows that after excluding these tracks, the treatment effect rises to 0.06 percentiles and becomes significant at the 1 percent confidence level.

50. At the time of the baseline survey, we told households that we would contact them after the allocation for a follow-up survey. We requested that they save a copy of their official track preference rankings in preparation for this. When we conducted the follow-up, we asked households to find their copy and read off the ranking. We therefore believe that the preference rankings reported in the follow-up reasonably approximate those submitted.

51. We set $\text{ppr}_{ij,\text{fs}}$ equal to 0 for tracks that households do not rank, since students cannot be assigned to these tracks. In particular, if a household ranks all J_i tracks in its town, its least-preferred track has $\text{ppr}_{ij,\text{fs}} = 1/J_i$. Unranked tracks should thus have a value of $\text{ppr}_{ij,\text{fs}}$ that is less than $1/J_i$; we choose 0.

52. By including X_{ij}, δ_1 measures the average slope of conditional-on-X_{ij} best-fit lines between $\text{ppr}_{ij,\text{fs}}$ and V^P_{jt} for households in the control group. $\delta_1 + \delta_2$ measures the average slope of these lines for treated households.

Effects continue to grow for tracks that were farther back in the baseline preference ranking. For instance, for tracks other than a household's five top baseline choices, the effect is 0.07 and significant at a 1 percent confidence level.

Table 14: Effects on the association between value added and households' preference rankings

	x^{th} most-preferred track in the baseline					
	All tracks	Two most-preferred	$\geq$ 3rd-most-preferred	$\geq$ 4th-most-preferred	$\geq$ 5th-most-preferred	$\geq$ 6th-most-preferred
Value added: treated	0.049*	-0.072	0.062***	0.064***	0.068***	0.069***
	(0.026)	(0.103)	(0.023)	(0.023)	(0.023)	(0.023)
Association: baseline	0.434	0.018	0.269	0.179	0.102	0.055
Association: follow-up	0.345	0.067	0.213	0.168	0.149	0.141
Clusters	76	76	76	76	76	76
Students	1,533	1,523	1,533	1,533	1,533	1,514
Student-tracks	20,029	2,937	17,092	15,849	14,779	13,938

The table presents results from regression (6). The values in the row labeled "Value added: treated" are the estimates and standard errors for δ_2. The columns provide results for different sets of tracks. "Two most-preferred" refers to the two tracks that households ranked highest in the baseline survey. "$\geq$ 3rd-most-preferred" are all tracks other than the two that are most preferred. The remaining columns are defined analogously. The regressions include indicators for the interaction between a track's position in a household's baseline preference ranking and whether the household is in the treatment group. In calculating these indicators, we create separate groups for tracks that households left unranked in the baseline survey. "Association: baseline" is the slope coefficient from a regression of the percentile preference rank from the baseline survey, $\text{ppr}_{ij,\text{bs}}$, on $\text{pr}(\text{V}_{jt}^{P})$. "Association: follow-up" is the slope coefficient from a regression of $\text{ppr}_{ij,\text{fs}}$ on $\text{pr}(\text{V}_{jt}^{P})$. Standard errors are clustered by the middle school treatment-control pairs within which we conducted the randomization.

Tables A18 and A19 (page 62) replicate Table 14 for households with low- and high-achieving children. These again display patterns similar to those in Section 3.2. For both groups of households, treatment effects are insignificant for the two tracks that households ranked highest in the baseline survey. For the remaining tracks, effects are small and mostly insignificant for households with high-achieving children; meanwhile, they are sizable and significant at a 1 percent confidence level for households with low-achieving children.

Thus, the story with regard to treatment effects on preference rankings matches that with respect to effects on beliefs. Providing information caused households to re-order their preference rankings only for tracks other than their two top baseline choices. In addition, these effects were small for households with high-achieving children.

3.4 Interpreting the effects on the value added of students' tracks

The results for beliefs and preference rankings help to explain the heterogeneity in treatment effects on the value added of students' tracks. For high-achieving students, providing information had little impact on beliefs or preference rankings. As a result, it did not cause these students to attend tracks with higher value added. For low-achieving students, the information did affect beliefs and preference rankings, but only for tracks that were initially less preferred. Thus, for this group, treatment effects differ depending on whether a student was eligible for his or her top baseline choices. The treatment had little influence on track assignments for students who were eligible for one of these choices, while it had significant effects for students who were not.

As mentioned in Section 3.1, the fact that households were more receptive to information for tracks other than their top baseline choices is likely a consequence of the approach that they used to rank tracks. The tracks that households ranked highest were ones that they thought would be feasible and that they thus expected their child to attend.[53] It may be that households were less attached to their beliefs and preference rankings for the other tracks because they did not think those tracks would be relevant.

A separate question is why responses were larger for households with low-achieving children than for those with high-achieving children. One possible explanation is that households with high-achieving children may have been more certain of their beliefs and preference rankings at the time of the baseline survey. If so, our intervention may have come too late in their decision-making process. To assess this story, we calculate the share of each type of household reporting (in the baseline survey) that they were uncertain, somewhat certain, or very certain of their preference ranking. The results are in Table A20 (page 62). They reveal that households with high-achieving children were relatively more certain, but not by a large amount. Namely, for this group, 48 percent reported being very certain and 6 percent reported being uncertain; for households with low-achieving children, these percentages are 36 and 11, respectively.

A second candidate explanation is that households with low-achieving children may be more trusting of information provided by outside authority figures. If this were the case, then treatment effects would be larger for low-achieving students even after conditioning on certainty. Table A21 (page 63) presents results for these effects. It displays effects on beliefs and preference rankings for subsets of students defined by the interaction of achievement and certainty.[54] The results reveal two points. First, effects are larger for households who reported being uncertain or somewhat certain than for those who were very certain. Second, even within these groups, effects are larger for low-achieving students than for high-achieving ones.

Thus, the evidence suggests that both stories play a role in explaining why households with low-achieving children were more receptive to the information. These households were less likely to have settled on their beliefs and preference rankings when the intervention occurred. In addition, they exhibited larger responses conditional on their degree of certainty.

3.5 Effects on other characteristics of students' tracks

We conclude this section by examining whether the treatment affected other characteristics of students' tracks, in addition to value added. We are interested in whether households made tradeoffs in order to attend higher-value added tracks.

To do this, we re-estimate (4) using a variety of track characteristics as outcomes. Table 15

53. Recall that more than 95 percent of households expected their child to be admitted to at least one of their two top baseline choices (Appendix A4, page 77).

54. In Table A21, the sample of tracks excludes a household's two top baseline choices. This is because the intervention generated no response for these choices, as discussed previously.

Table 15: Effects on the characteristics of students' tracks

	Value added	**Selectivity**	**Location quality**	**Curricular focus**		
				Humanities	Math & science	Technical
Panel A: Eligible for at least one of two top baseline choices						
Treated	0.007	0.009	0.015	-0.013	0.012	0.001
	(0.024)	(0.017)	(0.020)	(0.017)	(0.016)	(0.010)
Clusters	78	78	78	78	78	78
Students	2,054	2,054	1,978	2,054	2,054	2,054
Panel B: Ineligible for two top baseline choices						
Treated	0.184***	0.006	0.073	0.039	0.030	-0.069**
	(0.065)	(0.050)	(0.086)	(0.030)	(0.027)	(0.033)
Clusters	76	76	76	76	76	76
Students	638	638	492	638	638	638

The table presents results from regression (4) for a variety of outcome variables. The outcomes in columns 1 and 2 are, respectively, predicted value added and minimum transition score. These are both in standard deviation units. The outcome in column 3 is a household's baseline score for a track's location quality. Finally, the outcomes in columns 4-5 relate to a track's curricular focus. Regressions control for values of the outcome variable for the track that a household would have been assigned to based on their baseline preference ranking. This is the feasible track that the household ranked the highest in the baseline survey. The regressions include indicators for students who did not rank any feasible tracks in this survey. Standard errors are clustered by the middle school treatment-control pairs within which we conducted the randomization.

(page 36) presents the results. Column 1 is for the valued added of a student's track, replicating results in Section 3.1. Column 2 is for the track's selectivity, MTS_{jt}, and column 3 is for the household's baseline quality score for the track in terms of location, s^L_{ij}. Columns 4-6 capture the track's curricular focus—humanities, math and science, or technical subjects. Panel A is for students who were eligible for one of their two top baseline choices; Panel B is for students who were not.

Panel A reveals that the treatment had no impact on any track characteristic for students who were admitted to one of their top baseline choices. Panel B indicates that the treatment also had limited effect on characteristics other than value added for students who were rejected by their top choices. In particular, the treatment caused these students to attend tracks with 0.2 standard deviations worth of additional value added. Yet, it did not affect the selectivity or location quality of their tracks. The only tradeoff that students were forced to make had to do with a track's curricular focus. The treatment caused students in Panel B to be 7 percentage points less likely to attend a technical track. The reason for this is that, conditional on selectivity, technical tracks tend to have lower value added than humanities or math and science tracks.[55]

4 How do preferences for other track characteristics affect demand for value added?

We next explore how demand for value added is constrained by preferences for other track characteristics. While the analysis in this section is non-experimental, it helps to interpret the causal results from Section 3.

55. The relationship between value added and selectivity by curricular focus is in Figure A5 (page 51).

We study two questions. First, we attempt to disentangle the relative contributions of preferences and beliefs in causing households to leave value added on the table. Specifically, we ask, holding fixed households' preferences for track characteristics, how would their track choices change if they had fully accurate beliefs about value added? Second, we assess how strongly held are households' preferences. In particular, we examine whether providing information on value added causes households to care more about this track characteristic—such as by signaling its importance. If so, then providing information could have larger effects than would be predicted via its impact on beliefs alone.

We investigate these questions in three steps. First, we estimate households' preferences for track characteristics. Second, we predict households' track choices under accurate beliefs. Third, we gauge to what extent the experimental treatment effects operate through impacts on preferences rather than beliefs.

4.1 Estimating preferences for track characteristics

We start by estimating households' preferences for track characteristics. We do so by using their baseline track quality scores to explain their baseline preference rankings.

Specifically, we write household i's utility from track j, U_{ij}, as a linearly separable function of its scores for the track, s_{ij}^q, on quality dimensions q:

$$U_{ij} = \sum_q \beta_q \cdot s_{ij}^q + \epsilon_{ij}. \tag{7}$$

β_q is the increase in utility associated with a one-unit increase in a household's quality score, and ϵ_{ij} is the idiosyncratic component of utility that is not explained by the household's scores. Importantly, households' quality scores represent their *beliefs* about track characteristics, not the *true values* of these characteristics. Thus, by using quality scores, the β_q coefficients directly measure preferences. If, instead, we were to use true values, then the β_q coefficients would reflect a combination of preferences and the correlation between beliefs and true values.[56]

We assume that ϵ_{ij} is independent and follows a Type-1 Extreme Value distribution, and fit equation (7) to households' preference rankings using a rank-ordered logit.[57] Table A22

56. There are two ways in which equation (7) may fail to recover preferences. First, there may be an omitted quality dimension that causes ϵ_{ij} to be correlated with the s_{ij}^q covariates. In this case, the β_q coefficients will capture a combination of the effects of the included and omitted track characteristics. We do not think this is a major issue because we have quality scores on a large number of relevant dimensions. Second, the β_q coefficients will also be biased if utility depends on the *precision* of beliefs. For instance, if a household's utility is strictly concave in the score for a track's value added, then the household will gain more utility from a track that it knows has a score of 4 than from one that it believes has a score of either 3 or 5 with equal probability. We ignore this issue because we do not have data on the precision of households' beliefs.

57. A rank-ordered logit is a series of multinomial logits corresponding to each choice in a preference ranking. In practice, we do not use all the constituent multinomial logits in estimation. Instead, we use just those for a household's two top choices. That is, we maximize the probability that a household prefers track r_{i1} to all other tracks in the town times the probability that the household prefers track r_{i2} to all other tracks except r_{i1}.

(page 64) presents the results. The coefficients indicate that households care about a variety of track characteristics. Column 1 presents a specification that includes quality scores for a track's location, whether it is used by a student's siblings and friends, its peer quality, its curriculum, and its value added on passing the baccalaureate exam. All coefficients are statistically significant. Four are similar in magnitude: those for location (0.28), siblings and friends (0.37), peer quality (0.34), and value added (0.34). However, these are only a third the size of the coefficient for curriculum (0.93).

Next, Columns 2-4 substitute scores for alternative dimensions of value added: value added on attending a good college, value added on labor market wages, and teacher quality. Column 5 presents the horse race, including scores for all value added dimensions at once. The results suggest that households care about each type of value added, but are especially interested in value added on college quality and wages.[58] Throughout all specifications, the coefficients for the other track characteristics remain stable.

Finally, columns 6 and 7 supplement the quality scores by controlling for the true values of a few track characteristics that households can easily observe. Column 6 adds controls for a track's curricular focus, interacted with the student's gender.[59] Column 7 adds the absolute difference between the track's selectivity and the student's transition score.[60] These additional covariates generate a slight improvement in explanatory power. However, they cause little change in the coefficients for households' quality scores.

Next, we explore whether there is heterogeneity in preferences based on a student's academic achievement. Tables A23 and A24 (pages 65 and 66, respectively) replicate Table A22 for low- and high-achieving students. The results for the two groups are mostly similar, but there are a few notable differences. First, the coefficients on peer quality are large and statistically significant for high-achieving students, but small and insignificant for low-achieving ones. Next, high-achieving students like math and science tracks the most, while low-achieving students like these tracks the least. Finally, the models have different degrees of explanatory power for the two groups.

Specifically, this likelihood is:

$$\Pr[r_{i1}, r_{i2} | \mathcal{J}_i, \{s^q_{ij}\}_{q,j}] = \prod_{l=1}^{2} \frac{\exp[\sum_q \beta_q \cdot s^q_{ir_{il}}]}{\sum_{k \in \mathcal{J}_i \setminus \{r_{im}: m<l\}} \exp[\sum_q \beta_q \cdot s^q_{ik}]}. \tag{8}$$

We maximize the sum of the log of (8) over the set of survey households. We focus on only the first two choices because most households appear to have thought carefully about these by the time of the baseline survey. As we discuss later, our main conclusions are not sensitive to how many choices we include.

58. For instance, the horse race puts equal and statistically significant coefficients on value added on college quality and wages and small and statistically insignificant coefficients on the other value added dimensions.

59. Households can observe a track's curricular focus just by reading the track's name. Thus, it is reasonable to assume that they have correct beliefs about this track characteristic.

60. We include this covariate in an attempt to capture the phenomenon that, for their top choices, households tend to choose some of the more selective tracks among those that they think will be feasible. We have shown previously that households' beliefs about track selectivity and about their own child's transition score are not fully accurate. Nonetheless, households still seem to have a strong sense of this covariate.

Depending on the specification, we explain about 40 percent of the variation in track choices for high-achieving students, and only about 20 percent for low-achieving ones.

Finally, one concern in estimating equation (7) is that few households provide quality scores for all the tracks in their towns. Missing scores could introduce bias if households' propensity to score tracks depends on how much they like them. Table A25 (page 67) gauges the impact of missing scores. It compares the results in Tables A22-A24 with those for two alternative specifications: one in which we use only the subset of households without missing scores and another in which we impute the missing scores using a random forest.[61] The results reveal that the coefficients are similar across specifications.

4.2 How would track choices change under accurate beliefs?

We next examine to what extent households' track choices are distorted by the inaccuracy of their beliefs. To do this, we use the preference model to predict the value added of the tracks students would attend under alternative sets of beliefs. By comparing these predictions, we can calculate how much of the value added that households leave on the table is due to beliefs and how much is due to caring about other features of tracks.[62]

Specifically, we consider three sets of beliefs. The first set, "baseline scores" (BS), captures beliefs at the time of the baseline survey; it simply uses household's baseline quality scores.[63] The second and third sets approximate beliefs at the time that households submit their preference rankings. "Correct peer" (CP) reflects the information available to households in the control group. It uses baseline scores for all quality dimensions except peer quality. For peer quality, it substitutes the within-town quintile of a track's selectivity. "Correct peer and pass" (CPP) reflects the information available to households in the treatment group. It replaces scores for both peer quality and value added on passing the baccalaureate exam with the corresponding within-town quintiles of these track characteristics.

We predict the value added of students' tracks by combining the preference estimates from equation (7) with a given set of quality scores. In particular, we calculate a weighted average of the value added of each track in a household's feasible set. The weights are equal to the

61. For each quality dimension, we predict a household's score for a track using covariates including: i) characteristics of the track, ii) characteristics of the student, and iii) quality scores for the track from other households in the same town or middle school as the student. We replace missing scores with these predictions.

62. Importantly, the analysis in this section holds constant households' feasible choice sets. Thus, it disentangles the roles of preferences and information in driving choice behavior in the *current* choice setting. If households were somehow made to have correct beliefs, the choice setting would change due to dynamic effects on track selectivity, teacher sorting, value added, etc. We leave these dynamic effects for future work.

63. For missing quality scores, we substitute the predictions from the random forest.

probability that the household prefers the track, given the scores. For "Baseline scores", this is:

$$V_{i,\mathrm{BS}} \equiv \sum_{j \in \mathcal{J}_i^e} \mathrm{sd}(V_{jt}^P) \cdot \frac{\exp[\sum_q \hat{\beta}_q \cdot s_{ij}^q]}{\sum_{k \in \mathcal{J}_i^e} \exp[\sum_q \hat{\beta}_q \cdot s_{ik}^q]}.$$

The formulas for "Correct peer", $V_{i,\mathrm{CP}}$, and "Correct peer and pass", $V_{i,\mathrm{CPP}}$, are similar, but with the alternative sets of scores.

Before turning to results, we validate our approach comparing our predictions with households' observed choices. Specifically, we compare $V_{i,\mathrm{BS}}$ with the value added of the track a student would attend under the household's baseline preference ranking.[64] For households in the control group, we compare $V_{i,\mathrm{CP}}$ with the value added of students' actual tracks.[65] We show these comparisons for three versions of the preference model. Our preferred specification, "Preferred," uses the model from Column 7 of Table A22. "No transition score distance" drops the absolute difference between the track's selectivity and the student's transition score (using the model from Column 6 of Table A22). Finally, "Just quality scores" drops all the true track characteristics (using the model from Column 1 of the table).[66] Figure A7 (page 53) presents the comparisons. It reveals that $V_{i,\mathrm{BS}}$ and $V_{i,\mathrm{CP}}$ closely match the value added of the tracks that households choose. The fit is considerably better for our preferred specification than for the other two.[67] As a result, we use this specification in the main results.

The main results are in Figure 4 (page 41) and Table 16 (page 42). Figure 4 is similar to Figure 2 in Section 1.2. It displays the value added of students' tracks in relation to the options in their feasible choice sets. Further, it shows how these patterns vary based on a student's academic achievement.[68] Table 16 provides the corresponding summary statistics.

