Why do households leave school value added “on the table”? The roles of information and preferences

Robert Ainsworth*

Rajeev Dehejia+

Cristian Pop-Eleches±

Miguel Urquiola±

April 2022

Abstract

Administrative data reveal that Romanian households could choose schools with 1 s.d. worth of additional value added. Why do households leave value added “on the table”? We study two possibilities: households may lack *information* about value added, or they may have *preferences* for other school traits. First, we elicit households' beliefs about the value added of the schools in their town. These beliefs explain less than one fifth of the variation in measured value added. We then inform randomly selected households about schools' value added. This improves the accuracy of their beliefs and leads them to choose higher value added schools. The effect is stronger for low-achieving students and for students not admitted to their top two choices. Finally, we use a discrete choice model to estimate households' preferences for school attributes (as they perceive them). Households have strong preferences for peer quality and curriculum in addition to value added. Thus, households would not “max out” on value added even under perfect information: this only eliminates one quarter of the value added households leave unexploited.

For useful comments, we thank Matt Backus, Peter Bergman, Michael Best, Sarah Cohodes, David Figlio, John Friedman, Dong Woo Hahm, Rema Hanna, Douglas Harris, Caroline Hoxby, Kirabo Jackson, Francisco Martorell, Karthik Muralidharan, Suresh Naidu, Christopher Neilson, Daniele Paserman, Jonah Rockoff, Richard Romano, Bennett Smith-Worthington, Doug Staiger, and seminar participants at APPAM, Berlin, Bocconi, Boston, Bristol, Davis, ERMAS, EUI, Florida, NBER Summer Institute, Nottingham, and Princeton. 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 under number AEARCTR-0004496.

* University of Florida. + New York University. ± Columbia University.

Friedman (1955) argued that giving households freedom to choose schools would improve their children's learning. This simple idea underlies numerous programs that expand school choice. Yet research yields surprisingly mixed evidence on their effect. For example, voucher experiments show that choice can impact students' 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 evidence on the effects of choice among selective schools, Beuermann and Jackson (2020) state: "the lack of robust achievement effects of attending schools that parents prefer is something of a puzzle."

We investigate two possible explanations for this puzzle. First, a lack of *information* may prevent households from choosing schools with high value added. 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 a 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. Second, households may have *preferences* that lead them to prioritize school attributes other than value added. This possibility arises if school quality is multidimensional (Beuermann et. al 2019, Riehl et al. 2019).

Distinguishing between preferences and information is important. If information is the obstacle, then making it available would improve households' choices and, possibly, spur schools 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 to invest in other, perhaps less desirable, dimensions of quality (Rothstein 2006).

We explore the distinction between preferences and information by studying high school admissions in Romania. To conduct our analysis, we obtained administrative data on fifteen admissions cohorts. We also implemented surveys and ran an informational experiment.

Two features of the Romanian school system make it an advantageous setting. 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." These tests allow us to calculate schools' academic value added. The particular outcome we focus on—performance on the baccalaureate exam—is of central importance to Romanian students.

The second advantageous feature is the student assignment mechanism: a serial dictatorship. Each student receives a score before applying to high school. An algorithm 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; thus, its dominant strategy is to rank truthfully according to its preferences (Chade and Smith 2006). The serial dictatorship lets us: (i) observe the high schools that a student could attend and (ii) be confident that the one the student enrolls in is her most preferred. Further, 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, administrative data allow us to calculate value added and to see the *outcome* of household decision-making.

To probe the *mechanics* of this decision-making, we visited middle schools and collected a baseline survey. This survey occurred at school-sponsored information sessions held to help households apply to high school. In the survey, we interviewed parents to obtain the school preference rankings that they intended to submit. We also asked them to evaluate the high schools in their town along dimensions including location, peer quality, curriculum, and different types of value added.

We ran our experiment at the end of these sessions. At randomly selected treatment schools, we distributed a ranking of the town's high schools based on academic value added. After the assignment process was complete, we obtained students' official school assignments. We also conducted an endline survey, interviewing parents by phone to gather their submitted school preference rankings and to again elicit their beliefs about schools' value added.

This setup yields four findings, which we use to organize the exposition and frame the paper's contribution. They are as follows.

1. Households leave value added on the table

Schools with higher academic value added face higher demand. The correlation between a school's value added and the selectivity of its admissions cutoff is 0.56.[1] In addition, households choose options that are above average by value added in their feasible choice sets. Nonetheless, they leave considerable value added unexploited. Both low- and high-achieving students could gain, on average, about 1 s.d. worth of additional value added—or a 12 percentage point increase in the probability of passing the baccalaureate exam. By contrast, households come closer to maximizing selectivity. For this school trait, they leave only about 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 in which researchers can: (i) measure all schools' value added, and (ii) precisely observe the set of schools among which each household chooses, as well as the one it most prefers.

2. Households have limited knowledge of academic value added

When asked to score schools on academic value added, households' scores are off by an average of 1.1 within-town quintiles and explain only 17% of the variation. In contrast, households have more awareness of selectivity. Their scores for this school trait have a mean absolute error of 0.9 within-town quintiles and explain 33% of the variation. Finally, households with high-achieving

1. 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 there are positive peer effects.

children have more accurate beliefs than those with low-achieving children.

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

3. Households respond to information on value added (with heterogeneity)

Our treatment improved the accuracy of households' beliefs and caused them to assign higher preference ranks to high-value added schools. Thus, on average, it induced students to attend schools with 0.05 s.d. worth of additional value added. That said, the treatment had larger effects on beliefs and preference ranks for households with low-achieving students. In addition, it did not alter beliefs or ranks for the two options that a household ranked the highest in the baseline. As a result, its effects on students' school assignments were heterogeneous. Notably, for low-achieving students who were rejected by their two top choices, the treatment resulted in enrollment at schools with 0.2 s.d. worth of additional value added. This implies a 2.5 percentage point (9.8%) increase in their probability of passing the baccalaureate exam. By contrast, for all other students, the treatment had no impact on school assignments.

These results address 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 our knowledge, the first experimental distribution of information on school value added.

4. Households' preferences for other school traits limit their demand for value added

We use households' school preference rankings and their elicited beliefs about school traits to study their preferences for these traits. We first estimate preferences using a discrete choice model. We then disentangle the roles preferences and information play in causing households to leave value added on the table. Specifically, we compare predicted school assignments under accurate beliefs about academic value added with those under baseline beliefs. In different specifications, we predict that correcting households' beliefs would spur low- (high-) achieving students to attend schools with 0.13-0.18 (0.11-0.20) s.d. worth of additional value added. This is 17-25% (11-23%) of the value added that the households would leave unexploited under baseline beliefs. Households would not "max out" on value added under accurate beliefs due mostly to their preferences for curriculum and peer quality.

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 work assessing the roles of preferences and frictions in driving choices (Bergman et al. 2019; Hastings, Neilson, and Zimmerman 2018; Bergman, Chan, and Kapor 2020).

In the rest of the paper, Section 0 provides some necessary preliminaries, and Sections 1-4 report on findings 1-4, respectively. Section 5 concludes.

0 Preliminaries

We begin by describing the setting, the administrative data, our value added measures, our surveys, and our experiment.

0.1 Institutional setting

A few features of the Romanian setting are especially relevant to our analysis. First, in Romania, high schools cover grades 9-12 and are divided into *tracks*. These are self-contained units within schools that differ in their curriculum. Tracks fall into three broad categories: a) humanities tracks, b) math or science tracks, and c) "technical" tracks with applied themes such as business or agriculture.

Second, students are assigned to tracks via a centralized process known as a serial dictatorship. This process weights students' track preferences according to their academic performance in middle school (grades 5-8). Specifically, in 8$^{\text{th}}$ grade, each student takes a national high school entrance test. The student's score on this exam is combined with her middle school GPA to generate an admissions score, called the transition score. After finding out its child's transition score, a household submits a ranked list—or *preference ranking*—of its preferred tracks. The government then examines the preference rankings in the order of students' transition scores. It first takes the student with the highest score and assigns her to her most-preferred track. It then proceeds down the score distribution, assigning each student to her most-preferred track that is not yet at capacity.

Third, the track assignment process is incentive-compatible. Households' preference rankings can be of virtually unlimited length (up to 287 choices). As a result, the optimal strategy is to submit a list that truthfully reveals one's preferences.[2]

Fourth, a household's choice set is best thought of as the tracks in its town. Technically, households may rank any track in the country. However, it is uncommon for households to move for educational purposes. In addition, Romanian towns tend to be geographically distinct; thus, few students commute from one town to another. Further, Romanian towns are compact, and high schools are usually located in the town-center; thus, within-town commutes are rarely difficult. As a result of these features, we assume that households consider all the tracks in their town and do not consider options in other towns.[3]

2. Recent work notes that households may reasonably choose not to rank a track if they are certain it is out of reach for their child (Fack, Grenet, and He 2019; Artemov, Che, and He 2020). Below we show that our findings are robust to using empirical strategies that account for such "skipping".

3. In our baseline survey, over 93% of households said that they intended to apply only to tracks within their town. For these households, we find that within-town distances hardly affect track choice (Section 4). The one setting where distance may matter is Bucharest, which is by far the largest city. Following the Ministry of

Fifth, at the end of high school, students may elect to take a national standardized test known as the baccalaureate exam. The baccalaureate exam has high stakes: there are benefits both to passing it and to achieving a high score. Students who pass receive a baccalaureate diploma, which is necessary for admission to university—at less selective schools, it is the only requirement. A high score helps students access scholarships and prestigious universities (Borcan, Lindahl, and Mitrut 2017). Even for students who do not pursue higher education, performing well on the baccalaureate exam can be a strong labor market signal.

0.2 Administrative data

We have administrative data on the universe of students admitted to Romanian high schools. We use the data to calculate academic value added for each high school track and to examine whether households choose tracks with high value added. The data cover the 2004-2017 and 2019 cohorts. For all cohorts, they provide information on students' demographics, middle school, middle school GPA, scores on the transition exam, and assigned high school track. For 2004-2014, they also include performance on the baccalaureate exam. On average, a cohort includes about 144,000 students who live in about 400 towns and choose among about 3,800 tracks.[4]

Table 1: Summary statistics for the administrative data

	Mean	**Std. dev.**	**Students**
High school track:			
Number of students	61.8	47.0	2,162,736
Minimum transition score (MTS)	6.94	1.59	2,162,736
Student characteristics:			
Female	0.527	0.499	2,162,736
Transition score	7.70	1.35	2,162,736
Middle school GPA	8.65	0.97	2,162,736
Transition exam score	7.05	1.69	2,162,736
Baccalaureate performance:			
Took the exam	0.686	0.464	1,710,030
Passed the exam	0.533	0.499	1,710,030
Perfect score	0.001	0.025	1,710,030

The table provides summary statistics for the administrative data. Variables under "High school track" are characteristics of a student's track. Variables under "Baccalaureate performance" are available only for the 2004-2014 cohorts.

Table 1 (page 6) summarizes these data. One covariate in the table merits a special comment. A track's *minimum transition score* (MTS) is the score of the last student admitted. It is the track's admissions cutoff: students with higher scores are eligible to attend the track, while those with lower scores are not. The MTS is a direct measure of a track's selectivity; in addition, it is

Education, we divide Bucharest into six sub-town units. Our results are robust to excluding these units.

4. Appendix Table A1 presents the sample size by year. We impose three restrictions on the sample. First, we exclude 2018 due to a reporting issue. Second, we drop a small number of students who participate in vocational programs that do not offer a path to a baccalaureate diploma. Third, given that we are interested in track choice, we drop students who live in very small towns that offer only a single track.

a proxy for the demand that a track faces—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 accept students, it provides the tracks' MTS from the previous admissions round. Anecdotal evidence suggests that households pay attention to this information when determining their track preference rankings.

0.3 Value added

We calculate multiple measures of track value added; all relate to a track's effect on a student's performance on the baccalaureate exam. As mentioned, students choose whether to take this exam.[5] This means that value added calculated on students' scores could be biased by sample selection. Our main outcome, therefore, is an indicator for whether a student *passes* the exam; this variable is set to zero if the student fails the exam or does not take it. Two other outcomes are related to a student's score. First, the *percentile rank* of a student's baccalaureate performance is the percent of students in the admissions cohort who perform worse than the student, with all students who do not pass being assigned a value of 0.[6] Second, the *imputed exam score* is a version of the score that deals with missing scores using imputations: it fills in a value equal to the 33^{rd} percentile (in a cohort) among students who take but fail the exam.

We consider these three measures to be complementary. Value added on passing the exam is free of sample selection. Value added on the other outcomes allows more precise results for selective tracks in which large shares of students pass.

Our main value added measures vary by track and year; for robustness, we also calculate measures that vary by student characteristics. These measures allow a track-year to have different effects depending on whether a student is male or female or whether the student is better at math or language. It turns out that all our measures are highly correlated. For example, our main measure (a track-year effect on passing) has a correlation of over 0.9 with all the alternative measures (Appendix Table A2).

We detail our methodology for calculating value added in Appendices A1-A3. Three facts about it are worth mentioning. First, we rely on a traditional selection-on-observables model (Rothstein 2010; Angrist et al. 2017). Second, we validate our measures by comparing them with RD causal effects generated by track admissions cutoffs—this involves adapting Angrist et al. (2017) to an RD setting. We show that all of our measures closely match the causal effects; in addition, the measures that do not allow for heterogeneity by student characteristics perform

5. In our cohorts, 69% of high school students attempted the exam, 53% passed it, and 0.1% achieved a perfect score (Table 1). Appendix Figure A1 shows how these values vary with a student's transition score.

6. This outcome is motivated by the benefit structure of the baccalaureate exam. First, the benefits from a given score depend to a large extent on the fraction of students who perform worse. As such, households may be interested in the percentile rank of performance associated with the score. Next, there are no benefits to taking the exam but failing it. That is, students who fail the exam gain the same result as those who do not attempt it. Both groups do not perform better than any other students, and thus we assign them a percentile rank of 0.

just as well as those that do. Third, we deal with measurement error by calculating Empirical Bayes (EB) posterior means.

One complication in our setting is that we cannot observe baccalaureate outcomes for post-2014 admissions cohorts; as a result, we cannot directly calculate value added for these cohorts. We handle this by forecasting the missing years using machine learning.[7]

Our analysis uses value added variables which we label V_{jt} (for track-year effects) or V_{jgt} (for effects that vary by student). In this notation, j indexes tracks, t indexes years, and g indexes student types. For the years in which we can estimate value added (2004-2014), these variables equal the Empirical Bayes posteriors; for the years in which we cannot (2015-2017, 2019), they equal the machine learning forecasts.

In terms of magnitudes, we find that Romanian tracks differ significantly in value added. For instance, in 2019, a one standard deviation increase in true value added was equivalent to a 12 percentage point increase in the probability of passing the baccalaureate exam.

0.4 Baseline survey

To gain insight into households' beliefs and preferences, we conducted a baseline survey. We interviewed parents of 8$^{\text{th}}$ graders to collect: (i) their beliefs about the attributes of tracks in their towns, and (ii) their intended track preference rankings.[8] We did this at information sessions held by middle schools to inform parents about the high school application process. These occur about a month before households submit their track preference rankings.

To select our sample, we had to choose towns and 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. Second, among these towns, we chose those in which value added was easiest to forecast (Appendix Table A3 provides summary statistics for these towns). To choose middle schools, we randomly selected among those that had at least 15 students and in which it was logistically feasible for our surveyors to visit the information sessions. We wanted to minimize spillovers and general equilibrium effects from our experiment. As a result, we visited only a fraction of middle schools in each town—an average of 11% and never more than a third. Our sample covered 194 middle schools in 48 towns. In 2019—the year in which we conducted the survey—the towns 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.

We asked parents to score the tracks in their town—on a scale of 1 to 5—on a variety of dimensions.[9] Table 2 (page 9) lists those we covered. We first asked about two school attributes

7. We obtain the forecasts using a local linear forest (Athey et al. 2019). Our model uses a track's past value added and multiple track and student traits. We assess the model by making out-of-sample predictions in years in which we observe value added. The model predicts almost 80% of the variation in tracks' true value added.

8. Appendix A4 analyzes these rankings. It shows that households do not omit "out of reach" tracks. It also shows that households rank tracks from multiple curricular categories.