The results indicate that inaccurate beliefs play only a limited role in accounting for why households leave value added on the table. First, correcting households' beliefs about peer

64. This is the value added of the feasible track that the student ranked the highest in the baseline survey.

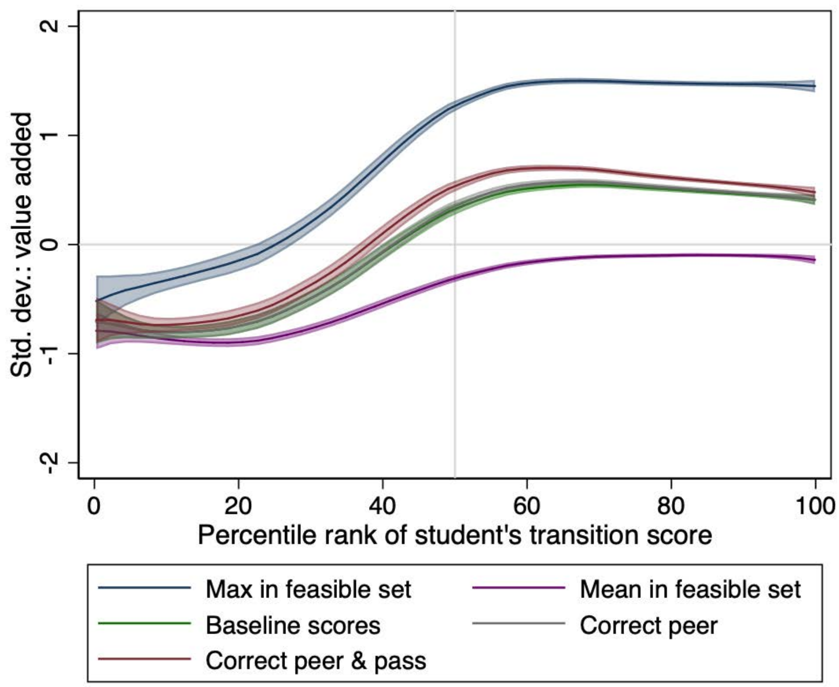
65. The analogous exercise for households in the treatment group would compare $V_{i,\mathrm{CPP}}$ with the value added of students' tracks. However, this would be inappropriate because the treatment did not fully influence households' beliefs about value added. For instance, it had no effect on beliefs regarding the two tracks that households ranked the highest in the baseline survey.

66. In reality, the specifications differ in two ways from those in Table A22. First, we allow coefficients to vary based on whether a student is low- or high-achieving, as in Tables A23 and A24. Second, we estimate the models after imputing missing quality scores with the predictions from the random forest, as in Table A25. Thus, the coefficients for our preferred specification are those from columns 6 and 9 of Table A25.

67. For instance, for $V_{i,\mathrm{BS}}$, for our preferred specification, the line of best fit has a slope coefficient of 0.98 and an R-squared of 0.61. For "Just quality scores", the slope and R-squared are, respectively, 0.91 and 0.55.

68. In particular, Figure 4 presents local linear regressions of a variety of variables on the percentile rank of a student's transition score. The blue line is for the value added of the track with the maximum value added in the student's feasible set, and the purple line is for the mean value added in the set. The green, gray, and brown lines are for our predictions for the value added of students' tracks for the three sets of beliefs. These are, respectively, $V_{i,\mathrm{BS}}$, $V_{i,\mathrm{CP}}$, and $V_{i,\mathrm{CPP}}$. The differences between the green, gray, and brown lines represent the mean change in value added due to changes in beliefs. The difference between these lines and the blue line is the mean potential increase in value added for a given set of beliefs. In the figure, the green line, "Baseline scores", and the gray line, "Correct peer", overlap almost perfectly.

Figure 4: The value added of students' tracks for alternative sets of beliefs



The figure shows the value added of students' tracks under alternative sets of beliefs. Specifically, it plots the relationship between the percentile rank of the student's transition score and the following variables. The blue (purple) line is for the maximum (mean) value of value added in the student's feasible set. The green, gray, and brown lines are for $V_{i,\mathrm{BS}}$, $V_{i,\mathrm{CP}}$, and $V_{i,\mathrm{CPP}}$, respectively. The lines are calculated using local linear regressions. The green and gray lines are nearly identical. The sample is similar to that for the experimental treatment effects from Section 3. However, it excludes 15 students who did not score above the admissions cutoffs for any of the tracks that existed in both 2018 and 2019. (These students were assigned to tracks that were newly created in 2019.)

quality would have almost no effect on the value added of students' tracks. Table 16 shows that the mean difference between $V_{i,\mathrm{CP}}$ and $V_{i,\mathrm{BS}}$ is 0.004 s.d. for low-achieving students and 0.023 s.d. for high-achieving ones. By contrast, correcting beliefs about value added would have a modest effect. The mean difference between $V_{i,\mathrm{CPP}}$ and $V_{i,\mathrm{CP}}$ is 0.131 s.d. for low-achieving students and 0.107 s.d. for high-achieving ones. Thus, even with correct beliefs about both peer quality and value added, households would still leave substantial value added on the table. For low-achieving students, the mean potential increase in value added is 0.713 s.d. for baseline scores (BS), 0.710 s.d. for correct peer (CP), and 0.579 s.d. for correct peer and pass (CPP). Meanwhile, for high-achieving students, the analogous values are 0.983 s.d., 0.960 s.d., and 0.852 s.d. We assume that "Correct peer" represent households' choices in the normal institutional context in which households are not provided with information on value added. Thus, for low-achieving students, inaccurate beliefs about value added account for 18 percent of the value added that is left on the table. For high-achieving students, the corresponding share is 11 percent.[69]

69. Table A26 (page 68) tests the sensitivity of the results to alternative specifications of the preference model. It provides results for "no transition score distance" and "just quality scores". It also provides results for models using different numbers of choices in estimating the rank-ordered logit. For each specification, the table presents the mean difference between $V_{i,\mathrm{CPP}}$ and $V_{i,\mathrm{CP}}$. It reveals that this difference is insensitive to the number of choices

Table 16: The effect of alternative sets of beliefs on the value added of students' tracks

	Students	Change in VA		Potential increase in VA		
		$V_{i,CP} - V_{i,BS}$	$V_{i,CPP} - V_{i,CP}$	$V_{i,BS}$	$V_{i,CP}$	$V_{i,CPP}$
All students	2,677	0.016	0.116	0.882	0.867	0.751
Low-achieving	997	0.004	0.131	0.713	0.710	0.579
High-achieving	1,680	0.023	0.107	0.983	0.960	0.852

The table presents summary statistics on our predictions for the value added of students' tracks under alternative sets of beliefs. "Change in VA" is the mean difference between the listed variables. "Potential increase in VA" is the mean difference between (i) the maximum value of value added in students' feasible sets and (ii) the listed variables. The sample is similar to that for the experimental treatment effects from Section 3. However, it excludes 15 students who did not score above the admissions cutoffs for any of the tracks that existed in both 2018 and 2019. (These students were assigned to tracks that were newly created in 2019.)

The results in Table 16 may seem small given the size of the treatment effects in Section 3. For instance, for the students for whom the treatment had an impact—low-achieving students who were not admitted to their top two baseline choices—the treatment effect was 0.2 s.d. This value is larger than the mean difference between $V_{i,CPP}$ and $V_{i,CP}$ for low-achieving students (0.131 s.d.).[70] A possible explanation for the discrepancy is that the treatment may affect choices via channels other than the accuracy of households' beliefs. In particular, it may also affect the precision of beliefs or households' preferences for track characteristics. These effects would operate through changes in the nature of the preference model itself, rather than in the values of the quality scores. We explore this explanation in the next section.

4.3 Was the treatment effect due to changes in beliefs or preferences?

This section examines whether the treatment effect operated via changes in beliefs or preferences. If the treatment operated through beliefs, then it would cause treated households to update their track preferences in accordance with our preference model. These households would engage in a reasoning process where they, in effect, calculate new values of utility for each track by multiplying the preference coefficients, β_q, by new values of track quality scores. By contrast, if treatment operated via other channels, then households' track preferences would diverge from the predictions of the preference model. Notably, if the treatment caused households to care more about value added, then their track preferences would align with this characteristic more tightly than predicted.

We explore this distinction by running a horse race. We evaluate whether households' final track preference rankings are better explained by predictions from the preference model or by raw values of track value added. Specifically, we do the following. First, we predict preference rankings under "Correct peer and pass." As discussed, this set of quality scores reflects the infor-

used. However, the difference is somewhat sensitive to which covariates are included in the model. Namely, the difference is slightly larger for "no transition score distance" and significantly larger for "just quality scores." Nonetheless, in all specifications, the difference is still quite modest—the largest value is 0.23 s.d.

70. In fact, for low-achieving who were not admitted to their top two baseline choices, the mean difference between $V_{i,CPP}$ and $V_{i,CP}$ is even smaller: 0.094 s.d.

mation available to treated households at the time they submit their final preference rankings.[71] Next, we calculate a percentile preference ranking, $\text{ppr}_{ij,\text{CPP}}$, by dividing our predicted ranking by the number of tracks in the town. Finally, we assess whether the treatment effect on preference rankings is more associated with $\text{ppr}_{ij,\text{CPP}}$ or with the percentile rank of value added, $\text{pr}(\text{V}^P_{jt})$.

Namely, we run a modified version of equation (6) that includes both $\text{ppr}_{ij,\text{CPP}}$ and $\text{pr}(\text{V}^P_{jt})$:

$$\text{ppr}_{ij,\text{fs}} = (\delta_1 + \delta_2 \cdot T_i) \cdot \text{pr}(\text{V}^P_{jt}) + (\delta_3 + \delta_4 \cdot T_i) \cdot \text{ppr}_{ij,\text{CPP}} + (\delta_{X,1} + \delta_{X,2} \cdot T_i)' \cdot X_{ij} + \delta_{ij}. \quad (9)$$

Here, δ_1 and δ_3 capture how control households changed their preference rankings between the baseline and endline surveys. These coefficients will be non-zero if control households learned or reasoned in a way that caused their preference rankings to become more correlated with $\text{ppr}_{ij,\text{CPP}}$ or $\text{pr}(\text{V}^P_{jt})$. Next, δ_2 and δ_4 represent the additional change in preference rankings for treated households due to the treatment. If this additional change is relatively more associated with raw values of value added, then δ_2 will be larger than δ_4. Conversely, if the additional change is better captured by our predicted preference rankings under correct beliefs, then δ_4 will be larger.[72]

Table 17 (page 44) presents the results. The values are for versions of regression (9) that exclude a household's two top baseline choices.[73] The first column replicates the results from Section 3.3. It includes only $\text{pr}(\text{V}^P_{jt})$, as in equation (6). The second column replaces $\text{pr}(\text{V}^P_{jt})$ with $\text{ppr}_{ij,\text{CPP}}$. Finally, the third column is the horse race; it includes both variables simultaneously. The results reveal that, for the full sample of students, the treatment effect is better explained by the predicted preference rankings, $\text{ppr}_{ij,\text{CPP}}$, than by raw value added, $\text{pr}(\text{V}^P_{jt})$. In the horse race, all of the effect loads onto the former variable.

However, the remaining columns reveal important differences between low- and high-achieving students. For low-achieving students, the treatment caused substantial and statistically significant increases in both the extent to which households preferred tracks with higher value added (Column 4) and in the association between actual and predicted preference ranks (Column 5). Further, in the horse race, all of the treatment effect loads onto $\text{pr}(\text{V}^P_{jt})$ (Column 6). For high-achieving students, the treatment significantly increased the association with predicted preference ranks (Column 8), but exerted little impact on demand for value added (Column 7). Moreover, in the horse race, all of the treatment effect loads onto $\text{ppr}_{ij,\text{CPP}}$ (Column 9).

71. We predict preference rankings using a simulation procedure. First, we calculate a household's utility for a track under $\epsilon_{ij} = 0$ by multiplying the preference coefficients by the given set of quality scores. Call this quantity $u_{ij,\text{CPP}}$. Second, we draw independent Type-1 Extreme Value (T1EV) errors for each track and add them to the $u_{ij,\text{CPP}}$ values. Third, we rank the tracks according to the resulting utilities. This generates a simulated preference ranking under the given draw of T1EV errors. Fourth, we repeat Steps 2 and 3 for a large number of draws. Finally, for each track, we average its simulated ranks.

72. Further, if preference rankings for treated households mirror $\text{ppr}_{ij,\text{CPP}}$, then $\delta_3 + \delta_4$ will be 1 and $\delta_1 + \delta_2$ will be 0. By contrast, if their preference rankings mimic the value added rankings that we provided, then $\delta_1 + \delta_2$ will be 1 and $\delta_3 + \delta_4$ will be 0. In practice, neither of these will happen because few households rank all tracks.

73. Recall that the treatment had no effect on households' beliefs or preference ranks for these tracks.

Table 17: Explaining treatment effects on preference rankings
using value added and predicted preference rankings

	All students			**Low-achieving**			**High-achieving**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Value added,	0.096***		-0.065***	0.002		-0.167***	0.145***		0.003
$\text{pr}(\text{V}^P_{jt})$	(0.018)		(0.024)	(0.028)		(0.041)	(0.021)		(0.027)
Value added: treated,	0.062***		-0.022	0.132***		0.105**	0.024		-0.099**
$T_i \cdot \text{pr}(\text{V}^P_{jt})$	(0.023)		(0.029)	(0.039)		(0.046)	(0.027)		(0.038)
Predicted preference rank,		0.248***	0.308***		0.181***	0.348***		0.267***	0.265***
$\text{ppr}_{ij,\text{CPP}}$		(0.027)	(0.037)		(0.039)	(0.066)		(0.029)	(0.036)
Predicted preference rank: treated,		0.115***	0.138***		0.141***	0.037		0.112***	0.204***
$T_i \cdot \text{ppr}_{ij,\text{CPP}}$		(0.031)	(0.040)		(0.051)	(0.068)		(0.033)	(0.044)
Clusters	76	76	76	74	74	74	75	75	75
Students	1,533	1,533	1,533	571	571	571	962	962	962
Student-tracks	17,092	17,092	17,092	6,083	6,083	6,083	11,009	11,009	11,009

The table presents results from regression (9). The regressions include indicators for the interaction between a track's position in a household's baseline preference ranking and whether the household is in the treatment group. In calculating these indicators, we create separate groups for tracks that households left unranked in the baseline survey. Columns 1, 4, and 7 replicate the third columns of Tables 14, A18, and A19, respectively. The sample is equal to the set of students in the follow-up survey. Standard errors are clustered by the middle school treatment-control pairs within which we conducted the randomization.

In short, the results suggest that for households with high-achieving students, the treatment mainly impacted beliefs. These households were more likely to reason through changes in their preference rankings by combining updated beliefs with stable preferences for track characteristics. Owing to the small effect of the treatment on their beliefs and to the multi-dimensional nature of their preferences, the treatment hardly affected whether they preferred high-value added tracks. By contrast, for households with low-achieving students, the treatment impacted both beliefs and preferences. Thus, it caused a sizeable change in their demand for track value added.

5 Conclusion

Friedman (1955) argued that giving households freedom to choose schools would improve educational performance. At first pass this is a straightforward claim, since it extends standard results from markets for consumer goods to education. Yet rigorous empirical work has produced mixed results on Friedman's prediction. For example, voucher experiments suggest that choice can impact students' measured skills in ways that are highly positive (Bettinger et al. 2017), highly negative (Abdulkadiroglu, Pathak, and Walters 2018), or modest (Muralidharan and Sundararaman 2015). Considering analogous mixed evidence on the effects of access to selective schools, Beuermann and Jackson (2020) observe that "the lack of robust achievement effects of attending schools that parents prefer is something of a puzzle."

We have brought new types of data and an experiment to bear on two possible explanations for this puzzle. The first is that a lack of information prevents households from choosing high value added schools. Analogous possibilities arise in recent work considering frictions in other markets in which households make choices with long term implications, such as selecting neighborhoods

(Bergman, Chan, and Kapor 2020). The second is that households have preferences that lead them to prioritize school attributes other than value added. This possibility arises in theoretical and empirical work suggesting this may be rational when school quality is multidimensional (MacLeod and Urquiola 2019; Beuermann et al. 2019).

Our results suggest that at least in Romania, both of these possibilities are relevant. First, we find that essentially all types of households make school choices that leave value added on the table. Second, our experiment shows that distributing information can affect households' school rankings, placements, and value added. While the overall effect is modest, it is substantial for households unable to access their top two choices, i.e., those with lower-achieving children. Finally, our preference estimates and simulations suggest that this heterogeneity originates in three factors: First, the treatment had a larger effect on beliefs for lower achieving households; second, this affected only rankings for choices outside the top two; third, the treatment also affected preferences for low-achieving households.

Overall, these findings are consistent with conclusions that emerge elsewhere. For instance, Abdulkadiroglu, Angrist, and Pathak (2014) and Abdulkadiroglu et al. (2020) find that there is an overall correlation between value added and selectivity in New York City, although not necessarily among elite schools. Our key results are also consistent with work finding that information on value added may affect school markets less than data on absolute achievement (Mizala and Urquiola 2013; Imberman and Lovenheim 2016).

That said, the effects of information could be larger or smaller in other settings. On the one hand, Romanian public schools are relatively homogenous in terms of resources. This may make it difficult for households to observe value added. On the other hand, the towns we studied contain fairly standardized markets, with a clear value added measure and few other constraints on choice, such as cost or distance. This suggests that the market mechanism may work even less well in other settings.

Finally, we note some issues left for further research. First, the effects of information might be larger and of a general equilibrium nature if information can be delivered in greater doses and in a more sustained fashion than we did. Second, our results suggest that at least in Romania, households attach a lot of weight to their top school choices. This might generalize to other settings, where at least anecdotally households tend to focus on "favorite" schools. The origins and implications of such behavior may be relevant to understanding school markets. Third, our results leave open questions on whether information interventions change only students' information sets, as opposed to affecting their preferences; these might have different implications in terms of wellbeing and schooling outcomes.

References

Abdulkadiroglu, Atila, Joshua Angrist, and Parag Pathak. 2014. "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools." *Econometrica* 82 (1): 137–196.

Abdulkadiroglu, Atila, Parag A. Pathak, Jonathan Schellenberg, and Christopher R. Walters. 2020. *Do Parents Value School Effectiveness?* Technical report.

Abdulkadiroglu, Attila, Parag Pathak, and Christopher R. Walters. 2018. "Free to choose: Can school choice reduce student achievement?" *American Economic Journal: Applied Economics* 10 (1): 175–206.

Ajayi, Kehinde F., Wila H. Friedman, and Adrienne M. Lucas. 2017. "The Importance of Information Targeting for School Choice." *American Economic Review: Papers and Proceedings* 107 (5): 638–643.

Allende, Claudia, Franciso Gallego, and Christopher Neilson. 2019. *Approximating the Equilibrium Effects of Information Interventions.* Mimeo. Princeton University.

Andrabi, Tahir, Jishnu Das, and Asim Khwaja. 2017. "Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets." *American Economic Review* 107 (6): 1535–15635.