9. In Romania, scales of 1-5 are often interpreted in terms of quintiles. However, we were careful not to ask

that many studies find households value: 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 types of value added related to college and labor market success. Finally, we asked parents to score tracks on teacher quality, on whether the 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 2: Track characteristics covered in the baseline 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.

A number of checks suggests these scores are credible. First, the means and standard deviations are similar for the various quality dimensions (Table A5). Second, the scores have a reasonable across-dimension correlation matrix (Table A6). For instance, the largest correlations are 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 the track. Third, the value added scores have an intuitive relationship with the other scores: in a multivariate regression, they are explained by scores for teacher quality, curriculum, and peer quality, but not by scores for location or siblings and friends (Appendix A5).

Table 3 (page 10) describes other variables in the survey, revealing a few notable facts. First, households did not rank or score all the tracks in their towns. On average, they assigned ranks to 42% and quality scores to 35%.[10] Second, at the time of the baseline survey, households differed in the degree to which they had settled on their track choices: 39% were "very certain" of their preference rankings, while 46% were "somewhat certain" and 15% were "uncertain". Third, households with low-achieving children tended to be less certain than those with high-achieving children (Appendix A6). Fourth, students in the baseline survey had similar characteristics as those in the administrative data (Table 1).

households to group tracks into equal-sized bins. Instead, we requested that they assign each track whatever score they thought was appropriate. Thus, the scores roughly correspond to a household's expectation about a track's quintile, rounded to the nearest integer. This way of scoring tracks can incorporate information about a household's uncertainty. For instance, if a household has imprecise beliefs and is unable to differentiate among tracks, it can assign each a score of 3. By contrast, a confident household can assign a distribution of scores that approximates quintiles. Appendix Table A4 summarizes the frequency with which households assign scores of each value. The frequencies are all close to 0.2—although households tend to assign more high than low scores.

10. Despite this, we find that households considered tracks from across the selectivity distribution (Appendix A4). That is, it is not the case that households with low-achieving children omitted all selective tracks; nor is it the case that households with high-achieving children omitted all non-selective ones.

Table 3: Summary statistics for the baseline survey

	Mean	**Std. dev.**	**Students**
High school application process:			
Num. of tracks in the town	13.1	4.7	3,898
Share of tracks ranked	0.424	0.324	3,898
Share of tracks scored on passing the bacc.	0.353	0.413	3,898
Share of tracks scored on peer quality	0.363	0.417	3,898
Very certain of preference ranking	0.389	0.488	3,898
Somewhat certain of preference ranking	0.459	0.498	3,898
Student characteristics:			
Female	0.519	0.500	3,898
Mother's years of schooling	12.0	2.2	3,759
Transition score	7.72	1.41	3,746
Middle school GPA	9.05	0.85	3,769
Transition exam score	7.35	1.61	3,830

The table describes the baseline survey. The sample consists of 3,898 students in 194 middle schools in 48 towns.

0.5 Experiment and endline survey

We ran an experiment and an endline survey to explore the impact of providing households with information on value added.

The experiment took place during the middle school information sessions where we conducted the baseline survey. In advance of the sessions, we split the middle schools into treatment and control groups using a clustered randomization process.[11] At the end of the baseline survey, we distributed an informational flyer. 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 and included a ranking of the tracks in the town by our value added forecasts. Respondents were allowed to keep the flyers.[12]

After the high school allocation, we obtained students' track assignments from the Ministry of Education. In addition, we phoned households to conduct the endline (or "follow-up") survey. In this survey, we collected the final track preference rankings that households submitted and asked them to again score tracks on a scale of 1 to 5 in terms of value added. The data on track assignments lets us see how the information affected the tracks that students attend. The

11. We matched pairs of middle schools within towns based on school characteristics. We then randomized within these matched pairs. Appendix A7 provides details.

12. Example flyers are in Appendix Figures A2-A3. For all households, the flyer included Figure A2; for treated households, it also included Figure A3. Our intervention focused only on value added with respect to passing the baccalaureate exam. We stated that our rankings reveal "which tracks most effectively improve students' chances of passing the baccalaureate exam relative to their 9$^{\text{th}}$ grade starting points". It is possible that this type of value added is not of interest to students with very high or very low chances of passing. However, we believe this is unlikely. Appendix A5 shows that households' beliefs about a track's value added on passing the exam are highly correlated with their beliefs about the track's value added on college quality or on wages. Thus, households may have interpreted our information as a clear signal of value added on these other outcomes.

follow-up survey lets us probe the mechanics by which the information influenced choices.[13]

Table 4: Summary statistics and balance tests for the experiment

Covariate	Summary statistics		Balance tests			
	Mean	Std. dev.	Coef.	Std. error	Clusters	Students
Assigned to a high school track	0.846	0.361	0.023	0.020	78	3,186
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 passing the bacc.	0.411	0.421	-0.014	0.032	78	2,692
Share of tracks scored on peer quality	0.422	0.424	-0.005	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
Student characteristics:						
Female	0.530	0.499	0.015	0.022	78	2,692
Mother's years of schooling	12.3	2.0	0.111	0.102	78	2,625
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	7.54	1.50	0.148	0.106	78	2,692
In the follow-up survey	0.569	0.495	-0.014	0.026	78	2,692

The table presents summary statistics and balance tests for the experiment. The sample includes 3,186 students in 170 middle schools in 45 towns. "Assigned to a high school track" indicates whether a student received a track assignment in the main allocation. The sample for the other rows is limited to students for whom this variable equals 1. "Coef." is the coefficient in a regression of the listed variable on the treatment indicator. It measures the difference between the means for the treatment and control groups. Standard errors are clustered by middle school treatment-control pairs. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4 (page 11) presents summary statistics and balance tests for the experiment. The first row displays the share of students in the experimental sample who were assigned to a high school track. It shows that 85% of students were assigned, with an insignificant difference of 2.3 percentage points between treatment and control groups.[14] The remaining rows exclude students who were not assigned. Balance tests for this sample suggest that the randomization succeeded. The differences between treatment and control groups are small relative to the variables' standard deviations, and none are statistically significant. Moreover, a test of joint statistical significance returns a p-value of 0.722.

Comparing Tables 3 and 4 shows that students in the experiment are mostly representative of those in the baseline survey. However, they ranked and scored a larger share of tracks. Also, they were slightly more likely to be certain of their preference rankings, and they had slightly higher transition scores.[15]

13. In the time between the creation of the matched pairs and the baseline survey, some middle schools withdrew their permission for our study. For every school where this occurred, we still conducted the baseline survey in the other school in the matched pair. However, we removed the pair from the experimental sample. Thus, while the baseline 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.

14. We matched students with data on track assignments by name and middle school. Students do not appear in these data if they do not submit a track preference ranking. A small number of 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 main allocation. The remaining students either drop out or attend vocational schools.

15. Appendix Table A10 compares all the samples used in the paper.

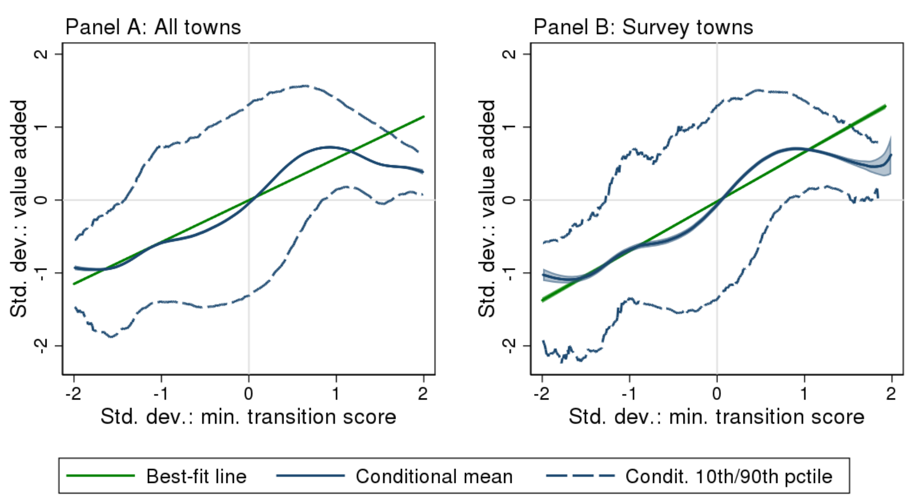
1 Households leave value added on the table

The first question we study is whether households choose tracks with high academic value added. Previous work considers this question in a variety of settings. We consider it further for four reasons. First, it reveals whether Romanian households gain academic benefits from their choices. Second, it shows if there is scope to increase their benefits by providing information. Third, it clarifies the representativeness of our setting: comparing our results with the literature provides a sense of whether our findings in later sections are likely to be externally valid. Fourth, in contrast to other studies, we can estimate the value added for all schools and observe each student's feasible choice set from among these—thus we can quantify the value added households leave unexploited.

We address the question 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. This approach is similar to the prior literature and lets us benchmark our results. We next calculate the precise amount of value added households leave unexploited. This analysis exploits our knowledge of households' choice sets and has not been done before. We note that in both cases the analysis is descriptive. It illuminates whether the tracks that households choose happen to have high value added—not whether households make choices based on value added.

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

Figure 1: The relationship between value added, 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, 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" are 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.

Figure 1 (page 12) shows that the relationship is strongly positive for less- and moderately selective tracks, but it is actually negative for highly selective tracks. The figure plots the conditional mean and the conditional 10^{th} and 90^{th} percentiles of standardized value added, V_{jt}, for given standardized values of minimum transition scores, MTS_{jt}.[16] It also includes a best-fit line from a linear regression. The resulting non-linearity holds in the survey towns (Panel B) and within curricular categories (Figure A4).[17]

Table 5 (page 13) quantifies the results in Figure 1: it presents coefficients from regressions of standardized V_{jt} on standardized MTS_{jt}. The values in the rows labeled "All tracks" match the slopes of the best-fit lines in Figure 1. The remaining rows capture the non-linearity in the figure—they are coefficients from regressions that split the sample by tercile of selectivity. The overall correlation between value added and selectivity is 0.56. But for the most-selective third of tracks, a one standard deviation increase in selectivity is associated with a 0.24 standard deviation decrease in value added. Panel B replicates the pattern for towns in the baseline survey.

Table 5: Regressions of standardized value added estimates on standardized selectivity

Sample	**Coefficient**	**Std. error**	**Town-years**	**Track-years**	**Students**
Panel A: All towns					
All tracks	0.560	0.005	5,969	57,521	2,162,736
By tercile of selectivity:					
Least selective	0.396	0.017	5,710	24,934	723,446
Moderately selective	1.07	0.028	4,325	17,207	723,023
Most selective	-0.239	0.024	2,420	15,380	716,267
Panel B: Survey towns					
All tracks	0.663	0.010	720	11,253	424,508
By tercile of selectivity:					
Least selective	0.406	0.042	717	4,319	135,007
Moderately selective	1.21	0.054	718	3,898	162,887
Most selective	-0.161	0.041	676	3,036	126,614

The table quantifies the results in Figure 1. It presents coefficients from regressions of standardized value added, V_{jt}, on standardized minimum transition score, MTS_{jt}. The coefficients in the rows labeled "All tracks" match the slopes of the best-fit lines in Figure 1, and can be interpreted as correlation coefficients. "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; standard errors are clustered by town-year.

These results are similar to prior work. For instance, in New York City high schools, the overall correlation between value added and peer quality is 0.59 (Abdulkadiroglu et al. 2020). In American higher education, the correlation between a college's selectivity and its earnings value added is 0.63 (Chetty et al. 2020). Also, in various locations, the most selective high schools do

16. V_{jt} here is our main value added measure: a track-year effect on the probability of passing the baccalaureate.

17. The pattern also holds regardless of the value added measure used (Figure A5). This mitigates a potential concern regarding Figure 1. In particular, the negative relationship between value added and selectivity for highly selective tracks could be due to 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 these tracks' measured value added. This cap is less likely to be binding for value added on the percentile rank of a student's exam performance or on a student's exam score—for instance, only 0.1% of students achieve a perfect score (Table 1). The non-linearity persists using these alternative measures.

not seem to boost achievement relative to students' fallback options (Abdulkadiroglu, Angrist, and Pathak 2014; Dobbie and Fryer 2014; Abdulkadroglu et al. 2017). Thus, Romania appears to be representative of other settings.[18]

We next characterize households' choices in relation to their available options. We exploit our knowledge of a household's feasible choice set—the tracks in the town that the student is eligible to attend, given her transition score and the admissions cutoffs. We conduct the analysis in two ways. First, we compare the value added of the track each household chooses with the value added of its other options. Second, we compute the amount by which each household could increase the value added it receives by switching to its highest-value-added option. We present a parallel analysis for selectivity, asking whether households favor tracks with high-achieving peers.

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, V_{jt}, or selectivity, MTS_{jt}) divided by the number of tracks that are available to the student.[19] 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 switching.

Table 6: 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, V_{jt}	67.1	61.1	72.9	67.2	59.9	74.2
Selectivity, MTS_{jt}	81.0	74.9	86.9	79.7	74.6	84.8
Panel C: Mean potential increase among feasible tracks (std. dev.)						
Value added, V_{jt}	1.01	1.01	1.01	0.91	0.93	0.88
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 are calculated for students with multiple feasible options; they display means for the "percentile rank of the student's track among feasible tracks" and the "potential increase among feasible tracks." 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.

Table 6 (page 14) shows that households choose above-average tracks by value added, but leave substantial value added unexploited. Panel A lists the percent of students who have only one track in their feasible set and hence no choice (2% in the full sample). The remaining panels

18. Appendix A9 provides additional replication of existing work. It studies choice behavior using a discrete choice model similar to that in Abdulkadiroglu et al. (2020) and Beuermann et al. (2019). Broadly speaking, we replicate previous findings. As in Abdulkadiroglu et al. (2020), we find that over the full sample, value added does not explain households' utility for tracks after conditioning on selectivity. As in Beuermann et al. (2019), we find that it does—to an extent—for households with high-achieving children.

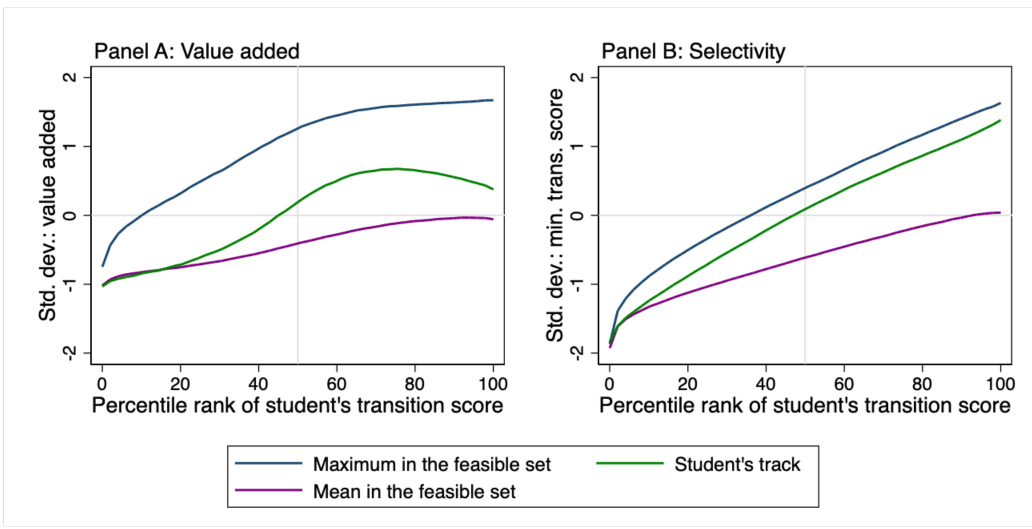
19. A value of 100 indicates that the household chooses the best option (by value added or selectivity).

concern the students with choice. Panel B reveals that on average these students attend tracks at the 67$^{\text{th}}$ percentile of value added among their feasible sets. Panel C shows that if they switched to their value added-maximizing options, they would gain, on average, 1 standard deviation of value added. In 2019, this was equal to a 12 percentage point increase in the probability of passing the baccalaureate exam.

Table 6 also shows that households come much closer to "maxing out" on selectivity. Over the full sample, students on average attend tracks with selectivity at the 81$^{\text{st}}$ percentile among their feasible tracks. The average potential increase in selectivity is only 0.32 standard deviations.

The remaining columns show that choice patterns do not vary greatly with students' academic achievement. In particular, the results are mostly similar for students with transition scores in the bottom half ("low-achieving") and top half ("high-achieving") of the within-year distribution.