Angrist, Joshua D., Peter D. Hull, Parag A. Pathak, and Christopher R. Walters. 2017. "Leveraging Lotteries for School Value-Added: Testing and Estimation." *The Quarterly Journal of Economics* 132, no. 2 (February): 871–919.

Athey, Susan, Rina Friedberg, Julie Tibshirani, and Stefan Wager. 2019. *Local linear forests.* Technical report. Stanford University.

Bergman, Peter, Eric Chan, and Adam Kapor. 2020. *Housing Search Frictions: Evidence from Detailed Search Data and a Field Experiment.* Working Paper. Columbia University, March.

Bergman, Peter, Raj Chetty, Stefanie DeLuca, Nathaniel Hendren, Lawrence F Katz, and Christopher Palmer. 2019. *Creating Moves to Opportunity: Experimental Evidence on Barriers to Neighborhood Choice.* Working Paper, Working Paper Series 26164. National Bureau of Economic Research, August.

Bettinger, Eric, Michael Kremer, Maurice Kugler, Carlos Medina, Christian Posso, and Juan Esteban Saavedra. 2017. *Can educational voucher programs pay for themselves?* Mimeo. Harvard University.

Beuermann, Diether, C. Kirabo Jackson, Laia Navarro-Sola, and Francisco Pardo. 2019. *What is a Good School, and Can Parents Tell? Evidence on the Multidimensionality of School Output.* Technical report. Northwestern University.

Beuermann, Diether W., and C. Kirabo Jackson. 2020. "Do Parents Know Best? The Short and Long-Run Effects of Attending the Schools that Parents Prefer." *Journal of Human Resources,* no. 2.

Borcan, Oana, Mikael Lindahl, and Andreea Mitrut. 2017. "Fighting Corruption in Education: What Works and Who Benefits?" *American Economic Journal: Economic Policy* 9, no. 1 (February): 180–209.

Burgess, S., E. Greaves, A. Vignoles, and D. Wilson. 2015. "What Parents Want: School Preferences and School Choice." *Economic Journal* 125 (587): 1262–1289.

Cattaneo, Matias D., Rocio Titiunik, Gonzalo Vazquez-Bare, and Luke Keele. 2016. "Interpreting Regression Discontinuity Designs with Multiple Cutoffs." *The Journal of Politics* 78, no. 4 (October): 1229–1248.

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

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson. 2016. "Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector." *The American Economic Review* 106 (8): 2110–2144.

Chetty, Raj, John N Friedman, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter. 2018. *The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility.* Working Paper, Working Paper Series 25147. National Bureau of Economic Research, October.

Chetty, Raj, John N Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan. 2020. "Income Segregation and Intergenerational Mobility Across Colleges in the United States*." *The Quarterly Journal of Economics* 135, no. 3 (February): 1567–1633.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104, no. 9 (September): 2593–2632.

Chetty, Raj, and Nathaniel Hendren. 2018. "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates." *The Quarterly Journal of Economics* 133, no. 3 (February): 1163–1228.

Corcoran, Sean P., Jennifer L. Jennings, Sarah R. Cohodes, and Carolyn Sattin-Bajaj. 2018. *Leveling the Playing Field for High School Choice: Results from a Field Experiment of Informational Interventions.* Technical report. National Bureau of Economic Research Working Paper No. 24471.

Dahl, Gordon B., Dan-Olof Rooth, and Anders Stenberg. 2020. "Long-Run Returns to Field of Study in Secondary School." *Working paper* (May).

Deming, David J. 2014. "Using School Choice Lotteries to Test Measures of School Effectiveness." *American Economic Review* 104, no. 5 (May): 406–11.

Friedman, Milton. 1955. "The Role of Government in Education." In *Economics and the Public Interest,* edited by Robert Solow. NJ: Trustees of Rutgers College.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica* 69 (1): 201–209.

Hastings, Justine, Thomas Kane, and Douglas Staiger. 2005. *Parental Preferences and School Competition: Evidence from a Public School Choice Program.* Mimeo. National Bureau of Economic Research Working Paper No. 11805.

Hastings, Justine, Christopher A. Neilson, and Seth D. Zimmerman. 2018. *The Effects of Earnings Disclosure on College Enrollment Decisions.* Mimeo. National Bureau of Economic Research Working Paper No. 21300.

Hastings, Justine, and Jeffrey Weinstein. 2008. "Information, School Choice, and Academic Achievement: Evidence from Two Experiments." *Quarterly Journal of Economics* 123, no. 4 (November): 1373–1414.

Imberman, Scott, and Michael Lovenheim. 2016. "Does the market value value-added? Evidence from housing prices after public release of school and teacher value added." *Journal of Urban Economics* 91:104–121.

Kirkeboen, Lars J., Edwin Leuven, and Magne Mogstad. 2016. "Field of Study, Earnings, and Self-Selection." *The Quarterly Journal of Economics* 131, no. 3 (May): 1057–1111.

MacLeod, W. Bentley, and Miguel Urquiola. 2019. "Is education consumption or investment? Implications for the effects of school competition." *Annual Review of Economics* 11:563–589.

Mizala, Alejandra, and Miguel Urquiola. 2013. "Parental choice and school markets: The impact of information on school effectiveness." *Journal of Development Economics* 103:313–335.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2015. "The Aggregate Effect of School Choice: Evidence from a Two-stage Experiment in India." *The Quarterly Journal of Economics* 130 (3): 1011–1066.

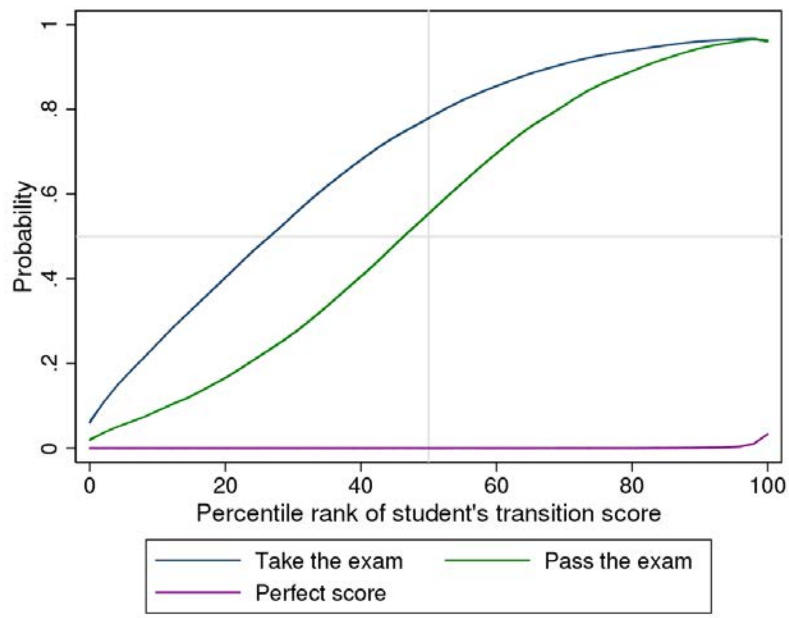
Rothstein, Jesse. 2006. "Good Principals or Good Peers: Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition Among Jurisdictions." *American Economic Review* 96 (4): 1333–1350.

———. 2010. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *The Quarterly Journal of Economics* 125, no. 1 (February): 175–214.

———. 2017. "Measuring the Impacts of Teachers: Comment." *American Economic Review* 107, no. 6 (June): 1656–84.

Additional figures (not for publication)

Figure A1: Baccalaureate exam outcomes by student's transition score



The figure presents local linear regressions of students' baccalaureate exam outcomes versus their transition scores. The horizontal axis represents the student's within-year percentile rank by transition score. "Take the exam," "Pass the exam," and "Perfect score" are indicator variables.

Figure A2: Page 1 of the information sheet provided to treated survey respondents

Information form

Town		**Code town:**	
School		**Code school:**	
Class			

You are receiving this information form because you agreed to participate in the study of the admission process for high schools in Romania. This study is done by CCSAS with the approval of the Ministry of Education in collaboration with researchers at New York University in the United States of America.

In order to help you and your child make the best choices during the admission process, we wanted to share some information with you.

The information on the admission process is available online:

1.) Government order Nr. 4829/2018 from August 30, 2018 on the admission process in 2019-2020 is available here:
http://ismb.edu.ro/documente/examene/admitere/2019/1_ORDIN_nr_%204829_30_08_2018.pdf

2.) The admission application form is available here:
http://ismb.edu.ro/documente/examene/admitere/2019/1_Fisa_Admitere_2019.pdf

3.) Information on admission scores in previous years are available here:
www.admitere.edu.ro

The figure displays the first half of the information sheet provided to parents in treated middle schools (at the conclusion of the baseline survey). The information sheet for parents in control middle schools included all the text until the fourth bullet point.

Figure A3: Page 2 of the information sheet provided to treated survey respondents

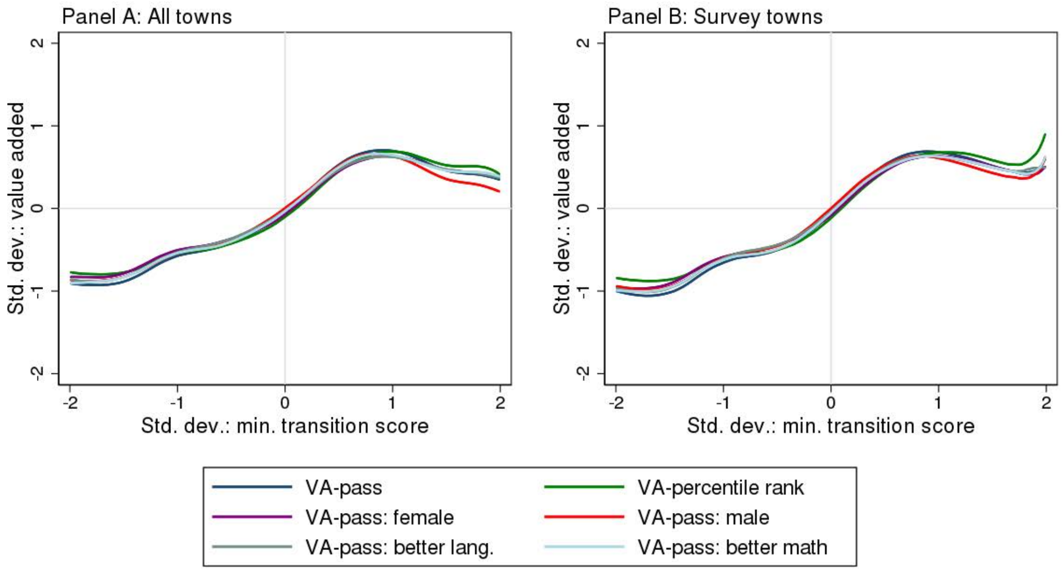
A team of economists at New York University has analyzed data in your hometown, Sebes Alba. They have calculated which tracks most effectively improve students' chances of passing the baccalaureate exam relative to their 9th grade starting points.

Rank of most effective track	Name of School	Name of track
1	NATIONAL HIGH SCHOOL "LUCIAN BLAGA" SEBES	Math-Computer Science
2	NATIONAL HIGH SCHOOL "LUCIAN BLAGA" SEBES	Natural Science
3	TECHNOLOGICAL HIGH SCHOOL SEBES	Economics
4	NATIONAL HIGH SCHOOL "LUCIAN BLAGA" SEBES	Social Science
5	GERMAN HIGH SCHOOL SEBES	Natural Science
6	TECHNOLOGICAL HIGH SCHOOL SEBES	Textile Industry
7	NATIONAL HIGH SCHOOL "LUCIAN BLAGA" SEBES	Philology-English
8	NATIONAL HIGH SCHOOL "LUCIAN BLAGA" SEBES	Philology

In case you have questions about the data and information provided, please call the headquarters of CCSAS at 0744393121 or 0729634372.

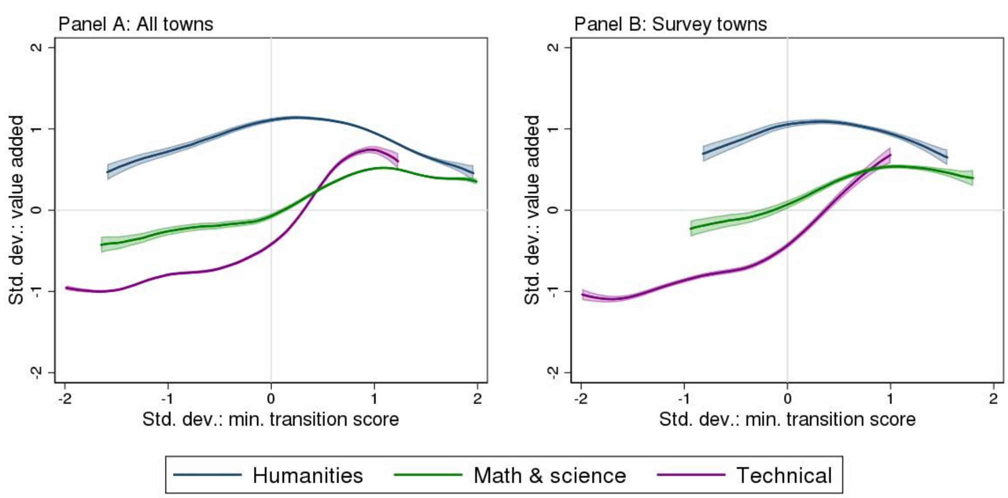
The figure displays the second half of the information sheet provided to parents in treated middle schools (at the conclusion of the baseline survey).

Figure A4: The rel. between V.A. and selectivity: robustness to alternative V.A. measures



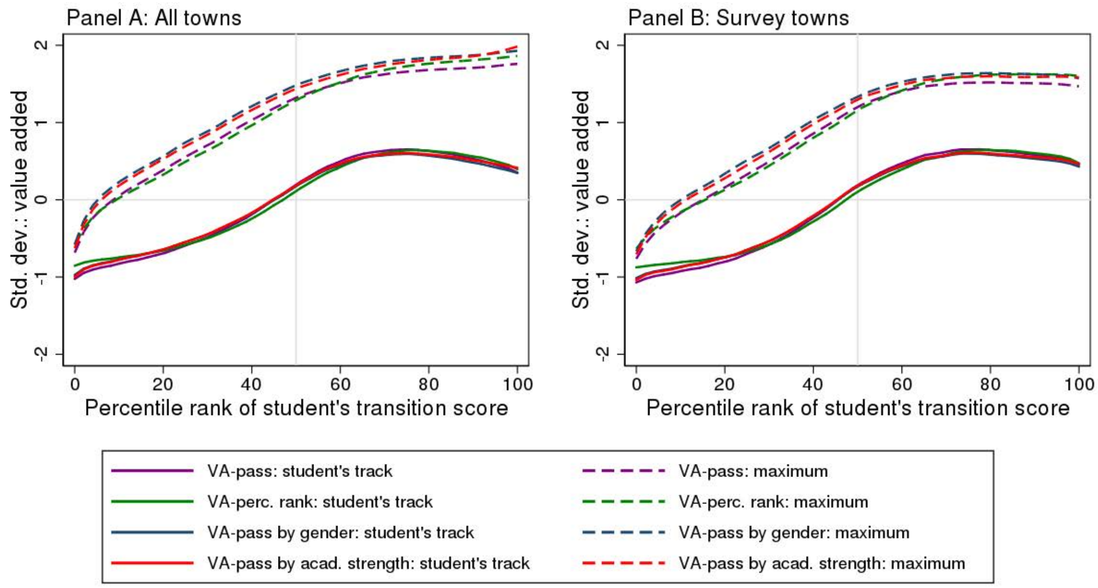
The figure replicates Figure 1 for alternative value added measures. Specifically, it presents local linear regressions of standardized values of various value added measures on standardized values of minimum transition score, MTS_{jt}. The value added measures are: (i) "VA-pass": a track-year effect on the probability of passing the baccalaureate exam; (ii) "VA-percentile rank": a track-year effect on the percentile rank of a student's exam performance; and (iii) track-year effects on the probability of passing the exam that vary by a student's gender or relative academic strength. See Section 0.3 for definitions of each value added measure. See Figure 1 for additional details on methodology and sample construction.

Figure A5: The relationship between value added and selectivity by track category



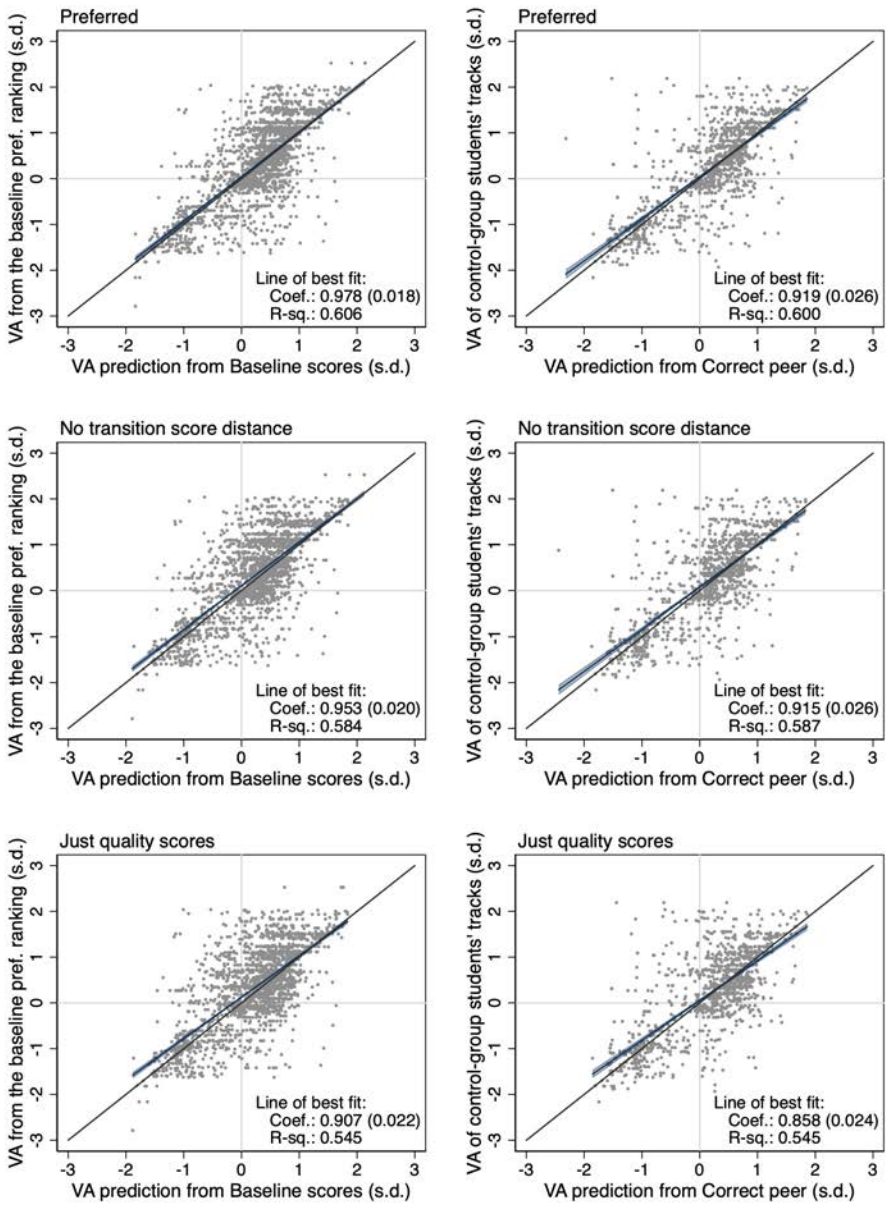
The figure presents the relationship between value added and selectivity for subsets of tracks by a track's curriculum. Specifically, it replicates the local linear regressions in Figure 1 separately for tracks with curricular focuses in humanities, math and science, or technical subjects. See Figure 1 for additional details on methodology and sample construction.