Figure 2: Choice patterns by transition score



The figure shows how choice patterns vary with a student's transition score. It plots the relationship between the percentile rank of the student's transition score and three variables. The blue line is the maximum value of value added, V_{jt}, or selectivity, MTS_{jt} in the student's feasible set. The purple line is the mean value in the 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 6 for additional details.

This limited heterogeneity is illustrated in Figure 2 (page 15). It 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 the student's feasible set (in standard deviations of value added or selectivity), (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 for students with a given transition score. For both value added and selectivity, the potential increases are relatively constant across the transition score distribution. However, they are smaller for the lowest-achieving students, who have limited choice. In addition, for value added, the potential increase is larger for the highest-achieving students.[20]

Finally, we find that households leave significant value added unexploited even within curricular categories. In other words, it does not seem that households sacrifice value added because they willingly exchange it for this other track characteristic.[21]

2 Households have limited knowledge of value added

The previous section showed that households leave value added on the table. We now investigate whether this may be because households have inaccurate beliefs about value added. To do so, we make use of the baseline survey. We compare households' elicited beliefs from this survey with the values of track attributes that we observe as researchers (the "measured values"). To our knowledge, this is the first such comparison in the literature.

In the baseline survey, we elicited beliefs by asking households to score tracks on a variety of dimensions on a scale of 1 to 5 (Table 2). We compare these scores with our measured values along two dimensions. First, we compare households' scores for a track's value added on passing the baccalaureate exam with our forecast for this characteristic, V_{jt}. As a benchmark, we also compare households' scores for a track's peer quality with the track's prior-year selectivity, MTS_{jt-1}. The concept of selectivity is well-understood in Romania; in addition, households can view each track's prior-year selectivity on the official admissions website. Thus, this benchmark reflects scores under a scenario of easy access to information.[22]

We characterize households' scores in three ways. First, we quantify the *accuracy* of the scores: we calculate the mean absolute difference between a household's score and the within-town quintile of a track's measured value.[23] This quantity reveals the average amount by which households' scores are incorrect. Second, we quantify the *bias* of the scores: we calculate the mean difference between the scores and the tracks' quintiles. This value shows whether the scores tend to be too high or too low. Finally, we regress the quintiles on the scores. In the regression, the slope coefficient reveals how differences in scores map to differences in quintiles. The R-squared measures how much of the variation in quintiles can be explained by the scores.

20. Appendix Figure A6 replicates Figure 2 using alternative value added measures, with similar results.

21. Appendix Table A12 replicates Table 6 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 the average student attends a track with value added (selectivity) at the 66^{th} (80^{th}) percentile among this restricted choice set. On average, these students could gain increases in value added (selectivity) of 0.55 (0.26) standard deviations.

22. Arguably, households' peer quality scores should reflect a track's *current-year* rather than prior-year selectivity. A rational household might combine the data on prior-year selectivity with its knowledge of recent changes in tracks' traits. In our main analysis, we refrain from using current-year selectivity because this variable may be impacted by our experiment. Nonetheless, we find that results would be similar if we were to use it.

23. We compare scores with within-town quintiles because the scores are on a scale of 1 to 5. As discussed, a quality score can be roughly interpreted as a household's rounded expectation about a track's quintile. It is possible that households assigned scores based on the national—rather than within-town—distribution of tracks. However, we believe this is not the case. In results not shown, we replicated the analysis using national quintiles. We find that doing so causes the scores to have less predictive power.

If households' scores were fully accurate, the slope coefficient and R-squared would both be 1.[24]

Table 7: The accuracy and bias of households' quality scores

	All tracks			**Two most-preferred tracks**		
	All students	Low-achieving	High-achieving	All students	Low-achieving	High-achieving
Panel A: Accuracy (mean abs. dif.)						
Value added: s_{ij}^{V} v. quint(V_{jt})	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: Bias (mean dif.)						
Value added: s_{ij}^{V} v. quint(V_{jt})	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 summarizes the accuracy and bias of households' scores. Panel A describes accuracy: it displays the mean absolute difference between a household's score and the within-town quintile of the associated track characteristic. Panel B describes bias: it exhibits the mean difference between these quantities. "Two most-preferred tracks" are the two that the household ranked highest at baseline. The sample drops: (i) student-track observations where the respondent did not score the track on both value added and peer quality and (ii) 152 students with missing transition scores.

The results are in Tables 7 and 8 (pages 17 and 17). Table 7 summarizes accuracy and bias for two sets of tracks: (i) all those that a household scored and (ii) the two that the household reported as being most preferred. We include this latter set because households may know more about their favored options. Table 8 provides the regression results using all scored tracks.

Table 8: Regressing track attributes on households' quality scores

	All students		**Low-achieving**		**High-achieving**	
	quint(V_{jt})	quint(MTS_{jt-1})	quint(V_{jt})	quint(MTS_{jt-1})	quint(V_{jt})	quint(MTS_{jt-1})
Score: VA-pass, s_{ij}^{V}	0.416***		0.380***		0.435***	
	(0.019)		(0.032)		(0.018)	
Score: Peers, s_{ij}^{PQ}		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 regressions of measured values of track characteristics on households' scores. quint(·) is the within-town quintile of the given variable. The notes to Table 7 describe the sample. Standard errors are clustered by middle school. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The tables reveal a few notable points. First, households have relatively limited knowledge of tracks' value added on passing the baccalaureate exam. On average, households' scores for this characteristic are off by 1.1 quintiles (Table 7). A 1 point increase in a household's score is associated with only a 0.42 quintile increase in measured value added, and households' scores explain only 17% of the variation in quintiles (Table 8).

24. These quantities are not mechanically related. To see this, suppose a household assigns all tracks but one a score of 3. Suppose it then assigns the correct score to the remaining track. In this example, the slope coefficient would be 1 while the R-squared would be small.

Second, households are more—if still imperfectly—aware of track selectivity. On average, households' peer quality scores are off by 0.90 quintiles in predicting prior-year selectivity (Table 7). A 1 point increase in a peer quality score is associated with a 0.57 quintile increase in the true value, and households' scores explain 33% of the variation (Table 8).

Third, households with high-achieving children have more accurate scores than those with low-achieving children. For the former, value added (peer quality) scores are off by an average of 1.09 (0.84) quintiles and explain 20% (39%) of the variation. For the latter, values are 1.19 (1.01) quintiles and 12% (23%) of the variation.

Fourth, households' scores concerning their two most-preferred tracks are only slightly more accurate than their scores concerning all tracks (Table 7). Further, these scores are biased; households tend to think their favored tracks are better than they actually are. For instance, for value added, households with low-achieving (high-achieving) children on average over-estimate the quality of their preferred tracks by 0.64 (0.63) quintiles. For prior-year selectivity, the bias is an over-estimate of 0.65 (0.17) quintiles.[25]

3 Households respond to information on value added

The previous sections showed that households leave value added on the table and have only partially accurate beliefs regarding this attribute. We now test whether informing households about value added can influence their track choices. This could occur if information causes households to update their beliefs or if it alters their preferences over track characteristics (e.g., by making value added more salient).

3.1 Effects on students' assigned tracks

Our main outcome is the academic value added of the track that a student attends. In order to calculate the treatment effect on this outcome, we estimate:

$$\mathrm{sd}(\mathrm{V}_i) = \eta_0 + \eta_1 \cdot T_i + {\eta_X}' \cdot X_i + \eta_i. \tag{1}$$

Here, $\mathrm{sd}(\mathrm{V}_i)$ is the 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.[26] The coefficient of interest is η_1; it captures the average treatment effect of providing information.

Using this specification, Table 9 (page 19) shows that the intervention had a substantial effect, but only among households with low-achieving children. Over the full sample ("All students"), providing information caused students to attend tracks with value added that was higher by 0.05

25. Appendix A10 shows that these results are highly robust.

26. In our primary 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 the 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.

Table 9: 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	0.58	1.45	-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 (1). Low- (high-) achieving students are those with transition scores in the bottom (top) half of the national distribution. "Effect in percentage points" is the effect on the probability that a student passes the baccalaureate exam. We calculate this by multiplying the effect in standard deviation units by the 2019 standard deviation of true value added. "Predicted pass rate" is the share of students in the regression sample who are predicted to pass. We calculate this 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. Standard errors are clustered by the middle school treatment-control pairs within which we conducted the randomization. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

standard deviations (significant at a 10% level). This amounts to an increase in the probability of passing the baccalaureate exam of 0.58 percentage points, which is small relative to the 63% predicted pass rate. For low-achieving students, the treatment effect is 0.12 standard deviations (significant at a 5% level). This is a 1.45 percentage point increase in the probability of passing, as compared to a 29% predicted pass rate. For high-achieving students, the treatment effect is virtually zero and statistically insignificant. These results are robust to a variety of alternative specifications of regression (1).[27] They are also not confounded by informational spillovers.[28]

We next investigate whether treatment effects vary based on whether a student was eligible for the tracks he or she preferred in the baseline. Over 95% of households expected their child to be admitted to at least one of the two tracks that they ranked highest (Appendix A8). 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 children's assignments. In 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.

The results (Table 10, page 20) are consistent with this story: the treatment had little impact for students who were admitted to one of their top baseline choices, and a large impact for students who were not. In Table 10, the columns indicate which baseline choice a student was eligible for.[29] The effects are statistically insignificant and close to zero for students who

27. For example, controlling for different covariates and using a difference-in-difference design; see Table A13.

28. Namely, if treated households shared information with others in the town, including some in the control group, the effects would be biased toward zero. Appendix A11 tests for spillovers by examining whether effects are larger in towns in which we visited a smaller fraction of middle schools. We find no evidence that they are.

29. The first column is for students who scored above the cutoff for their most-preferred baseline choice. The second column is for students who scored above the cutoff for their second-most-preferred baseline choice, but not for their top choice. The remaining columns are for students who were eligible for only their third-most-preferred

Table 10: 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	0.23	-0.86	2.21	2.08	2.05	2.28
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 (1) for subsets of students by eligibility for the tracks 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.

were eligible for either their most- or second-most-preferred baseline choice. For the remaining students, the effects are always significant and range from 0.17 to 0.19 standard deviations. These translate into increases in the probability of passing the baccalaureate exam of over 2 percentage points (substantial relative to predicted pass rates of 28-33%).

Table 11: Effects on value added by additional student characteristics

	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	0.08	0.42	0.00	0.01	0.16	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	2.21	2.45	-0.28	2.32	2.16	1.85	2.65
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 (1) 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. See the notes to Table 9 for additional details on the regressions.

We next examine how the heterogeneity in Table 10 interacts with student characteristics. We estimate regression (1) for different types of students, always distinguishing between those who were eligible for their two top baseline choices and those who were not. Table 11 (page 20) contains the results. Panel A provides the effects for the students who were eligible. The impacts for these students remain small and statistically insignificant regardless of achievement,

track or worse, fourth-most-preferred or worse, etc.

gender, or mother's schooling. Panel B is for the ineligible students. The effects for these students vary little by gender or mother's schooling, but they do vary by achievement. For low-achieving students, the effect is 0.20 standard deviations (significant at a 1% level). This is a 2.45 percentage point increase in the probability of passing, over a 25% predicted pass rate. Meanwhile, for high-achieving students, the effect is statistically insignificant and close to 0.

Table 12: Effects on other characteristics of students' tracks

	Value added	**Selectivity**	**Peer SES**	**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	0.017	0.015	-0.013	0.012	0.001
	(0.024)	(0.017)	(0.013)	(0.020)	(0.017)	(0.016)	(0.010)
Clusters	78	78	78	78	78	78	78
Students	2,054	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	-0.009	0.073	0.039	0.030	-0.069**
	(0.065)	(0.050)	(0.052)	(0.086)	(0.030)	(0.027)	(0.033)
Clusters	76	76	76	76	76	76	76
Students	638	638	638	492	638	638	638

The table estimates regression (1) for different outcomes. Columns 1 and 2 refer to value added and the minimum transition score. Column 3 refers to the average transition score in the middle schools of a track's students. This is a measure of track peer SES because Romania has neighborhood-based middle schools. The outcomes in Columns 1-3 are all in standard deviation units. The outcome in Column 4 is a household's baseline score for a track's location quality. Columns 5-7 refer to a track's curricular focus. Regressions control for values of the outcome variable for the feasible track that the household's ranked highest at baseline. This is the track to which the household would have been assigned based on its baseline ranking. The regressions also include indicators for students who did not rank any feasible tracks. Standard errors are clustered by middle school treatment-control pairs.

We do not find evidence that households made large tradeoffs in order to attend higher-value added tracks. Table 12 (page 21) re-estimates (1) using additional track characteristics as outcomes. As before, we provide results separately for students who were and were not eligible for their most-preferred baseline choices. The results indicate that the treatment had little impact on track characteristics other than value added. This is the case even for students who were rejected by their top choices. Notably, these students attended tracks with 0.18 standard deviations worth of additional value added; yet they did not experience any difference in selectivity, peer socio-economic status, or location quality. The only tradeoff that these students were induced to make relates to curricular focus: they were 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 academic value added than humanities or math and science tracks (Figure A4).[30]

30. The fact that the treatment induced students to switch out of technical tracks is a potential concern. The impacted students were largely low-achieving (Table 11), and it is possible that technical tracks—being more career-oriented—are a good option for students unlikely to succeed in college. In this story, these tracks' low test-related value added would mask higher wage value added. We are unable to measure wage value added. Nonetheless, we explore this story using our data on households' beliefs. For each value added dimension we asked about at baseline, we calculate the mean quality score (across households) for a given track. These variables exploit the "wisdom of the crowd" to measure value added. We use the variables as outcomes in regressions akin

In short, the treatment had heterogeneous effects on the value added of students' tracks and little effect on other track characteristics. For value added, the treatment had no impact for high-achieving students or for low-achieving students who were admitted to one of their preferred baseline choices. By contrast, it had large impacts for low-achieving students who were rejected by these choices. We next try to understand this heterogeneity by using the follow-up survey to investigate effects on beliefs and track preference rankings.

3.2 Effects on beliefs regarding value added

This subsection presents treatment effects on beliefs—it explores whether providing information increased the accuracy of households' scores for academic value added. Before turning to results, we note that there are two ways in which this analysis may understate effects on beliefs. First, it is possible for information to influence the precision of households' beliefs without changing their scores.[31] Second, the scores may contain measurement error: the follow-up survey took place a few weeks after households submitted their preference rankings; by this time, households may have forgotten some of what they knew when they were deciding their track preferences. Despite these caveats, our sense is that treatment effects on quality scores are a useful proxy—and possibly a lower bound—for those on beliefs.

To conduct the analysis, we estimate:

$$|\text{quint}(\text{V}_{jt}) - s^V_{ij,\text{fs}}| = \eta_0 + \eta_1 \cdot T_i + {\eta_X}' \cdot X_{ij} + \eta_{ij}. \tag{2}$$

Here, $|\text{quint}(\text{V}_{jt}) - s^V_{ij,\text{fs}}|$ is the absolute difference between: (i) the value added of track j in units of within-town quintiles, $\text{quint}(\text{V}_{jt})$, 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 (1), 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.

The results indicate that the treatment increased the accuracy of households' value added scores, but only for their less-preferred tracks. In Table 13 (page 23), the first column is for all tracks, while the others distinguish by a track's position in a household's baseline preference ranking. For the full set of tracks, the treatment led to a statistically insignificant improvement in accuracy of 0.06 quintiles—small relative to the mean inaccuracy of about one quintile. For the tracks that households' ranked highest and second-highest in the baseline survey, effects are

to those in Table 12. If the tracks that students switched out of were believed to have high wage value added, then treatment effects on mean quality scores for this outcome would be smaller than on scores for academic value added. The results do not support this (Table A14). They show positive and similarly sized treatment effects on scores for all value added dimensions. That is, the treatment induced students to switch to tracks believed to be better for both academics and wages.

31. Suppose that a track's true score is 4, and that during the baseline a household believed the track had an equal chance of being a 3, 4, or 5. In this case, the household would assign a score of 4 but would be uncertain about its decision. Our treatment would remove the uncertainty, although the score would stay the same.

close to zero. For the remaining tracks, improvements are sizable and are all significant at either the 1% or 5% confidence level. These results also reveal that the changes in accuracy grow larger for tracks that are farther down a household's baseline preference ranking. For instance, among tracks other than a household's two top baseline choices, the improvement is 0.10 quintiles; among tracks other than a household's five top baseline choices, it rises to 0.18 quintiles.