Figure A6: Choice patterns by transition score: robustness to alternative value added measures



The figure replicates Figure 2 for alternative value added measures. The dotted lines represent local linear regressions of the maximum value of value added in the student's feasible set versus the within-year percentile rank of the student's transition score. The solid lines are local linear regressions of the value of value added in the track the student attends versus the student's percentile rank. The value added measures are: (i) a track-year effect on the probability of passing the baccalaureate exam, (ii) a track-year effect on the percentile rank of a student's exam performance, and (iii) track-year effects on the probability of passing that vary by a student's gender or relative academic strength. See Section 0.3 for definitions of each value added measure and Figure 2 for details on methodology and sample construction.

Figure A7: Comparing our predictions with the value added of students' observed choices



The figure shows how our predictions for the value added of students' tracks compare with the value added of students' observed choices. The plots in column 1 compare $V_{i,BS}$ with the value added of the track a student would attend under the household's baseline preference ranking. The plots in column 2 are just for students in the control group. They compare $V_{i,CP}$ with the value added of students' actual tracks. The three rows represent different specifications of the preference model. See Section 4.2 for definitions of "Preferred", "No transition score distance", and "Just quality scores". The black line is a 45-degree line, and the blue line is the line of best fit. The slope coefficient and R-squared for the line of best fit are shown in the bottom-right corner of each plot. The sample is similar to that for the experimental treatment effects from Section 3. However, it excludes 15 students who did not score above the cutoffs of any of the tracks that existed in both 2018 and 2019. (These students were assigned to tracks that were newly created in 2019.) In addition, the plots in column 1 exclude students who did not rank any feasible tracks in the baseline survey.

Additional tables (not for publication)

Table A1: Administrative data sample size, by year

Year	Towns	High schools	Tracks	Students
2004	426	1,247	3,691	185,383
2005	405	1,223	3,500	146,712
2006	386	1,195	3,284	136,671
2007	383	1,192	3,259	134,692
2008	476	1,305	4,851	172,174
2009	438	1,261	4,470	170,087
2010	417	1,226	4,018	164,146
2011	437	1,242	4,506	187,442
2012	410	1,207	4,234	146,114
2013	420	1,208	4,269	141,934
2014	378	1,144	3,784	124,675
2015	368	1,129	3,649	121,880
2016	362	1,116	3,541	115,902
2017	351	1,098	3,427	109,694
2018	348	1,084	3,342	110,101
2019	312	1,015	3,038	105,230
Mean	395	1,181	3,804	142,052
Total	6,317	18,892	60,863	2,272,837
Distinct	512	1,402	13,420	2,272,837

The table presents summary statistics on the administrative data by year. It restricts the sample to towns that have at least two tracks in the given year. "Mean" is the average number of the listed quantity during 2004-2019. "Total" is the sum of the quantity over those years. "Distinct" is the number of distinct towns, high schools, tracks, and students. The sample varies year to year because tracks go in and out of existence reflecting factors like changes in student enrollment, the emergence of technical fields, and instructor availability.

Table A2: Correlations of alternative V.A. measures with track-year effects on passing the baccalaureate exam

Value added measure	Correlation	Town-years	Track-years	Students
Percentile rank of exam performance	0.944	4,576	43,866	1,710,030
Pass the exam:				
Female	0.924	4,572	41,435	1,677,023
Male	0.915	4,575	43,216	1,704,417
Better at language	0.937	4,567	42,622	1,700,886
Better at math	0.929	4,575	43,587	1,708,946

The table presents correlations between estimates for our main value added measure with those for alternative measures. The main measure is a track-year effect on a student's probability of passing the baccalaureate exam. The alternative measures are: i) a track-year effect on the percentile rank of a student's performance, and ii) track-year effects on the probability of passing the exam that vary by student gender or relative academic strength. See Section 0.3 for details. Correlations are weighted by student.

Table A3: Covariates used in the prediction model: covariates of the track

Covariate	**Lags**								
	0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Curricular focus	Y								
Language	Y								
Number of students		Y	Y	Y		Y	Y		
Transition score: minimum		Y	Y						
Transition score: maximum			Y						
Transition score: average			Y	Y		Y	Y	Y	
Transition score: std. dev.			Y	Y					
Middle school GPA: average			Y						
Middle school GPA: std. dev.			Y						
Transition exam: Math score: average			Y	Y					
Transition exam: Math score: std. dev.			Y						
Transition exam: Romanian score: average			Y						
Transition exam: Romanian score: std. dev.			Y	Y					
Share female			Y	Y		Y	Y		
Number of students in students' middle schools: average			Y						
Number of students in students' middle schools: std. dev.			Y						
Average transition score in students' middle schools: average			Y						
Average transition score in students' middle schools: std. dev.			Y						
Std. dev. of transition score in students' middle schools: average			Y						
Average GPA in students' middle schools: average			Y						
Average transition exam: Math score in students' middle schools: average			Y						
Average transition exam: Rom. score in students' middle schools: average			Y						
Rank of track in school by minimum transition score		Y							
Rank of track in curricular focus by minimum transition score		Y							
Rank of track in town by minimum transition score		Y							
Rank of track in town by average transition score			Y						
Rank of track in school by average transition score			Y						
Share of students who took the baccalaureate exam						Y	Y	Y	
Rank of track in town by value added						Y	Y	Y	
Value added						Y	Y		
Value added de-meaned by town-year						Y	Y	Y	Y
Standard error of value added						Y	Y		

The table lists covariates used in the local linear forest prediction model. Specifically, the subset of covariates concerning characteristics of the track being predicted. A "Y" indicates that the specified lag of the covariate is included in the model.

Table A4: Covariates used in the prediction model: covariates of the track's school

Covariate	**Lags**								
	0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Number of tracks	Y	Y				Y			
Number of academic tracks	Y	Y				Y			
Number of languages or social sciences tracks	Y								
Number of math tracks	Y								
Number of natural sciences or technical tracks	Y								
Number of Romanian-language tracks	Y								
Number of Hungarian-language tracks	Y								
Number of students		Y	Y	Y					
Transition score: minimum		Y	Y						
Maximum of tracks' minimum transition scores		Y	Y						
Transition score: average			Y	Y					
Transition score: std. dev.			Y						
Middle school GPA: average			Y						
Middle school GPA: std. dev.			Y						
Transition exam: Romanian score: average			Y						
Transition exam: Romanian score: std. dev.			Y						
Share female			Y						
Number of students in students' middle schools: average			Y						
Average transition score in students' middle schools: average			Y						
Average transition score in students' middle schools: std. dev.			Y						
Share of students who took the baccalaureate exam						Y	Y		
Value added de-meaned by town-year: average						Y	Y		
Value added: std. dev.						Y	Y		

The table lists covariates used in the local linear forest prediction model. Specifically, the subset of covariates concerning the high school of the track being predicted. A "Y" indicates that the specified lag of the covariate is included in the model.

Table A5: Covariates used in the prediction model: covariates of the track's town

Covariate	**Lags**								
	0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Number of schools	Y								
Number of tracks	Y	Y				Y			
Number of academic tracks	Y								
Number of languages or social sciences tracks	Y	Y				Y			
Number of math tracks	Y	Y				Y			
Number of natural sciences or technical tracks	Y	Y				Y			
Number of Romanian-language tracks	Y								
Number of Hungarian-language tracks	Y								
Number of students		Y	Y	Y					
Transition score: average			Y	Y					
Transition score: std. dev.			Y						
Middle school GPA: average			Y						
Middle school GPA: std. dev.			Y						
Transition exam: Romanian score: average			Y						
Transition exam: Romanian score: std. dev.			Y						
Share of students who took the baccalaureate exam						Y	Y		
Value added: average						Y	Y		
Value added: std. dev.						Y	Y		

The table lists covariates used in the local linear forest prediction model. Specifically, the subset of covariates concerning the town of the track being predicted. A "Y" indicates that the specified lag of the covariate is included in the model.

Table A6: Summary statistics on survey towns

County	Town	R-squared	2018		2019		Survey		
			Tracks	*Students*	*Tracks*	*Students*	*Students*	*Middle schools*	*Two-class schools*
Alba	Alba Iulia	0.905	16	504	15	476	132	6	2
Alba	Sebes	0.862	10	290	10	297	35	3	0
Arges	Campulung	0.819	13	423	11	420	67	4	0
Bacau	Moinesti	0.791	9	303	9	280	87	3	2
Bacau	Onesti	0.809	16	650	16	637	157	6	2
Bihor	Beius	0.607	11	307	10	322	72	2	2
Bistrita Nasaud	Bistrita	0.825	28	925	23	782	148	7	2
Brasov	Fagaras	0.915	10	323	9	273	117	3	2
Buzau	Ramnicu Sarat	0.741	12	476	13	445	113	4	2
Calarasi	Calarasi	0.901	24	666	20	709	161	8	2
Caras Severin	Resita	0.572	20	473	18	425	103	7	1
Cluj	Dej	0.883	10	300	10	299	80	4	1
Cluj	Gherla	0.658	10	261	10	265	37	2	0
Cluj	Turda	0.757	12	282	11	281	71	5	0
Constanta	Mangalia	0.563	10	336	9	252	145	5	2
Constanta	Medgidia	0.601	10	308	9	280	27	1	0
Covasna	Sfantul Gheorghe	0.856	20	396	20	437	43	2	1
Covasna	Tirgu Secuiesc	0.592	9	219	9	233	43	3	0
Dolj	Calafat	0.795	7	183	6	168	37	2	0
Galati	Tecuci	0.872	18	753	16	728	79	5	0
Giurgiu	Giurgiu	0.853	15	591	14	602	148	9	2
Gorj	Motru	0.590	11	362	9	308	53	3	0
Harghita	Gheorgheni	0.869	11	280	12	263	22	2	0
Harghita	Miercurea Ciuc	0.852	22	602	21	589	48	4	1
Harghita	Odorheiu Secuiesc	0.855	15	392	15	364	39	3	1
Harghita	Toplita	0.754	7	170	8	172	22	2	0
Hunedoara	Deva	0.921	19	369	19	353	102	5	1
Hunedoara	Hunedoara	0.753	11	364	10	308	91	6	0
Hunedoara	Petrosani	0.735	9	299	8	224	101	4	2
Ialomita	Slobozia	0.912	19	636	16	644	91	4	2
Ialomita	Urziceni	0.886	11	316	7	280	59	3	0
Iasi	Harlau	0.858	8	222	7	224	34	2	0
Iasi	Pascani	0.838	17	688	16	644	109	4	2
Iasi	Targu Frumos	0.798	7	222	6	196	49	3	0
Maramures	Sighetu Marmatiei	0.689	21	582	19	565	104	5	2
Mures	Sighisoara	0.816	14	307	14	301	55	4	0
Mures	Tarnaveni	0.585	8	231	8	194	46	3	0
Neamt	Roman	0.863	21	825	19	672	48	3	1
Prahova	Campina	0.738	16	530	16	554	76	4	0
Salaj	Zalau	0.898	22	759	21	741	125	7	1
Satu Mare	Carei	0.832	10	247	8	224	54	3	0
Suceava	Gura Humorului	0.409	8	304	9	289	48	3	0
Suceava	Radauti	0.800	16	672	18	672	114	4	2
Teleorman	Alexandria	0.678	15	699	16	746	88	4	2
Timis	Lugoj	0.540	14	427	12	373	131	6	0
Valcea	Dragasani	0.787	12	328	7	308	108	2	2
Vaslui	Birlad	0.873	20	758	18	694	158	8	2
Vrancea	Adjud	0.798	9	314	7	280	21	2	0
	Total	-	663	20,874	614	19,793	3,898	194	44
	Mean	0.773	13.8	435	12.8	412	81.2	4.0	0.9
	Min	0.409	7	170	6	168	21	1	0
	Max	0.921	28	925	23	782	161	9	2

The table presents summary statistics on towns included in the survey. R-squared is the fraction of the variation in value added, V_{jt}, explained by predicted value added, V^P_{jt}, for the town during 2008-2014 (see the notes to Table 2). "Two-class schools" indicates the number of middle schools in which we visited two classrooms.

Table A7: Summary statistics for households' baseline scores for track characteristics

	Students	Student-tracks	Mean	Std. dev.	Min	Max
Location	2,673	19,959	3.88	1.30	1	5
Siblings & friends	2,091	15,588	2.98	1.64	1	5
Peer quality	2,496	18,478	3.62	1.36	1	5
Curriculum	2,516	18,134	3.45	1.44	1	5
Teacher quality	2,478	17,940	3.87	1.28	1	5
VA: pass the bacc.	2,469	17,882	3.75	1.35	1	5
VA: college	2,406	17,451	3.56	1.43	1	5
VA: wages	2,343	17,260	3.53	1.39	1	5

The table describes households' baseline scores for track characteristics. The mean and standard deviation are weighted by student.

Table A8: Correlations between households' baseline scores for track characteristics

	Location	Siblings	Peers	Curriculum	Teachers	Pass bacc.	College	Wages
Location	1							
Siblings & friends	0.498	1						
Peer quality	0.576	0.608	1					
Curriculum	0.522	0.613	0.774	1				
Teacher quality	0.579	0.539	0.770	0.726	1			
VA: pass the bacc.	0.551	0.565	0.773	0.772	0.808	1		
VA: college	0.523	0.589	0.782	0.805	0.755	0.867	1	
VA: wages	0.514	0.577	0.732	0.751	0.737	0.812	0.850	1

The table shows correlations between respondents' scores for different track characteristics. Values are weighted by student.

Table A9: Year-specific correlations between value added and selectivity

Year	**Coefficient**	**Std. error**	**Towns**	**Tracks**	**Students**
2004	0.618	0.018	426	3,691	185,383
2005	0.452	0.025	405	3,500	146,712
2006	0.504	0.022	386	3,284	136,671
2007	0.550	0.018	383	3,259	134,692
2008	0.553	0.015	476	4,851	172,174
2009	0.609	0.012	438	4,470	170,087
2010	0.566	0.015	417	4,018	164,146
2011	0.575	0.014	437	4,506	187,442
2012	0.580	0.016	410	4,234	146,114
2013	0.535	0.017	420	4,269	141,934
2014	0.478	0.014	378	3,784	124,675
2015	0.531	0.016	368	3,649	121,880
2016	0.535	0.017	362	3,541	115,902
2017	0.508	0.019	351	3,427	109,694
2019	0.500	0.019	312	3,038	105,230

The table presents year-specific correlations between a track's value added and its selectivity. Specifically, it displays coefficients from regressions of standardized values of value added estimates, $\hat{\mathrm{V}}_{jt}$, on standardized values of minimum transition score, MTS_{jt}. The sample includes the full set of towns. See Figure 1 and Table 6 for additional details on methodology and sample construction.

Table A10: Summary statistics on households' track choices:
Feasible tracks with the same curricular category as the track the student attends

	All towns			Survey towns		
	All students	*Low-achieving*	*High-achieving*	*All students*	*Low-achieving*	*High-achieving*
Panel A: Percent of students with only one track in the choice set	11.7	15.3	8.0	7.3	9.8	4.9
Panel B: Mean percentile rank of student's track among tracks in the choice set						
Value added, $\hat{V}_{jt}$	65.3	64.1	66.4	67.0	63.7	70.0
Selectivity, MTS_{jt}	79.8	75.6	83.6	78.9	76.2	81.4
Panel C: Mean potential increase (std. dev.) among tracks in the choice set						
Value added, $\hat{V}_{jt}$	0.62	0.65	0.59	0.49	0.53	0.45
Selectivity, MTS_{jt}	0.26	0.31	0.21	0.26	0.28	0.23
Number of students	2,162,736	1,081,075	1,081,661	424,508	211,917	212,591

This table replicates Table 7 but uses a different choice set. The choice set in Table 7 is the set of tracks that a student is eligible to attend (i.e., the student's "feasible set", $\mathcal{J}_i^e$). By contrast, the choice set in this table is the subset of feasible tracks whose curricula fall into the same category as that of the student's track. Curricular categories are humanities, math and science, and technical subjects. See Table 7 for additional details on variable definitions and sample construction.

Table A11: Explaining within-town quintiles of track attributes using households' quality scores:
Households who scored all tracks

	All students		Low-achieving		High-achieving	
	quint(V^P_{jt})	quint(MTS_{jt-1})	quint(V^P_{jt})	quint(MTS_{jt-1})	quint(V^P_{jt})	quint(MTS_{jt-1})
Score: VA-pass	0.446***		0.420***		0.463***	
	(0.020)		(0.035)		(0.017)	
Score: Peers		0.611***		0.589***		0.631***
		(0.015)		(0.028)		(0.014)
R-sq.	0.19	0.37	0.15	0.30	0.23	0.42
Clusters	117	117	89	89	106	106
Students	811	811	308	308	503	503
Student-tracks	10,393	10,393	3,988	3,988	6,405	6,405

The table presents results analogous to those in Table 8. However, the sample is limited to survey respondents who provided quality scores for both value added and peer quality for all of the tracks in their towns. See Table 8 for additional details.

Table A12: Explaining within-town quintiles of track attributes using households' quality scores:
Tracks that would have been feasible in the prior year

	All students		Low-achieving		High-achieving	
	quint(V^P_{jt})	quint(MTS_{jt-1})	quint(V^P_{jt})	quint(MTS_{jt-1})	quint(V^P_{jt})	quint(MTS_{jt-1})
Score: VA-pass	0.425***		0.314***		0.438***	
	(0.018)		(0.027)		(0.018)	
Score: Peers		0.569***		0.378***		0.600***
		(0.013)		(0.021)		(0.012)
R-sq.	0.18	0.33	0.10	0.20	0.20	0.38
Clusters	186	186	158	158	177	177
Students	2,136	2,136	682	682	1,454	1,454
Student-tracks	13,691	13,691	3,261	3,261	10,430	10,430

The table presents results analogous to those in Table 8. However, the sample is limited to student-track observations in which the track would have been feasible for the student in the prior year. These are observations in which the student's transition score is greater than or equal to the track's prior-year minimum transition score, MTS_{jt-1}. See Table 8 for additional details.

Table A13: Explaining within-town quintiles of track attributes using households' quality scores: Households who are certain of their preference rankings

	All students		**Low-achieving**		**High-achieving**	
	quint(V^P_{jt})	quint(MTS_{jt-1})	quint(V^P_{jt})	quint(MTS_{jt-1})	quint(V^P_{jt})	quint(MTS_{jt-1})
Score: VA-pass	0.438***		0.388***		0.459***	
	(0.020)		(0.039)		(0.017)	
Score: Peers		0.583***		0.491***		0.622***
		(0.018)		(0.048)		(0.015)
R-sq.	0.20	0.35	0.13	0.22	0.23	0.42
Clusters	176	176	127	127	158	158
Students	1,042	1,042	309	309	733	733
Student-tracks	7,288	7,288	2,252	2,252	5,036	5,036

The table presents results analogous to those in Table 8. However, the sample is limited to survey respondents who reported being "very certain" of their preference rankings in the baseline survey. See Table 8 for additional details.

Table A14: Explaining track attributes (in std. dev.) using households' quality scores

	All students		**Low-achieving**		**High-achieving**	
	V^P_{jt} (s.d.)	MTS_{jt-1} (s.d.)	V^P_{jt} (s.d.)	MTS_{jt-1} (s.d.)	V^P_{jt} (s.d.)	MTS_{jt-1} (s.d.)
Score: VA-pass	0.306***		0.271***		0.325***	
	(0.0128)		(0.0234)		(0.0109)	
Score: Peers		0.353***		0.322***		0.375***
		(0.0212)		(0.0308)		(0.0213)
R-sq.	0.17	0.29	0.13	0.22	0.20	0.35
R-sq.: V_{jt}	0.14	-	0.10	-	0.16	-
Clusters	188	188	171	171	177	177
Students	2,370	2,370	883	883	1,487	1,487
Student-tracks	17,460	17,460	6,433	6,433	11,027	11,027

The table presents results from regressions of predicted value added, V^P_{jt}, and prior-year selectivity, MTS_{jt-1}, on households' quality scores. The regressions are similar to those in equation (3). However, the outcome variable is in standard deviations, rather than within-town quintiles. "R-sq." is the R-squared from explaining the listed outcome variable. "R-sq.: V_{jt}" adjusts for the fact that we observe only a prediction for value added, V^P_{jt}, not the true value, V_{jt}. See Appendix A1.4 for an explanation of how we calculate "R-sq.: V_{jt}". See Table 8 for additional details on sample construction. Standard errors are clustered by middle school.

Table A15: Average treatment effects on value added: robustness to alternative specifications

	(1) Baseline	(2) Final	(3) Change	(4) Final	(5) Change	(6) Final
Treated	-0.018	0.034	0.052**	0.048*	0.063**	0.054**
	(0.043)	(0.044)	(0.025)	(0.025)	(0.026)	(0.026)
Effect in percentage points	-0.21	0.40	0.61	0.56	0.74	0.63
Predicted pass rate	62.9	62.9	62.9	62.9	62.9	62.9
Controls:						
Indicator ranking a feasible track in the baseline survey	Y	Y	Y	Y	Y	Y
Value added of the most-preferred feasible track in the baseline survey				Y		Y
Fixed effects for middle school treatment-control pair					Y	Y
Clusters	78	78	78	78	78	78
Students	2,692	2,692	2,692	2,692	2,692	2,692

The table presents results from various versions of regression (4). In the first column, the outcome variable is the value added of the feasible track that the student ranked the highest in the baseline survey. The regression in this column is thus a balance test. In the columns labeled "Final", the outcome variable is the value added of the track the student attends. The results in Column 4 correspond to those in the first column of Table 10. Finally, in the columns labeled "Change", the outcome is the difference between the value added of the track the student attends and the value added of the feasible track that the student ranked the highest in the baseline survey. These columns thus represent difference-in-difference regressions. The covariates in each specification are listed under "Controls". Standard errors are clustered by the middle school treatment-control pairs within which we conducted the randomization.

Table A16: Effects on the accuracy of households' value added scores: low-achieving students

	x^{th} most-preferred track in the baseline						
	All tracks	Most-preferred	2nd-most-preferred	$\geq$ 3rd-most-preferred	$\geq$ 4th-most-preferred	$\geq$ 5th-most-preferred	$\geq$ 6th-most-preferred
Treated	-0.053	0.034	0.064	-0.122	-0.152*	-0.172*	-0.174*
	(0.057)	(0.098)	(0.109)	(0.076)	(0.082)	(0.091)	(0.098)
Mean abs. difference: baseline	1.09	0.88	1.13	1.20	1.24	1.31	1.31
Mean abs. difference: follow-up	1.19	0.94	1.21	1.27	1.33	1.35	1.34
Clusters	74	71	68	74	74	74	74
Students	569	411	314	511	461	416	383
Student-tracks	1,886	411	314	1,161	960	820	729

The table presents results analogous to those in Table 13. However, the sample is limited to students with transition scores in the bottom half of the national distribution. See the notes to Table 13 for additional details.

Table A17: Effects on the accuracy of households' value added scores: high-achieving students

	x^{th} most-preferred track in the baseline						
	All tracks	Most-preferred	2nd-most-preferred	$\geq$ 3rd-most-preferred	$\geq$ 4th-most-preferred	$\geq$ 5th-most-preferred	$\geq$ 6th-most-preferred
Treated	-0.050	0.025	-0.092	-0.067	-0.076	-0.098	-0.139*
	(0.031)	(0.046)	(0.064)	(0.044)	(0.056)	(0.065)	(0.077)
Mean abs. difference: baseline	0.99	0.96	1.05	0.98	1.02	1.06	1.07
Mean abs. difference: follow-up	0.89	0.82	0.94	0.91	0.94	0.96	0.97
Clusters	75	74	75	75	74	74	73
Students	956	852	648	841	673	551	485
Student-tracks	3,084	852	648	1,584	1,140	907	758

The table presents results analogous to those in Table 13. However, the sample is limited to students with transition scores in the top half of the national distribution. See the notes to Table 13 for additional details.

Table A18: Effects on the association between value added and households' preference rankings: low-achieving students

	x^{th} **most-preferred track in the baseline**					
	All tracks	Two most-preferred	≥ 3rd-most-preferred	≥ 4th-most-preferred	≥ 5th-most-preferred	≥ 6th-most-preferred
Value added: treated	0.095**	-0.131	0.132***	0.127***	0.123***	0.114***
	(0.040)	(0.127)	(0.039)	(0.040)	(0.040)	(0.039)
Association: baseline	0.291	0.018	0.128	0.062	0.020	-0.002
Association: follow-up	0.221	0.120	0.113	0.094	0.080	0.080
Clusters	74	74	74	74	74	74
Students	571	565	571	571	571	567
Student-tracks	7,167	1,084	6,083	5,633	5,259	4,968

The table presents results analogous to those in Table 14. However, the sample is limited to students with transition scores in the bottom half of the national distribution. See the notes to Table 14 for additional details.

Table A19: Effects on the association between value added and households' preference rankings: high-achieving students

	x^{th} **most-preferred track in the baseline**					
	All tracks	Two most-preferred	≥ 3rd-most-preferred	≥ 4th-most-preferred	≥ 5th-most-preferred	≥ 6th-most-preferred
Value added: treated	0.022	-0.040	0.024	0.033	0.038	0.045*
	(0.030)	(0.130)	(0.027)	(0.026)	(0.026)	(0.025)
Association: baseline	0.514	0.019	0.343	0.242	0.148	0.088
Association: follow-up	0.414	-0.043	0.264	0.203	0.180	0.167
Clusters	75	75	75	75	75	75
Students	962	958	962	962	962	947
Student-tracks	12,862	1,853	11,009	10,216	9,520	8,970

The table presents results analogous to those in Table 14. However, the sample is limited to students with transition scores in the top half of the national distribution. See the notes to Table 14 for additional details.

Table A20: Households' certainty in their preference rankings during the baseline survey

	All students	**Low-achieving**	**High-achieving**
Share who reported being:			
Uncertain	0.08	0.11	0.06
Somewhat certain	0.49	0.54	0.46
Very certain	0.43	0.36	0.48
Students	2,692	1,012	1,680

The table presents summary statistics on the share of households in the experimental sample who reported (in the baseline survey) that they were "uncertain", "somewhat certain", or "very certain" of their track preference rankings.

Table A21: Effects on beliefs and preference rankings by households' baseline certainty

	Uncert. or somewhat certain			**Very certain**		
	All students	Low-achieving	High-achieving	All students	Low-achieving	High-achieving
Panel A: Treatment effects on the accuracy of value added quality scores						
Treated	-0.139**	-0.208**	-0.107*	-0.037	0.038	-0.014
	(0.054)	(0.082)	(0.059)	(0.068)	(0.139)	(0.062)
Mean abs. difference: baseline	1.06	1.18	0.98	1.07	1.25	0.99
Mean abs. difference: follow-up	1.11	1.31	0.94	0.99	1.21	0.87
Clusters	76	74	69	75	54	73
Students	767	340	427	585	171	414
Student-tracks	1,605	773	832	1,140	388	752
Panel B: Treatment effects on preference rankings						
Value added: treated	0.084***	0.155***	0.048	0.028	0.086	0.000
	(0.031)	(0.054)	(0.038)	(0.030)	(0.073)	(0.031)
Association: baseline	0.249	0.108	0.341	0.295	0.165	0.345
Association: follow-up	0.197	0.094	0.262	0.234	0.149	0.265
Clusters	76	74	69	75	58	73
Students	861	368	493	672	203	469
Student-tracks	9,614	3,934	5,680	7,478	2,149	5,329

The table presents treatment effects on beliefs and preference rankings, distinguishing by a household's degree of certainty in their preference ranking at the time of the baseline survey. "Uncert. or somewhat certain" are households who reported being uncertain or somewhat certain of their preference rankings during this survey. "Very certain" are households who reported already being very certain. Panel *A* presents results from regression (5), as in Table 13. Panel *B* presents results from regression (6), as in Table 14. The sample is for tracks other than a household's two top baseline choices. See the notes to Tables 13 and 14 for additional details.

Table A22: Households' preferences for track attributes: all students

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Households' baseline quality scores:							
Location	0.276^{***}	0.292^{***}	0.277^{***}	0.281^{***}	0.289^{***}	0.276^{***}	0.261^{***}
	(0.0689)	(0.0688)	(0.0687)	(0.0655)	(0.0685)	(0.0691)	(0.0735)
Siblings and friends	0.336^{***}	0.326^{***}	0.319^{***}	0.344^{***}	0.311^{***}	0.334^{***}	0.330^{***}
	(0.0480)	(0.0477)	(0.0482)	(0.0479)	(0.0479)	(0.0484)	(0.0501)
Peer quality	0.344^{***}	0.317^{***}	0.318^{***}	0.380^{***}	0.298^{***}	0.302^{***}	0.201^{***}
	(0.0692)	(0.0659)	(0.0691)	(0.0665)	(0.0743)	(0.0697)	(0.0692)
Curriculum	0.931^{***}	0.789^{***}	0.877^{***}	0.986^{***}	0.763^{***}	0.929^{***}	0.914^{***}
	(0.0708)	(0.0691)	(0.0681)	(0.0729)	(0.0678)	(0.0708)	(0.0720)
VA: pass the bacc.	0.337^{***}				0.0130	0.341^{***}	0.321^{***}
	(0.0819)				(0.0829)	(0.0823)	(0.0827)
VA: college		0.519^{***}			0.347^{***}		
		(0.0730)			(0.0819)		
VA: wages			0.485^{***}		0.320^{***}		
			(0.0638)		(0.0694)		
Teacher quality				0.180^{**}	-0.0255		
				(0.0883)	(0.0800)		
True track characteristics:							
Math and science: female						0.112	0.0236
						(0.112)	(0.114)
Math and science: male						0.361^{***}	0.365^{***}
						(0.0942)	(0.0955)
Technical: female						-0.453^{*}	-0.321
						(0.272)	(0.289)
Technical: male						0.296	0.478^{**}
						(0.197)	(0.200)
Distance between transition score & cutoff: $\lvert TS_i - MTS_{jt} \rvert$							-0.557^{***}
							(0.0593)
R-sq.	0.33	0.33	0.33	0.32	0.34	0.33	0.35
Clusters	150	150	150	150	150	150	150
Students	1,170	1,157	1,151	1,168	1,137	1,170	1,170
Student-tracks	11,575	11,395	11,382	11,573	11,220	11,575	11,575

The table presents results from the preference model, equation, (7). The model is estimated by maximizing the log-likelihood corresponding to equation (8). "Households' baseline quality scores" are the quality scores that households assign to tracks in the baseline survey. "True track characteristics" are true values of track characteristics. The first four variables under "True track characteristics" are interactions between a track's curricular focus and a student's gender. Humanities: female and Humanities: male are the dropped groups. The sample is limited to students in experimental middle schools. Standard errors are clustered by middle school.

Table A23: Households' preferences for track attributes: low-achieving students

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Households' baseline quality scores:							
Location	0.234**	0.259**	0.222*	0.248**	0.254**	0.234**	0.227*
	(0.116)	(0.112)	(0.113)	(0.111)	(0.113)	(0.113)	(0.117)
Siblings and friends	0.335***	0.305***	0.316***	0.351***	0.289***	0.352***	0.350***
	(0.0802)	(0.0756)	(0.0809)	(0.0804)	(0.0772)	(0.0812)	(0.0819)
Peer quality	0.0642	0.103	0.0761	0.126	0.109	0.122	0.0978
	(0.0931)	(0.0878)	(0.0941)	(0.0903)	(0.106)	(0.0950)	(0.0924)
Curriculum	0.711***	0.611***	0.717***	0.774***	0.616***	0.662***	0.642***
	(0.117)	(0.109)	(0.113)	(0.114)	(0.107)	(0.111)	(0.115)
VA: pass the bacc.	0.265**				0.123	0.262**	0.263**
	(0.111)				(0.101)	(0.110)	(0.112)
VA: college		0.340***			0.170		
		(0.111)			(0.121)		
VA: wages			0.317***		0.227**		
			(0.0969)		(0.0951)		
Teacher quality				0.0406	-0.103		
				(0.132)	(0.116)		
True track characteristics:							
Math and science: female						-0.816***	-0.673***
						(0.177)	(0.172)
Math and science: male						-0.471***	-0.310**
						(0.141)	(0.151)
Technical: female						-0.582**	-0.491*
						(0.289)	(0.295)
Technical: male						0.205	0.329
						(0.243)	(0.262)
Distance between transition score & cutoff: $\lvert \text{TS}_i - \text{MTS}_{jt} \rvert$							-0.400***
							(0.0712)
R-sq.	0.21	0.21	0.21	0.20	0.22	0.23	0.24
Clusters	119	118	117	119	117	119	119
Students	394	382	387	394	376	394	394
Student-tracks	3,966	3,806	3,889	3,971	3,756	3,966	3,966

The table presents results analogous to those in Table A22, but for low-achieving students. See the notes to Table A22 for additional details.

Table A24: Households' preferences for track attributes: high-achieving students

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Households' baseline quality scores:							
Location	0.333***	0.343***	0.348***	0.340***	0.337***	0.348***	0.328***
	(0.0861)	(0.0897)	(0.0871)	(0.0821)	(0.0893)	(0.0872)	(0.0939)
Siblings and friends	0.322***	0.328***	0.307***	0.325***	0.315***	0.311***	0.318***
	(0.0633)	(0.0674)	(0.0639)	(0.0619)	(0.0673)	(0.0658)	(0.0692)
Peer quality	0.542***	0.473***	0.508***	0.562***	0.429***	0.413***	0.291***
	(0.0779)	(0.0815)	(0.0820)	(0.0805)	(0.0833)	(0.0839)	(0.0882)
Curriculum	1.085***	0.937***	1.009***	1.147***	0.886***	1.082***	1.091***
	(0.0917)	(0.0896)	(0.0830)	(0.0923)	(0.0917)	(0.0916)	(0.0952)
VA: pass the bacc.	0.430***				-0.0365	0.402***	0.367***
	(0.109)				(0.114)	(0.109)	(0.108)
VA: college		0.626***			0.447***		
		(0.0923)			(0.103)		
VA: wages			0.572***		0.342***		
			(0.0776)		(0.0892)		
Teacher quality				0.296***	0.0602		
				(0.0903)	(0.0957)		
True track characteristics:							
Math and science: female						0.448***	0.296**
						(0.121)	(0.129)
Math and science: male						0.907***	0.800***
						(0.135)	(0.143)
Technical: female						-0.607	-0.373
						(0.405)	(0.413)
Technical: male						-0.276	0.0889
						(0.291)	(0.243)
Distance between transition score & cutoff: $\lvert TS_i - MTS_{jt} \rvert$							-0.475***
							(0.0930)
R-sq.	0.40	0.41	0.40	0.40	0.41	0.41	0.43
Clusters	136	135	135	136	135	136	136
Students	776	775	764	774	761	776	776
Student-tracks	7,609	7,589	7,493	7,602	7,464	7,609	7,609

The table presents results analogous to those in Table A22, but for high-achieving students. See the notes to Table A22 for additional details.