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	$\geq$ 3rd-most-preferred	$\geq$ 4th-most-preferred	$\geq$ 5th-most-preferred	$\geq$ 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 (2). The values in the row labeled "Treated" are the estimates for η_1. 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 ranked second highest. "$\geq$ 3rd-most-preferred" are all tracks other than the two most preferred. 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 value added, $\text{quint}(\text{V}_{jt})$, and (ii) the household's baseline score for the track on this dimension, s^V_{ij}. "Mean abs. difference: baseline" is the mean absolute difference between $\text{quint}(\text{V}_{jt})$ and s^V_{ij} for the sample. Similarly, "Mean abs. difference: follow-up" is the mean absolute difference between $\text{quint}(\text{V}_{jt})$ and $s^V_{ij,\text{fs}}$. Standard errors are clustered by the middle school treatment-control pairs within which we conducted the randomization.

This pattern holds regardless of whether households have low- or high-achieving children. In each case, providing information increased the accuracy of scores, but only for tracks that households did not initially prefer (Tables A15 and A16). That said, magnitudes are twice as large for households with low-achieving children as for those with high-achieving ones.

3.3 Effects on track preference rankings

We next analyze whether information caused households to assign higher preference ranks to tracks with higher value added. To do this, we calculate treatment effects 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:

$$\text{ppr}_{ij,\text{fs}} = (\delta_1 + \delta_2 \cdot T_i) \cdot \text{pr}(\text{V}_{jt}) + (\delta_{X,1} + \delta_{X,2} \cdot T_i)' \cdot X_{ij} + \delta_{ij}. \quad (3)$$

Here, $\text{ppr}_{ij,\text{fs}}$ is household i's percentile preference rank for track j, as reported in the follow-up survey.[32] It is calculated by dividing a track's rank in the household's preference ranking

32. 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 ranking. During the follow-up, we asked households to find their copy and read off the ranking. We therefore believe that the rankings reported in the follow-up closely approximate those submitted. For instance, 99% of respondents report that their child attends the track that we observe them attending in the administrative data.

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.[33] Next, $\text{pr}(\text{V}_{jt})$ 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.[34]

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 (3). The values in the row labeled "Value added: treated" are the estimates for δ_2. The columns provide results for different sets of tracks. "Two most-preferred" refers to the two tracks that households ranked highest at baseline. "$\geq$ 3rd-most-preferred" are all tracks other than the two 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 ranking and whether the household is in the treatment group. "Association: baseline" is the slope coefficient from a regression of the percentile preference rank from the baseline survey, ppr_{ij}, on $\text{pr}(\text{V}_{jt})$. "Association: follow-up" is the slope coefficient from a regression of $\text{ppr}_{ij,\text{fs}}$ on $\text{pr}(\text{V}_{jt})$. Standard errors are clustered by the middle school treatment-control pairs within which we conducted the randomization.

The results exhibit a pattern similar to that for effects on the accuracy of scores (Section 3.2): providing information increased the association between preference ranks and value added, but only among tracks that households did not initially prefer. Table 14 (page 24) shows that among all tracks, the treatment caused the association to be higher by 0.05 percentiles (significant at the 10% level). For the two tracks that a household ranked highest in the baseline survey, the effect is insignificant and of the wrong sign. After excluding these tracks, the effect rises to 0.06 percentiles and is significant at the 1% confidence level. Moreover, the effect continues to grow for tracks that are farther back in the baseline preference ranking.

This pattern persists for households with low- and high-achieving children. For both, effects exist only among tracks other than the two top baseline choices. That said, for households with high-achieving children, the effects are small and mostly insignificant; for those with low-achieving children, they are sizable and significant at a 1% level (Tables A17 and A18).

33. We set $\text{ppr}_{ij,\text{fs}}$ equal to 0 for tracks that households do not rank, since students cannot be assigned to these. 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.

34. To see this, note that δ_1 is the average slope of conditional-on-X_{ij} best-fit lines between $\text{ppr}_{ij,\text{fs}}$ and V_{jt} for households in the control group; $\delta_1 + \delta_2$ is the average slope of these lines for treated households.

3.4 Discussion

The results for beliefs and preference rankings help to explain the heterogeneity in impacts on the value added of students' tracks. For high-achieving students, providing information had modest effects on beliefs and little effect on 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, impacts on track assignments differ depending on whether a student was eligible for her top baseline choices. The treatment had no influence on assignments for students who were eligible for these choices, but it had significant impacts for students who were not.

As noted, 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.[35] 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. This behavior is consistent with evidence that searching for information on school quality is costly (Arteaga et al. 2021).

A separate question is why responses were larger for households with low-achieving children than for those with high-achieving children. One explanation is that households with high-achieving children may have been more certain of their beliefs and rankings at the time of the baseline. If so, our intervention may have come too late in their decision-making process. We assess this by comparing the two groups' self-reported certainty about their baseline preference rankings. Households with high-achieving children were indeed more certain; 45% reported being very certain, 43% were somewhat certain, and 12% were uncertain. For households with low-achieving children, the corresponding percentages are 33%, 50%, and 17% (Appendix A6).

A second explanation is that households with low-achieving children may be more trusting of information provided by outside authority figures. If so, treatment effects on beliefs and preference rankings would be larger for low-achieving students even after conditioning on certainty. Two results emerge in this regard (Table A19). First, for both low- and high-achieving students, effects are larger for households who reported being uncertain or somewhat certain than for those who were very certain. Second, within the two certainty groups, effects are larger for low-achieving students.

In short, 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.

35. Recall that more than 95% of households expected their child to be admitted to at least one of their two top baseline choices (Appendix A8).

4 Preferences for other traits limit demand for value added

We next explore how demand for academic value added is constrained by households' preferences for other track characteristics. We first estimate preferences—in a discrete choice model, we explain households' track preference rankings using their quality scores. We then predict how track choices would change if households' scores for value added were made to be fully accurate. This exercise allows us to decompose the value added that households leave unexploited into two components: one that is due to preferences and another that is due to inaccurate beliefs.

We caution that the results do not provide a firm upper bound on the impact of providing information. This is because households' choices may depend on the precision of their beliefs in ways not captured by the quality scores; in addition, providing information may influence preferences, such as by signaling the importance of value added. With these caveats in mind, we conclude by comparing the magnitude of our experimental treatment effects with those predicted under accurate quality scores.

4.1 Households' preferences for track characteristics

We estimate preferences for track characteristics by relating households' track preference rankings to their beliefs about the attributes of the tracks. For simplicity, we use baseline values of track preference rankings and beliefs.[36]

Specifically, we assume that households rank tracks according to expected utility.[37] We then write a household's baseline expected utility from a track as a linearly separable function of its baseline scores for the track on various quality dimensions. This is:

$$U_{ij} = \sum_q \beta_q \cdot s_{ij}^q + \epsilon_{ij}, \tag{4}$$

where U_{ij} is household i's baseline expected utility from track j, s_{ij}^q is the household's score for the track on dimension q, and ϵ_{ij} is an error term. The β_q coefficients reflect households' preferences for track attributes; they represent the change in expected utility associated with a one-unit increase in a given quality score. To estimate (4), we assume that the error term, ϵ_{ij}, is independent and follows a Type-1 Extreme Value distribution. We then fit the model to households' baseline preference rankings using a rank-ordered logit.[38]

36. Appendix A12, we instead use endline values, and we exploit experimental variation in beliefs. We show that all results are very similar.

37. This is weakly dominant since the track assignment mechanism is incentive compatible. The dominance is strict for tracks that a household believes its child has a chance of attending. In our main analysis, we assume that households consider all the tracks in their towns. The results are robust to excluding tracks that households may have considered "out of reach."