Table A25: Households' preferences for track attributes: robustness to missing baseline quality scores

	All students			Low-achieving			High-achieving		
	No imputations	Scored all tracks	Imputations	No imputations	Scored all tracks	Imputations	No imputations	Scored all tracks	Imputations
Households' baseline quality scores:									
Location	0.261***	0.159*	0.311***	0.227*	0.140	0.299***	0.328***	0.210**	0.349***
	(0.0735)	(0.0881)	(0.0680)	(0.117)	(0.155)	(0.0928)	(0.0939)	(0.0993)	(0.0975)
Siblings and friends	0.330***	0.409***	0.323***	0.350***	0.399***	0.244***	0.318***	0.419***	0.366***
	(0.0501)	(0.0605)	(0.0448)	(0.0819)	(0.124)	(0.0759)	(0.0692)	(0.0774)	(0.0639)
Peer quality	0.201***	0.153*	0.157***	0.0978	-0.0174	0.0391	0.291***	0.348***	0.264***
	(0.0692)	(0.0901)	(0.0505)	(0.0924)	(0.117)	(0.0688)	(0.0882)	(0.117)	(0.0671)
Curriculum	0.914***	0.986***	1.048***	0.642***	0.851***	0.976***	1.091***	1.087***	1.086***
	(0.0720)	(0.0871)	(0.0542)	(0.115)	(0.127)	(0.0812)	(0.0952)	(0.104)	(0.0673)
VA: pass the bacc.	0.321***	0.296***	0.289***	0.263**	0.237*	0.281**	0.367***	0.310**	0.332***
	(0.0827)	(0.100)	(0.0784)	(0.112)	(0.132)	(0.115)	(0.108)	(0.140)	(0.0903)
True track characteristics:									
Math and science: female	0.0236	-0.204	-0.0289	-0.673***	-1.197***	-0.817***	0.296**	0.196	0.290***
	(0.114)	(0.158)	(0.0798)	(0.172)	(0.218)	(0.113)	(0.129)	(0.170)	(0.0986)
Math and science: male	0.365***	0.198*	0.393***	-0.310**	-0.430**	-0.413***	0.800***	0.690***	0.956***
	(0.0955)	(0.118)	(0.0735)	(0.151)	(0.195)	(0.0916)	(0.143)	(0.188)	(0.101)
Technical: female	-0.321	-0.728*	-0.648***	-0.491*	-1.020**	-0.603***	-0.373	-0.615	-1.059***
	(0.289)	(0.419)	(0.136)	(0.295)	(0.439)	(0.159)	(0.413)	(0.657)	(0.227)
Technical: male	0.478**	0.441*	0.0604	0.329	0.216	0.0187	0.0889	0.223	-0.281
	(0.200)	(0.266)	(0.123)	(0.262)	(0.345)	(0.142)	(0.243)	(0.272)	(0.196)
Dist. between trans. score	-0.557***	-0.546***	-0.642***	-0.400***	-0.460***	-0.382***	-0.475***	-0.395***	-0.594***
& cutoff: $\lvert \text{TS}_i - \text{MTS}_{jt} \rvert$	(0.0593)	(0.0658)	(0.0527)	(0.0712)	(0.101)	(0.0538)	(0.0930)	(0.0968)	(0.0850)
R-sq.	0.35	0.39	0.35	0.24	0.30	0.24	0.43	0.47	0.44
Clusters	150	96	169	119	72	163	136	83	157
Students	1,170	553	2,664	394	199	993	776	354	1,671
Student-tracks	11,575	7,238	34,920	3,966	2,663	12,649	7,609	4,575	22,271

The table shows whether the coefficients from the preference model, equation (7), are robust to missing values for households' baseline quality scores. "No imputations" are specifications that ignore missing scores. They correspond to Columns 7 of Tables A22-A24. "Scored all tracks" are specifications that restrict the sample to households without any missing scores. "Imputations" are specifications that impute the missing scores using a random forest, as described in Section 4.1. See the notes to Table A22 for additional details on estimating the preference model.

Table A26: The effect of beliefs on the value added of students' tracks: robustness

	Change in value added: $V_{i,CPP} - V_{i,CP}$		
	All students	Low-achieving	High-achieving
Panel A: Preferred			
Top 1	0.122	0.160	0.100
Top 2	0.116	0.131	0.107
Top 3	0.112	0.130	0.101
Top 4	0.116	0.132	0.107
Panel B: No transition score distance			
Top 1	0.151	0.170	0.139
Top 2	0.144	0.139	0.147
Top 3	0.143	0.138	0.145
Top 4	0.149	0.142	0.153
Panel C: Just quality scores			
Top 1	0.190	0.189	0.191
Top 2	0.191	0.168	0.204
Top 3	0.195	0.164	0.213
Top 4	0.207	0.170	0.229

The table presents the mean difference between $V_{i,CPP}$ and $V_{i,CP}$ for alternative specifications of the preference model. See Section 4.2 for definitions of "Preferred", "No transition score distance", and "Just quality scores". The columns represent models in which we use different numbers of choices in estimating the rank-ordered logit. "Top 1" uses just a household's top choice, "Top 2" uses the household's two top choices, and analogously for "Top 3" and "Top 4". The sample is similar to that for the experimental treatment effects from Section 3. However, it excludes 15 students who did not score above the admissions cutoffs for any of the tracks that existed in both 2018 and 2019. (These students were assigned to tracks that were newly created in 2019.)

A1 Adjusting for measurement error

This section describes the strategies that we use to adjust for measurement error.

A1.1 The standard deviation of $\mathbf{V}_{jt}$ based on $\widehat{\mathbf{V}}_{jt}$

With finite data, we obtain estimates of value added, $\hat{\mathrm{V}}_{jt}$, rather than the true values, V_{jt}. Suppose that the estimates are equal to the true values plus independent measurement error:

$$\hat{\mathrm{V}}_{jt} = \mathrm{V}_{jt} + \varepsilon_{jt},$$

with $\varepsilon_{jt} \perp \mathrm{V}_{jt}$. By independence, we have:

$$\mathrm{Var}[\hat{\mathrm{V}}_{jt}] = \mathrm{Var}[\mathrm{V}_{jt}] + \mathrm{Var}[\varepsilon_{jt}].$$

$\mathrm{Var}[\varepsilon_{jt}]$ can be estimated as the average of the squared standard errors for the $\hat{\mathrm{V}}_{jt}$ estimates. Thus, we can calculate the standard deviation of true value added as:

$$\mathrm{SD}[\mathrm{V}_{jt}] = \sqrt{\mathrm{Var}[\hat{\mathrm{V}}_{jt}] - \mathrm{Var}[\varepsilon_{jt}]}.$$

This involves simply subtracting the average squared standard error from the variance of estimated value added and taking the square root. For tracks in set $\mathcal{S}$, we thus use the finite-sample formula:

$$\mathrm{SD}[\mathrm{V}_{jt} | jt \in \mathcal{S}] = \left(\sum_{jt \in \mathcal{S}} \frac{N_{jt}}{N_{\mathcal{S}}} [(\hat{\mathrm{V}}_{jt} - \sum_{jt \in \mathcal{S}} \frac{N_{jt}}{N_{\mathcal{S}}} \hat{\mathrm{V}}_{jt})^2 - \varepsilon_{jt}^2] \right)^{1/2},$$

where N_{jt} is the number of students in track j in cohort t, $N_{\mathcal{S}}$ is the total number of students in $\mathcal{S}$, and ε_{jt}^2 is the squared standard error for $\hat{\mathrm{V}}_{jt}$.

A1.2 R-squared for predicting $\mathbf{V}_{jt}$ using $\mathbf{V}_{jt}^P$

In assessing the quality of the predictions, we are interested in how well they predict true value added, not estimated value added. Specifically, the metric that we want is R-squared in predicting true value added:

$$R^2 = 1 - \frac{\mathrm{E}[(\mathrm{V}_{jt} - \mathrm{V}_{jt}^P)^2]}{\mathrm{Var}[\mathrm{V}_{jt}]}.$$

$\mathrm{Var}[\mathrm{V}_{jt}]$ can be estimated using the approach explained in the previous subsection. The other term can be obtained via the following derivation:

$$\begin{aligned} \mathrm{E}[(\hat{\mathrm{V}}_{jt} - \mathrm{V}_{jt}^P)^2] &= \mathrm{E}[(\mathrm{V}_{jt} + \varepsilon_{jt} - \mathrm{V}_{jt}^P)^2] \\ &= \mathrm{E}[(\mathrm{V}_{jt} - \mathrm{V}_{jt}^P)^2] + \mathrm{Var}[\varepsilon_{jt}]. \\ \Rightarrow \quad \mathrm{E}[(\mathrm{V}_{jt} - \mathrm{V}_{jt}^P)^2] &= \mathrm{E}[(\hat{\mathrm{V}}_{jt} - \mathrm{V}_{jt}^P)^2] - \mathrm{Var}[\varepsilon_{jt}]. \end{aligned}$$

Thus, for tracks in set $\mathcal{S}$, we calculate R-squared using the following finite-sample formula:

$$R_{\mathcal{S}}^2 = 1 - \frac{\sum_{jt \in \mathcal{S}} \frac{N_{jt}}{N_{\mathcal{S}}} [(\hat{\mathrm{V}}_{jt} - \mathrm{V}_{jt}^P)^2 - \varepsilon_{jt}^2]}{\sum_{jt \in \mathcal{S}} \frac{N_{jt}}{N_{\mathcal{S}}} [(\hat{\mathrm{V}}_{jt} - \sum_{jt \in \mathcal{S}} \frac{N_{jt}}{N_{\mathcal{S}}} \hat{\mathrm{V}}_{jt})^2 - \varepsilon_{jt}^2]}.$$

Here, again, N_{jt} is the number of students in track j in cohort t, $N_{\mathcal{S}}$ is the total number of students in $\mathcal{S}$, and ε_{jt}^2 is the squared standard error for $\hat{\mathrm{V}}_{jt}$.

A1.3 The standard deviation of V_{jt} based on V_{jt}^P

For the 2015-2019 admissions cohorts, we cannot estimate value added and instead only have predictions, V_{jt}^P. We would like nonetheless to calculate the standard deviation of the true effects, $\mathrm{SD}[\mathrm{V}_{jt}]$, for these years. To do this, we assume that the true effects are equal to the predictions plus independent forecast error:

$$\mathrm{V}_{jt} = \mathrm{V}_{jt}^P + \vartheta_{jt},$$

with $\vartheta_{jt} \perp \mathrm{V}_{jt}^P$. We calculate the variance of V_{jt} by assuming that V_{jt}^P has an R-squared in predicting V_{jt} equal to that observed for the 2008-2014 cohorts (0.793, Table 2). Specifically, we use the following derivation:

$$\begin{aligned}
R^2 &= \frac{\mathrm{Var}[\mathrm{V}_{jt}] - \mathrm{E}[(\mathrm{V}_{jt} - \mathrm{V}_{jt}^P)^2]}{\mathrm{Var}[\mathrm{V}_{jt}]} \\
&= \frac{\mathrm{Var}[\mathrm{V}_{jt}] - \mathrm{Var}[\vartheta_{jt}]}{\mathrm{Var}[\mathrm{V}_{jt}]} \\
&= \frac{\mathrm{Var}[\mathrm{V}_{jt}^P]}{\mathrm{Var}[\mathrm{V}_{jt}]} \\
\Rightarrow \mathrm{SD}[\mathrm{V}_{jt}] &= \left(\frac{1}{R^2} \cdot \mathrm{Var}[\mathrm{V}_{jt}^P]\right)^{1/2}.
\end{aligned}$$

For tracks in set $\mathcal{S}$, we thus use the following finite-sample formula:

$$\mathrm{SD}[\mathrm{V}_{jt} | jt \in \mathcal{S}] = \left(\frac{\sum_{jt\in\mathcal{S}} \frac{N_{jt}}{N_{\mathcal{S}}}(\mathrm{V}_{jt}^P - \sum_{jt\in\mathcal{S}} \frac{N_{jt}}{N_{\mathcal{S}}}\mathrm{V}_{jt}^P)^2}{R_{0814}^2}\right)^{1/2},$$

where $R_{0814}^2 = 0.793$.

A1.4 R-squared for beliefs about V_{jt}, proxied by V_{jt}^P

In Section 2, we are interested in assessing how well households' beliefs about value added reflect a track's true value added. However, we observe only a track's predicted value added, not its true value added. Let p_{ij}^{V} be the fitted value from a regression of V_{jt}^P on s_{ij}^{V}. We calculate R-squared with respect to explaining true value added as follows. R-squared is:

$$R^2 = 1 - \frac{\mathrm{E}[(\mathrm{V}_{jt} - p_{ij}^{\mathrm{V}})^2]}{\mathrm{Var}[\mathrm{V}_{jt}]}.$$

$\mathrm{Var}[\mathrm{V}_{jt}]$ can be calculated using the approach described in Section A1.3. The other term is:

$$\begin{aligned}
\mathrm{E}[(\mathrm{V}_{jt} - p_{ij}^{\mathrm{V}})^2] &= \mathrm{E}[(\mathrm{V}_{jt}^P + \vartheta_{jt} - p_{ij}^{\mathrm{V}})^2] \\
&= \mathrm{E}[(\mathrm{V}_{jt}^P - p_{ij}^{\mathrm{V}})^2] + 2 \cdot \mathrm{E}[(\mathrm{V}_{jt}^P - p_{ij}^{\mathrm{V}}) \cdot \vartheta_{jt}] + \mathrm{Var}[\vartheta_{jt}] \\
&= \mathrm{E}[(\mathrm{V}_{jt}^P - p_{ij}^{\mathrm{V}})^2] - 2 \cdot \mathrm{E}[p_{ij}^{\mathrm{V}} \cdot \vartheta_{jt}] + \mathrm{Var}[\vartheta_{jt}].
\end{aligned}$$

$\mathrm{E}[(\mathrm{V}_{jt}^P - p_{ij}^{\mathrm{V}})^2]$ can be calculated from the data. $\mathrm{Var}[\vartheta_{jt}]$ can be calculated as:

$$\mathrm{Var}[\vartheta_{jt}] = \mathrm{Var}[\mathrm{V}_{jt}] - \mathrm{Var}[\mathrm{V}_{jt}^P].$$

Finally, we assume $\mathrm{E}[p^{\mathrm{V}}_{ij} \cdot \vartheta_{jt}] = 0$; that is, households' scores are not correlated with the unforecastable component of track value added.[74] Thus, we calculate R-squared as:

$$R^2 = 1 - \frac{\mathrm{E}[(\mathrm{V}^P_{jt} - p^{\mathrm{V}}_{ij})^2] + \mathrm{Var}[\vartheta_{jt}]}{\mathrm{Var}[\mathrm{V}^P_{jt}] + \mathrm{Var}[\vartheta_{jt}]}.$$

The finite-sample formula is:

$$R^2 = 1 - \frac{\frac{1}{J}\sum_i \sum_{j\in\mathcal{J}_i}[(\mathrm{V}^P_{jt} - p^{\mathrm{V}}_{ij})^2 + \frac{1-R^2_{0814}}{R^2_{0814}}(\mathrm{V}^P_{jt} - \frac{1}{J}\sum_i \sum_{j\in\mathcal{J}_i} \mathrm{V}^P_{jt})^2]}{\frac{1}{J}\sum_i \sum_{j\in\mathcal{J}_i}(\mathrm{V}^P_{jt} - \frac{1}{J}\sum_i \sum_{j\in\mathcal{J}_i} \mathrm{V}^P_{jt})^2 / R^2_{0814}}.$$

Here, i is a survey respondent, and $J \equiv \sum_i J_i$ is the sum of the number of tracks in each respondent's town.

A2 Validating value added

In this section, we use admissions-cutoff RDs to validate our selection-on-observables value added measures. We first define the admissions-cutoff RD and then explain how it can be used to compare value added estimates with causal effects. We finally present results.

A2.1 The admissions-cutoff RD

As discussed by Kirkeboen, Leuven, and Mogstad (2016) and Dahl, Rooth, and Stenberg (2020), the admissions-cutoff RD captures a complicated treatment effect. To see this, consider the admissions-cutoff RD for track c in town $l(c)$ in cohort t. Let $\mathcal{F}^c_t$ be the set of "fallback" tracks to track c in cohort t. These are tracks in town $l(c)$ with admissions cutoffs (or minimum transition scores) that in cohort t are lower than that of track c: $\mathrm{MTS}_{ft} < \mathrm{MTS}_{ct}\ \forall\ f \in \mathcal{F}^c_t$. Calculate a running variable, m^c_i, for student i in town $l(c)$ and cohort t as the difference between the student's transition score, TS_i, and the track's minimum transition score: $m^c_i \equiv \mathrm{TS}_i - \mathrm{MTS}_{ct}$. Next, let $z^c_i \in \{0, 1\}$ be an offer to attend track c, which the student receives if his or her value of the running variable is positive, $m^c_i > 0$.[75] Finally, let $d^c_{ij}(z)$ denote whether student i would attend track j under $z^c_i = z$.

In our setting, the only way receiving an admissions offer can change track attendance is by inducing the student to attend track c. As a result, students can be classified as one of two types. "Type-f compliers" prefer track c to all fallbacks, followed by track f. These students attend track f if they do not receive an offer and attend track c if they do: $d^c_{if}(0) = d^c_{ic}(1) = 1$. By contrast, "type-$f$ never-takers" prefer track f to track c. Thus, these students attend track f regardless of whether they receive an offer: $d^c_{if}(0) = d^c_{if}(1) = 1$.

The admissions-cutoff RD is the difference in observed outcomes between students who score just above and just below the cutoff. Consider the RD for admissions to track c for students in cohort t. For reasons that will be apparent later, consider the RD only for students who fall into

74. This assumption need not hold. However, we think it is reasonable based on the evidence with respect to households' scores for peer quality. Specifically, we found that households' scores for peer quality are not more predictive of a track's current-year minimum transition score than they are for the track's prior-year value. This suggests that households do not have information on trends in peer quality that is not observable to researchers. Our assumption is that this is also the case for value added.

75. Students with $m_{ic} = 0$ receive an offer with probability between 0 and 1. We cannot observe which of these students receive the offer and choose not to attend the cutoff track and which do not receive the offer. As a result, we exclude these students from the analysis.

group g. This quantity is:

$$\text{RD}_{cgt} \equiv \lim_{\Delta \to 0} \{ \text{E}[y_i | m_i^c = \Delta, z_i^c = 1, i \in \mathcal{I}_{l(c)gt}] - \text{E}[y_i | m_i^c = -\Delta, z_i^c = 0, i \in \mathcal{I}_{l(c)gt}] \}.$$

Here, y represents a generic outcome and $\mathcal{I}_{l(c)gt}$ is the set of students in town $l(c)$ in cohort t who are in group g. The RD can be rewritten in terms of potential outcomes. Let y_{ij} be the potential value of outcome y if student i attends track j. Also, for notational simplicity, omit the conditioning on $\mathcal{I}_{l(c)gt}$. Then the admissions-cutoff RD can be rewritten:

$$\begin{aligned}
&\lim_{\Delta \to 0} \{ \text{E}[y_i | m_i^c = \Delta, z_i^c = 1] - \text{E}[y_i | m_i^c = -\Delta, z_i^c = 0] \} \\
&= \lim_{\Delta \to 0} \{ \text{E}[y_{ic} \cdot d_{ic}^c(1) + \sum_f y_{if} \cdot d_{if}^c(1) | m_i^c = \Delta, z_i^c = 1] - \text{E}[\sum_f y_{if} \cdot d_{if}^c(0) | m_i^c = -\Delta, z_i^c = 0] \} \\
&= \text{E}[y_{ic} \cdot d_{ic}^c(1) + \sum_f y_{if} \cdot d_{if}^c(1) | m_i^c = 0] - \text{E}[\sum_f y_{if} \cdot d_{if}^c(0) | m_i^c = 0] \\
&= \text{E}[\sum_f (y_{ic} - y_{if}) \cdot \mathbb{1}\{d_{if}^c(0) = d_{ic}^c(1) = 1\} + \sum_f (y_{if} - y_{if}) \cdot \mathbb{1}\{d_{if}^c(0) = d_{if}^c(1) = 1\} | m_i^c = 0] \\
&= \text{E}[\sum_f (y_{ic} - y_{if}) \cdot \mathbb{1}\{d_{if}^c(0) = d_{ic}^c(1) = 1\} | m_i^c = 0] \\
&= \sum_f \text{E}[y_{ic} - y_{if} | d_{if}^c(0) = d_{ic}^c(1) = 1, m_i^c = 0] \cdot \Pr[d_{if}^c(0) = d_{ic}^c(1) = 1 | m_i^c = 0].
\end{aligned}$$

Define the type-f treatment effect as the difference in a student's potential outcome at the cutoff track relative to track f: $y_{ic} - y_{if}$. Then, in words, the admissions-cutoff RD is a weighted sum of type-f local average treatment effects for type-f compliers at the cutoff. Weights,

$$\omega_{fgt}^c \equiv \Pr[d_{if}^c(0) = d_{ic}^c(1) = 1 | m_i^c = 0, i \in \mathcal{I}_{l(c)gt}],$$

are equal to the share of students at the cutoff who are type-f compliers.[76]

A2.2 RDs on two outcomes

Our strategy for validating the value added measures involves calculating RDs on two different outcomes. First, we calculate the RD on a student's performance on the baccalaureate exam: p_i. This is the traditional admissions-cutoff RD. Second, we calculate an RD on the value added of the track j_i^* that the student attends: $\hat{\text{V}}_{j_i^* gt}$. These RDs capture the following quantities:

$$\begin{aligned}
\text{RD}_{cgt}^p &= \sum_f \text{E}[p_{ic} - p_{if} | d_{if}^c(0) = d_{ic}^c(1) = 1, m_i^c = 0, i \in \mathcal{I}_{l(c)gt}] \cdot \omega_{fgt}^c \\
\text{RD}_{cgt}^{\text{V}} &= \sum_f (\hat{\text{V}}_{cgt} - \hat{\text{V}}_{fgt}) \cdot \omega_{fgt}^c.
\end{aligned}$$

Here, p_{ij} is the potential baccalaureate outcome from attending track j, $\hat{\text{V}}_{jgt}$ is track j's value added for students in group g in cohort t, and ω_{fgt}^c are weights. If our value added measure does not suffer from bias and if tracks exert a constant treatment effect on students in group g and cohort t, then $\text{E}[p_{ic} - p_{if} | d_{if}^c(0) = d_{ic}^c(1) = 1, m_i^c = 0, i \in \mathcal{I}_{l(c)gt}] = \hat{\text{V}}_{cgt} - \hat{\text{V}}_{fgt}$. Thus, under these conditions—and with infinite data—RDs calculated on the two outcomes would be the same.

76. In the derivation, the first equality is due to the definition of potential outcomes. The second equality is from the RD identification proof of Hahn, Todd, and Van der Klaauw (2001). The third equality is due to the fact that students are either type-f compliers or type-f never-takers. The fourth equality is a simple manipulation, and the final equality is due to the law of total expectation.

A2.3 Multi-year RDs

In practice, a track-specific RD for students of a particular type in a single year will be very noisy. In order to gain statistical power, we calculate RDs that aggregate over each group and cohort. As shown in the appendix to Cattaneo et al. (2016), these RDs are:

$$\text{RD}_c^m = \sum_t \sum_g \text{RD}_{cgt}^m \cdot \Pr[i \in \mathcal{I}_{l(c)gt} | m_i^c = 0, i \in \mathcal{I}_{l(c)}]$$

for $m \in \{p, \text{V}\}$. These RDs maintain the same structure as in the previous subsection. As before, if the value added measure is valid, then RDs on the two outcomes (p and V) should be equal.

A2.4 Estimation

We estimate the RDs using a local linear regression with a uniform kernel. We use a bandwidth equal to one standard deviation from the nation-wide transition score distribution in each cohort. Specifically for each cutoff c, we run the regression:

$$y_i = \lambda_t + \lambda_1 \cdot m_i^c + \mathbb{1}\{m_i^c \geq 0\} \cdot (\phi_0 + \phi_1 \cdot m_i^c) + u_i$$

for $i \in \mathcal{I}_{l(c)}$ with $|m_i^c| \leq 1$ and $m_i^c \neq 0$. Here, λ_t is an intercept that varies by cohort, $\lambda_1 \cdot m_i^c + \mathbb{1}\{m_i^c \geq 0\} \cdot \phi_1 \cdot m_i^c$ is a linear spline in the running variable, and ϕ_0 is the RD treatment effect (RD_c^y for outcome y).

A2.5 Comparing RDs

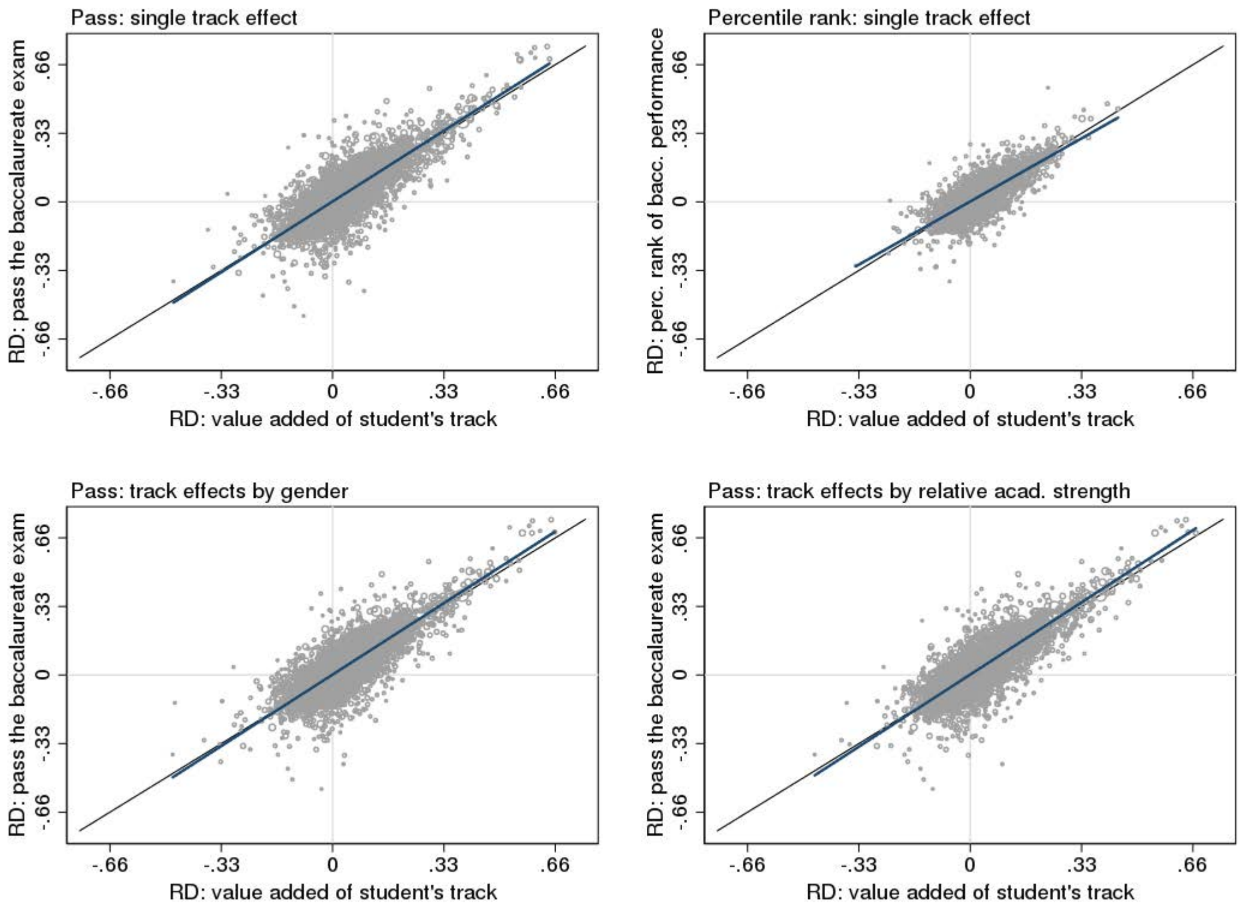
We then compare the RDs for the two outcomes. Figure A8 plots estimated RDs for baccalaureate outcomes, $\hat{\text{RD}}_c^p$, versus those for the value added of students' tracks, $\hat{\text{RD}}_c^{\text{V}}$. The figure includes plots for a variety of combinations of baccalaureate outcomes and value added measures. The top-left, bottom-left, and bottom-right plots are for the baccalaureate outcome of whether the student passes the exam. These plots use value added measures of, respectively, a single track effect by year on the probability of passing, track effects on this probability that vary by year and gender, and those that vary by year and relative academic strength. The top-right plot is for the percentile rank of a student's exam performance and uses a value added measure of a single track effect by year on this alternative baccalaureate outcome.

In the plots, each dot represents a different cutoff. The grey diagonal line is a 45-degree line, and the blue line is a line of best fit from a linear regression. If the RDs on value added are an unbiased predictor of the RDs on baccalaureate outcomes, then the best fit line will equal the 45-degree line. If the RDs on the two outcomes were always equal, then all the dots would fall on the 45-degree line. It can be seen that in each plot the best fit line closely matches the 45-degree line, but that the dots exhibit dispersion around these lines. Importantly, much of this dispersion could be due to noise in estimating the RDs.

In Table A27, we assess the similarity of the RDs using an approach that allows us to account for noise. Specifically, we calculate R-squared from predicting RDs on baccalaureate outcomes using RDs on value added. We present two different versions of R-squared. The first version is R-squared for the estimated RDs. This quantity is presented in the first column of the table. It captures the dispersion represented in Figure A8 and does not account for noise. It is:

$$R_{\text{raw}}^2 = 1 - \frac{\sum_c \frac{N_c}{N} (\hat{\text{RD}}_c^p - \hat{\text{RD}}_c^{\text{V}})^2}{\sum_c \frac{N_c}{N} (\hat{\text{RD}}_c^p - \sum_c \frac{N_c}{N} \hat{\text{RD}}_c^p)^2}. \tag{10}$$

Figure A8: Admissions-cutoff RDs



The figure plots estimates of admissions-cutoff RDs on baccalaureate outcomes, $\hat{\text{RD}}_c^p$, versus those on the value added of students' tracks, $\hat{\text{RD}}_c^{\text{V}}$. The grey line is a 45 degree line, and the blue line is a best fit from a linear regression. Values are weighted by the number of students with transition scores within 1 standard deviation of the cutoff. See Section A2.5 for additional details.

Here, N_c is the number of students in the estimation sample for cutoff c (i.e., $i \in \mathcal{I}_{l(c)}$ with $|m_i^c| \leq 1$ and $m_i^c \neq 0$), and N is the sum of the number of students in each cutoff's estimation sample. Next, the second version is R-squared for the true RDs. This quantity is presented in the second column of Table A27. It is calculated by purging R^2_{raw} of measurement error. Specifically, write $\hat{\text{RD}}_c^y = \text{RD}_c^y + \varepsilon_c^y$, where ε_c^y is measurement error. The true (or adjusted) R-squared is:

$$R^2_{\text{adj.}} = 1 - \frac{\sum_c \frac{N_c}{N}[(\hat{\text{RD}}_c^p - \hat{\text{RD}}_c^{\text{V}})^2 - (\varepsilon_c^p)^2 + 2 \cdot \varepsilon_c^p \cdot \varepsilon_c^{\text{V}} - (\varepsilon_c^{\text{V}})^2]}{\sum_c \frac{N_c}{N}[(\hat{\text{RD}}_c^p - \sum_c \frac{N_c}{N}\hat{\text{RD}}_c^p)^2 - (\varepsilon_c^p)^2]}, \tag{11}$$

Here, $(\varepsilon_c^y)^2$ is the squared standard error for the estimated RD for cutoff c on outcome y, and $\varepsilon_c^p \cdot \varepsilon_c^{\text{V}}$ is the covariance in the measurement error for the cutoff's RDs across the two outcomes.[77]

The values in Table A27 suggest that RDs on value added are highly similar to those on baccalaureate outcomes. Further, they indicate that much of the dispersion seen in Figure A8

77. As described in Appendix C.3.2 of Chandra et al. (2016), the covariance term can be calculated by stacking the RD regression equations for each outcome for cutoff c and selecting the appropriate element of the variance-covariance matrix.

Table A27: Comparing admissions-cutoff RDs

Value added measure	R-squared		Cutoffs	Student-cutoffs
	Raw	*Adjusted*		
Pass the exam:				
All	0.748	0.994	10,333	24,194,112
Gender	0.754	0.996	10,333	24,194,112
Relative academic strength	0.748	0.986	10,333	24,194,112
Percentile rank of exam performance	0.701	0.964	10,333	24,194,112

The table presents R-squared from explaining admissions-cutoff RDs on baccalaureate outcomes, RD^{p}_{ct}, using those on the value added of students' tracks, $\text{RD}^{\text{V}}_{ct}$. Raw R-squared is defined in equation (10). Adjusted R-squared is defined in equation (11). Values are weighted by the number of students with transition scores within 1 standard deviation of the cutoff.

is due to measurement error. The values in the first row of the table are for the baccalaureate outcome of passing the exam and a value added measure of a single track effect on this outcome. For this specification, the estimated RDs on value added explain 74.8% of the estimated RDs on passing. However, most of the unexplained variation is noise. After adjusting for measurement error, the R-squared jumps to 0.994. The next two rows keep the same baccalaureate outcome but use value added measures that vary by student type. They show that allowing value added to vary by a student's gender generates a slight improvement (adjusted R-squared of 0.996), while allowing it to vary by the student's relative academic strength causes a slight deterioration (adjusted R-squared of 0.986). The final row is for an alternative baccalaureate outcome. It uses the percentile rank of the student's exam performance, and the value added measure is a single track effect on this outcome. For this outcome, the adjusted R-squared remains extremely high (0.964).

A2.6 Comparing RDs using an IV approach

The second strategy that we use to compare the RDs is an adaptation of the procedure developed by Angrist et al. (2017). This involves using the admissions offers that students receive due to scoring above a cutoff as instruments in a regression of p_i (a baccalaureate outcome) on $\hat{\text{V}}_{j_i^* gt}$ (the value added of the student's track). In this regression, we stack observations for all cutoffs and include cutoff-year fixed effects and cutoff-specific controls for the running variable. The admissions offers generate exogenous variation in $\hat{\text{V}}_{j_i^* gt}$ due to the fact that some students who receive an offer attend the associated track. If on average over all cutoffs, an increase in value added due to scoring above a cutoff improves outcomes by the same amount, then the coefficient on $\hat{\text{V}}_{j_i^* gt}$ will equal 1. In addition, Angrist et al. (2017) note that a researcher can use an over-identification test to examine whether each cutoff would generate the same coefficient on its own. Thus, the procedure allows a researcher to both quantify the average bias and to examine whether there is heterogeneity in the bias across cutoffs.

Table A28 (page 76) presents results. With our large dataset, this exercise is computationally burdensome. Thus, we provide results only for our main value added measure of a track-year effect on the probability of passing the baccalaureate exam. In addition, we divide the cutoffs into ten random groups and calculate results separately for each group. The results in the table indicate that value added is unbiased on average, with IV coefficients that hover around 1. However, the results for the over-identification test generally allow us to reject that each cutoff would generate the same IV coefficient if used on its own.

In short, the results from our validation exercises indicate that our value added measures closely approximate causal effects. However, statistically speaking, the amount of bias is larger than what would be predicted by noise alone.

Table A28: Testing for bias using the Angrist et al. (2017) IV strategy

	Group									
	1	2	3	4	5	6	7	8	9	10
IV coefficient	1.03	1.05	1.05	1.02	1.05	1.07	1.01	1.09	0.97	1.04
	(0.019)	(0.019)	(0.020)	(0.019)	(0.020)	(0.019)	(0.020)	(0.021)	(0.020)	(0.020)
First-stage F statistic	29.4	28.4	26.8	27.2	25.1	26.7	24.1	24.7	25.7	24.9
Bias										
Wald statistic	2.13	8.24	5.45	1.65	5.89	12.8	0.11	18.1	1.78	4.35
p-value	0.144	0.004	0.020	0.199	0.015	0.000	0.744	0.000	0.182	0.037
Overidentification										
Hansen J statistic	1,339	1,430	1,440	1,556	1,410	1,453	1,459	1,415	1,425	1,505
degrees of freedom	1,033	1,032	1,032	1,033	1,032	1,032	1,033	1,032	1,032	1,032
p-value	¡0.001	¡0.001	¡0.001	¡0.001	¡0.001	¡0.001	¡0.001	¡0.001	¡0.001	¡0.001
Student-cutoffs	2,452,742	2,435,325	2,435,572	2,390,494	2,507,486	2,451,076	2,387,812	2,501,369	2,430,272	2,201,964

The table presents results from the strategy of Angrist et al. (2017), described in Section A2. Results are for the value added measure of a track-year effect on the probability of passing the baccalaureate exam. Cutoffs are divided into ten random groups, and results are presented separately for these groups. The "IV coefficient" is the coefficient on $\hat{V}_{j_i^* gt}$ in an instrumental variables regression of p_i on $\hat{V}_{j_i^* gt}$, cutoff-year fixed effects, and cutoff-specific controls for the running variable. "Bias" is a Wald test that the IV coefficient is equal to 1. "Overidentification" is the Sargan-Hansen test of over-identifying restrictions. It tests whether each instrument would generate the same IV coefficient if used on its own. The IV regression is estimated using two-stage least squares. All values are robust to heteroskedasticity.

A3 Details on the randomization

We conducted a clustered randomization that involved matching pairs of middle schools within towns, and then randomizing within pairs. We began with a target sample of 228 middle schools in 49 towns. Schools in the sample had either one or two classrooms.

We first conducted the randomization for the two-class schools. In our sample, towns had no more than two two-class schools. There were 25 towns with two two-class schools. In these towns, we paired the two-class schools and randomly selected one for treatment. Next, in two towns, there was one two-class school. In one of these towns, there was one two-class school and one one-class school. These were matched into a pair, with one school randomly assigned to treatment. In the other town, there was one two class-school and two one-class schools. These were matched into a three-school pair, with the one two-class school and the two one-class schools being restricted to have a different randomly assigned treatment.

We next randomized the one-class schools. We calculated the Mahalanobis distance among all one-class schools in each town, using as covariates: i) the number of students in the school, ii) the average transition score of students in the school, iii) the share of students in the school that were assigned to academic high-school tracks, and iv) the share of students in the school that were assigned to tracks with Romanian language of instruction. We then selected treatment-control pairs sequentially. In each iteration of the matching algorithm, we created a pair by selecting the two schools in the town with the lowest distance among the schools that did not already form part of a pair. Finally, we randomly assigned one element of the pair to treatment.

One complication for the matching algorithm was that some towns had an odd number of one-class schools. In these towns, we stopped the matching algorithm when there were three remaining one-class schools. We calculated the Mahalanobis distance of the covariates for each school in the triple to the average of the covariates of the other two schools in the triple. We split the triple into two groups based on which school had the lowest Mahalanobis distance to the average of the two other schools. We then randomly assigned one of the two groups in the triple to treatment.

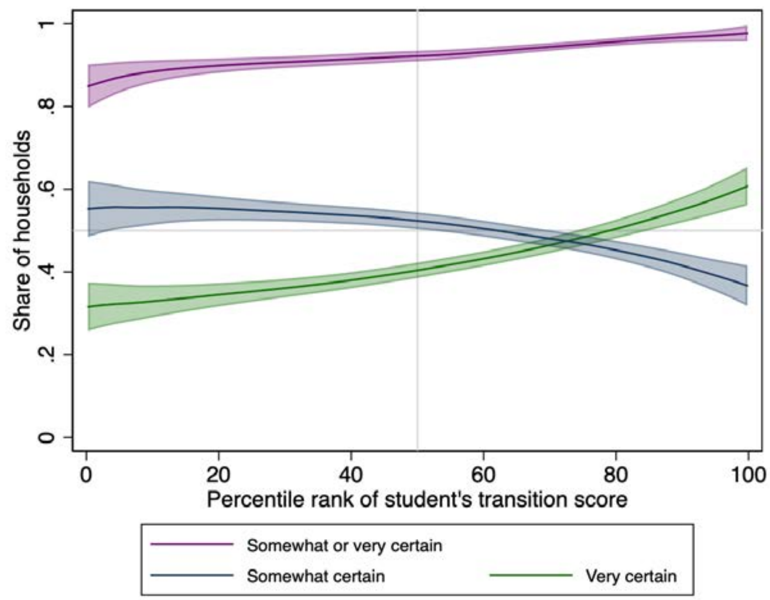
In the target sample, the treatment and control groups each consisted of 114 schools. Some of these schools did not agree to participate in the survey, and in some schools there were issues with implementation of the survey. When there was an issue with one school in a matched pair, we dropped the entire pair. Thus, the final experimental sample included 173 middle schools in 46 towns, of which 88 middle schools were in the treatment group and 85 were in the control group.

A4 Households' preference rankings and quality scores

This section presents stylized facts about how households ranked and scored tracks. We are interested in whether households' track preference rankings truthfully reflect their preferences and whether the household's town is the appropriate unit of analysis for its choice set.

First, we investigate how carefully households had thought about their track preference rankings at the time of the baseline survey. The survey occurred about a month before households were required to submit their rankings. Further, it occurred in information sessions that are used to explain the admissions process. It is thus possible that during the survey households had not yet seriously considered their options. We find that this is not the case. In particular, a large share of households self-report already being certain of their preference rankings. Table 4 (page 16) shows that 43 percent of households report being "very certain" and 50 percent report being "somewhat certain".[78]

Figure A9: Households' certainty about their baseline track preference rankings



The figure presents information on the share of survey households who, at the time of the baseline survey, report being somewhat certain or very certain of their track preference rankings. The category "Somewhat or very certain" is the sum of the categories "Somewhat certain" and "Very certain". The lines represent local linear regressions of the listed variables on the percentile rank of a student's transition score.

Figure A9 (page 78) examines whether there is any point in the distribution of student achievement where a large fraction of households report being uncertain. The figure plots the shares of households who are somewhat certain, very certain, or either somewhat or very certain against the national percentile rank of the student's transition score.[79] It indicates that households with low-achieving students are more likely to be somewhat certain, while those with high-achieving students are more likely to be very certain. However, over the entire transition score distribution, more than 85 percent and as many as 98 percent of households are either somewhat or very certain.

Second, we inspect the share of tracks that households rank and score. As noted in Section 0.4.2, households tend to assign preference ranks and quality scores to only a subset of the tracks in their towns. Namely, households rank an average of 43 percent of tracks and score an average of 37 percent on peer quality and 35 percent on value added on passing the baccalaureate exam (Table 4). Table A29 (page 79) provides additional statistics on this topic. Its first column shows that most households rank a significant share of tracks. 63 percent rank over a quarter of tracks, and 22 percent rank over three quarters. Only 8 percent of households rank no tracks. The fourth column displays the share of tracks that a household scores.[80] It can be seen that this distribution is more bimodal than that for the share of tracks that a household ranks, with most households assigning scores to either a small or large share of the tracks in their towns. Specifically, 60 percent

78. The remaining 7 percent chose either "somewhat uncertain" or "very uncertain".

79. As noted in Section 0.4.2, we were only able to match 83 percent of students in the baseline survey with the administrative data. For the 17 percent of students who were not matched, we cannot obtain official transition scores. Fortunately, in the survey we asked respondents to predict their children's scores. We impute missing transition scores using these predictions, and thus have scores for all but 152 students.

80. We define a household as scoring a track if it assigns scores for both peer quality and value added on passing.

of households score a quarter of the tracks or fewer, with 37 percent scoring no tracks. On the other hand, 28 percent score over three quarters of the tracks.

Table A29: Summary statistics on the share of tracks that a household ranks and/or scores

	Included in preference ranking			**Scored on pass and peers**		
	All students	*Low-achieving*	*High-achieving*	*All students*	*Low-achieving*	*High-achieving*
Mean share of tracks ranked / scored	0.44	0.42	0.45	0.36	0.33	0.38
Fraction of households ranking / scoring:						
No tracks	0.08	0.09	0.06	0.37	0.43	0.32
1-25 percent	0.30	0.33	0.27	0.23	0.22	0.24
26-50 percent	0.30	0.28	0.31	0.08	0.06	0.10
51-75 percent	0.11	0.09	0.12	0.04	0.04	0.04
> 75 percent	0.22	0.21	0.23	0.28	0.26	0.30
Number of students	3,746	1,564	2,182	3,746	1,564	2,182

The table describes the share of tracks that a survey household scores and/or ranks. A household is said to score a track if it assigns scores for both value added on passing the baccalaureate exam ("pass") and peer quality ("peers"). A household is defined as ranking a track if it includes it in its preference ranking. "Mean share of tracks scored/ranked" is the average share of tracks that a household scores or ranks. The remaining rows display the fraction of households that score or rank none of the tracks in their towns, 1-25 percent, 26-50 percent, 51-75 percent, and more than 75 percent. Low-achieving (high-achieving) students are those with transition scores in the bottom (top) half of the national distribution. The sample drops 152 students with missing values for both official transition score and respondent's prediction of transition score. As a result of dropping these students, values for "Mean share of tracks ranked / scored" differ from those in Table 4 (page 16). Households' choice sets exclude tracks that newly created in 2019.

Figure A10 (page 80) and the remaining columns of Table A29 show how the share of tracks ranked or scored varies with the student's transition score. They reveal that households with low-achieving students are more likely to not assign scores to any track. However, behavior is otherwise relatively similar across the transition score distribution.

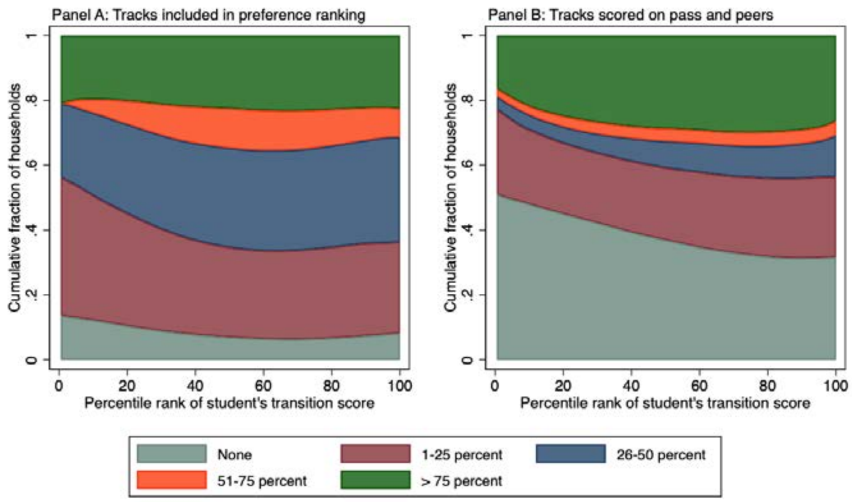
The third question we ask is whether households with low- and high-achieving children differ in the selectivity of the tracks that they rank and score. The fact that the serial dictatorship is incentive compatible means that it is weakly dominant for a household to assign a rank to each track that it prefers to the outside option of vocational school. Moreover, the dominance is strict if there is a non-zero chance that the student will be assigned to the track, conditional on ranking it. In practice, however, households may find it costly to evaluate tracks. As a result, they may consider only tracks that they believe their children are likely to attend. In this case, the relevant choice set for a household would not be the full set of tracks in a town, but instead a subset of them by selectivity, with the particular subset depending on the student's achievement.[81]

In contrast, we find that households rank and score tracks from across the selectivity distribution. Figure A11 summarizes the composition of the tracks that a household ranks and/or scores by the tracks' prior-year minimum transition scores.[82] The figure shows that households with low-achieving students assign ranks and/or score to tracks from each within-town quintile of selectivity at almost uniform rates. Households with high-achieving students are more likely to rank and/or score selective tracks than non-selective ones—among this group, on average about

81. For instance, a household with a low-achieving child may not rank and/or score highly selective tracks. Similarly, a household with a high-achieving child may not rank and/or score non-selective tracks if it cares about peer quality and thinks its child will be admitted to one of its top choices.

82. We use the prior-year (2018) value of minimum transition score as the measure of selectivity because this variable is observable by households at the time of the information sessions. Households can remember it from the 2018 allocation. In addition, it is published by the government when it announces the year's list of available tracks. Thus, it is likely to be more closely related to a household's beliefs about track selectivity than is the current-year (2019) version. Furthermore, the 2019 version may be influenced by our experiment.

Figure A10: The share of tracks that a household ranks and/or scores by student transition score



The figure shows how the share of tracks that a survey household ranks and/or scores varies with the student's transition score. Specifically, households are assigned to groups based on whether they ranked and/or scored none of the tracks in their towns, 1-25 percent of the tracks, 26-50 percent, 51-75 percent, or more than 75 percent. The colored areas in the figure represent the fraction of households in each group. The dividing lines are calculated using local linear regressions of indicators for group membership on the national percentile rank of student's transition score. See the notes to Table A29 for details on sample construction.

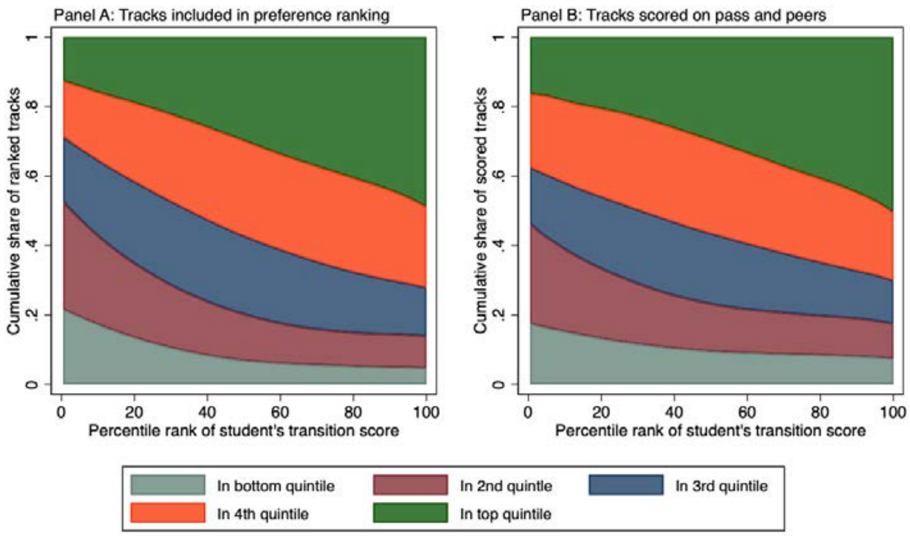
40 percent of the tracks that a household ranks and/or scores fall into the top quintile. However, these households still rank and score significant fractions of non-selective tracks, with on average about 20 percent of ranked and/or scored tracks coming from the bottom two quintiles.

Finally, we examine a household's behavior with respect to its most-preferred tracks. We explore whether households tend to select "reach" tracks that they do not believe will be feasible, or whether they instead choose options that they expect their child to be eligible to attend.[83] We also assess the accuracy of households' expectations. Figure A12 provides the results. The first panel of the figure shows results for a household's highest-ranked track, and the second panel shows them for its two highest-ranked tracks. The figure shows that a large majority of households select options that they expect to be feasible. Depending on the student's transition score, between 84 and 94 percent of households believe their child will be admitted to their most-preferred track and between 93 and 97 percent think their child will be admitted to at least one of their two most-preferred tracks. Consistent with their expectations, households with lower-performing students choose less selective tracks than do those with higher-performing ones. However, households tend to be overly optimistic about track feasibility. For students with transition scores in the bottom half of the distribution, only 40 percent would have been eligible for their top-ranked track based on the track's prior-year minimum transition score. Similarly, only 54 percent would have been eligible for one of their top two choices. Not until about the 70th percentile of the transition score distribution does the probability that a student is eligible catch up to households' expectations.

The evidence presented in this subsection thus broadly supports the assumptions that households truthfully rank tracks and that the town is the appropriate unit for defining a choice set. It also suggests that households overestimate the academic performance of their children when determining their track preference rankings.

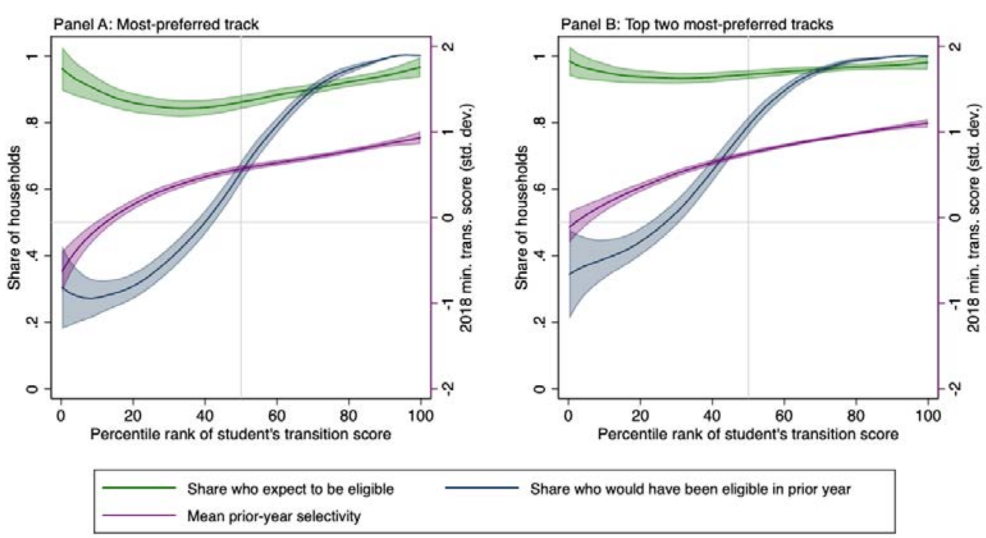
83. We highlight that the latter pattern of behavior does not imply that households are deviating from truthful revelation of preferences. Notably, it could be that households prefer tracks that are a "good fit" in terms of their child's achievement level.

Figure A11: The selectivity of ranked and/or scored tracks by student transition score



The figure provides information on the selectivity of the tracks that survey households consider. Specifically, among either the tracks that a household includes in its preference ranking (Panel A) or among those that the household scores on both peer quality and value added on passing the baccalaureate exam (Panel B), the figure summarizes the shares of tracks that fall into each within-town quintile of 2018 minimum transition score, MTS_{jt-1}. The dividing lines in the figure represent local linear regressions of a household's cumulative shares against the national percentile rank of the student's transition score. The sample drops respondents who didn't score any tracks on both peer quality and value added on passing the baccalaureate exam, as well as those who didn't include any tracks in their preference rankings. Respondents' choice sets exclude tracks that were newly created in 2019.

Figure A12: Summary statistics on a household's most-preferred tracks by student transition score



The figure provides information on the selectivity of households' most-preferred tracks. Panel A refers to each household's top-ranked track, and Panel B to its two highest-ranked tracks. The green lines describe the shares of households that expect their child to be eligible for their most-preferred track (Panel A) or for one of their two most-preferred tracks (Panel B). The blue lines describe the shares of households whose children would have been eligible based on selectivity in 2018. A household is in this latter group if the student's transition score is greater than or equal to a track's 2018 minimum transition score, MTS_{jt-1}. The purple line describes the mean 2018 selectivity of a household's most-preferred tracks in standard deviations, $\widetilde{\mathrm{MTS}}_{jt-1}$. We drop respondents who did not score any tracks on both peer quality and value added, as well as those who did not include any tracks in their preference rankings.

A5 Testing for informational spillovers

In this section, we investigate whether the experiment suffered from informational spillovers. In particular, it's possible that treated households shared the information on track value added with households in the control group. If so, treatment effects would be biased toward zero.

Our experimental set-up included factors that both decreased and increased the likelihood of spillovers. First, we tried to limit spillovers by visiting only a fraction of middle schools in each town. Across towns, we visited an average of 11% of middle schools and a maximum of 29%. On the other hand, our method for distributing information potentially facilitated spillovers. We provided treated households with informational flyers, which we allowed households to keep. Households may have given these flyers to others in their towns.

We test for spillovers by examining whether treatment effects differ in towns in which we visited a smaller or larger fraction of middle schools. If there are spillovers, then, all else equal, treatment effects should be smaller in towns where this fraction is larger. In these towns, there is more interaction between treated and control households and more opportunity for the information to be shared. Importantly, our test will be confounded if there are third factors that are correlated with both the fraction of schools that we visited and the magnitude of treatment effects. We think this is unlikely to be the case. In particular, we decided what fraction of schools to survey based on i) the share of schools with at least 15 students and ii) logistical considerations, such as whether the date of a school's information session was convenient for our surveyors. These variables have no obvious relationship with the magnitude of treatment effects, except via their effect on spillovers.

Specifically, to conduct the test, we partition the sample based on whether a student's town is in the bottom or top half by the share of schools surveyed. We then calculate treatment effects on the value added of students' tracks (regression (4)) separately for these two groups.

Table A30: Testing for spillovers in treatment effects

	All students			**Low-achieving**			**Low-achieving and ineligible**		
	All towns	Bottom	Top	All towns	Bottom	Top	All towns	Bottom	Top
Treated	0.048*	0.056*	0.037	0.121**	0.122*	0.118*	0.204***	0.184**	0.223**
	(0.025)	(0.033)	(0.039)	(0.049)	(0.072)	(0.067)	(0.069)	(0.084)	(0.109)
Clusters	78	37	41	78	37	41	76	36	40
Students	2,692	1,407	1,285	1,012	462	550	533	266	267

The table presents results from regression (4) for subsets of students by whether a student's town was in the bottom ("Bottom") or top ("Top") half by the share of middle schools surveyed. The columns for "All towns" replicate results from Section 3.1. "Low-achieving" are students with transition scores in the bottom half of the national distribution. "Low-achieving and ineligible" are low-achieving students who did not gain admission to either of their two top baseline choices. See the notes to Table 10 for additional details on the regressions.

The results are presented in Table A30. In the table, the first three columns are for the full sample of students and the remaining columns are for sub-samples that were found to have non-zero treatment effects in Section 3.1. The columns labeled "All towns" replicate results from Section 3.1, while the other columns distinguish between the share of schools surveyed. The results in the table provide no evidence of spillovers. Instead, treatment effects are shown to be similar in magnitude for each group of towns.