38. 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. We use just those for a household's top choices. In our main results, we consider a households' *two* top choices. That is, we maximize the probability that a household prefers its highest-ranked track, r_{i1}, to all other tracks in the town times the probability that the

Table 15: Households' preferences for track attributes

	(1)	(2)	(3)	(4)	(5)
Location	0.276***	0.292***	0.277***	0.281***	0.289***
	(0.0689)	(0.0688)	(0.0687)	(0.0655)	(0.0685)
Siblings and friends	0.336***	0.326***	0.319***	0.344***	0.311***
	(0.0480)	(0.0477)	(0.0482)	(0.0479)	(0.0479)
Peer quality	0.344***	0.317***	0.318***	0.380***	0.298***
	(0.0692)	(0.0659)	(0.0691)	(0.0665)	(0.0743)
Curriculum	0.931***	0.789***	0.877***	0.986***	0.763***
	(0.0708)	(0.0691)	(0.0681)	(0.0729)	(0.0678)
VA: pass the bacc.	0.337***				0.0130
	(0.0819)				(0.0829)
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)
R-sq.	0.33	0.33	0.33	0.32	0.34
Clusters	150	150	150	150	150
Students	1,170	1,157	1,151	1,168	1,137
Student-tracks	11,575	11,395	11,382	11,573	11,220

The table presents results from equation (4). The model is estimated by maximizing the log-likelihood corresponding to equation (5). The sample is limited to students in experimental middle schools. Standard errors are clustered by middle school.

Table 15 (page 27) presents the estimates for the β_q coefficients, which indicate that households care about a variety of track characteristics. Column 1 presents our benchmark model. It shows that households have similar preferences for a track's location (coef. estimate of 0.28), for whether a child's siblings and friends use the track (0.34), for the track's peer quality (0.34), and for the track's value added on passing the baccalaureate exam (0.34). By contrast, households have considerably stronger preferences for curriculum (0.93). All estimates are significant at a 1% confidence level.

household prefers its second-highest-ranked track, r_{i2}, to all other tracks except r_{i1}. This likelihood is:

$$\Pr[r_{i1}, r_{i2}|\mathcal{J}_i, \{s^q_{ij}\}_{q,j}] = \prod_{l=1}^{2} \frac{\exp[\sum_q \beta_q \cdot s^q_{ir_{il}}]}{\sum_{k\in\mathcal{J}_i\setminus\{r_{im}:m<l\}} \exp[\sum_q \beta_q \cdot s^q_{ik}]}. \tag{5}$$

We focus on the first two choices because most households appear to have settled on these by the time of the baseline survey. However, in Section 4.2, we also run specifications with different numbers of choices. In addition, we sometimes restrict attention to tracks that are plausibly feasible. When we do this, we re-define a household's top choices as the most preferred among this narrower set. There are three ways in which our approach may fail to recover preferences. First, we may be omitting a quality dimension that is correlated with both utility and the s^q_{ij} covariates. This concern is mitigated because we have scores on a large number of quality dimensions. Second, the quality scores may contain measurement error—we consider this issue in the main text. Third, it is possible that households are risk averse with respect to track characteristics. In this case, utility would not be linear in the characteristics—as in (4)—but rather strictly concave. Thus, expected utility would depend on the precision of households' beliefs in a manner not captured by quality scores. For instance, a household would gain more expected utility when it knows a track is a 4 than when it thinks the track has an even chance of being a 3, 4, or 5. We ignore this issue because accounting for it would require data on the full density of beliefs.

Columns 2-5 explore preferences for alternative dimensions of value added: value added on college quality, value added on wages, and a track's teacher quality. When included one at a time, all value added dimensions are statistically significant (Columns 2-4). However, the results suggest that households care most about value added on college and wages. For instance, in a horse race (Column 5), the β_q estimates are large and significant for these dimensions, but small and insignificant for passing the baccalaureate exam and for teacher quality. One interpretation is that households may see the two latter dimensions as inputs into the former two.

We find some heterogeneity in preferences between households with low- and high-achieving children (Table A20). First, the two groups differ in their preference for peer quality: coefficient estimates are large (0.43-0.56) and statistically significant for high-achievers, but small (0.06-0.13) and insignificant for low-achievers. Second, the groups' quality scores have differing degrees of explanatory power. Depending on the specification, these explain 40-41% of the variation in track choices for high-achievers and only 20-22% for low-achievers.

Our preference estimates are robust to a variety of potential issues. A first concern is that few households provide scores for all the tracks in their towns. Missing scores could introduce bias if a household's propensity to score a track depends on its preference for the track. We gauge the impact of missing scores using two approaches. First, we limit the sample to households without missing scores. Second, we impute the missing scores using a random forest.[39] In both cases, results are similar to those for our main specification (Table A21).

A second concern stems from the "skipping" issue highlighted by Fack, Grenet, and He (2019) and Artemov, Che, and He (2020). If households refrain from ranking tracks that they believe will not admit their child, then their rankings will not reflect their true preferences. To assess this issue, we run two specifications that exclude tracks that households may have considered "out of reach". Again, results are unchanged (Table A22). Importantly, the preference estimate for peer quality among low-achieving students remains small. Thus, this value appears to be a reflection of preferences rather than an artifact of skipping.

A final concern is that the quality scores may contain measurement error—that is, they may be a noisy proxy for the expectation of a household's beliefs. Measurement error would cause the preference estimates to be attenuated. To explore this concern, we use a horse race proposed by Kapor, Neilson, and Zimmerman (2020). The approach involves re-estimating the preference model while adding controls for measured values of track characteristics. Specifically, for the attributes of curriculum, peer quality, and academic value added, we have both quality scores and measured values. Thus, we can test the quality scores by including the measured values. If the quality scores are noisy, then the measured values may contain additional information

39. 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.

about households' beliefs, in which case they would provide additional explanatory power for expected utility.[40] The results suggest that measurement error is a modest issue (Table A23). The measured values are often statistically significant; nonetheless, they generate only small increases in R-squared. In addition, they exert limited impact on the coefficients for the quality scores.[41]

4.2 Track choices under accurate beliefs

Next, we simulate how track choices would change if households had accurate beliefs about academic value added.[42] Using the preference model, we predict choice outcomes under two sets of quality scores. The first set, *Inaccurate scores*, uses households' baseline scores for value added. The second set, *Accurate scores*, replaces these with within-town quintiles of measured value added.[43]

For each track in a household's feasible choice set, we predict the probability that the household would prefer the track, given the scores. We then use the probabilities to predict the value added of the track the student would attend. This latter prediction is a weighted average of the value added of each feasible track, with weights that are equal to the predicted preference probabilities. For *Inaccurate scores*, it is:

$$\mathrm{V}_{i,\mathrm{IS}} \equiv \sum_{j \in \mathcal{J}_i^e} \mathrm{sd}(\mathrm{V}_{jt}) \cdot \frac{\exp[\sum_q \hat{\beta}_q \cdot \tilde{s}_{ij}^q]}{\sum_{k \in \mathcal{J}_i^e} \exp[\sum_q \hat{\beta}_q \cdot \tilde{s}_{ik}^q]}.$$

Here $\hat{\beta}_q$ is a coefficient estimate from the preference model (4), $\mathcal{J}_i^e$ is household i's feasible choice set, and $\tilde{s}_{ij}^q$ is a score in *Inaccurate scores*. For the prediction for *Accurate scores*, $\mathrm{V}_{i,\mathrm{AS}}$, the formula is analogous but with correct scores for value added.

Under each set of scores, we produce four versions of our predictions. These reflect different assumptions about the preference model and about how households update their beliefs in response to information. Our first specification is titled "Just quality scores". It uses a preference model akin to that in Column 1 of Table 15, controlling for scores for location, siblings and

40. The measured values may be significant even absent measurement error. For example, they may be correlated with omitted quality dimensions. In this way, the test can provide evidence for measurement error but not proof.

41. For scores for location, siblings and friends, and curriculum, the coefficients are not changed at all. For scores for value added, coefficients fall by 25% (44%) for low- (high-) achieving students. For scores for peer quality, coefficients fall for high-achieving students and rise slightly for low-achieving ones.

42. The analysis in this section holds constant households' feasible choice sets. Thus, it lends insight into choice behavior, but it does not reveal the impacts of a large-scale policy of information provision. In particular, if all households were somehow made to have correct beliefs, then the choice setting would change due to dynamic effects on track selectivity, teacher sorting, value added, etc. We leave these effects for future work.

43. For the other quality dimensions, we always use households' baseline scores. The only exception is peer quality, where we substitute the within-town quintile of a track's selectivity. We make this substitution because households have access to information on selectivity and may absorb this information before making their final choices. We impute missing baseline scores using a random forest.

friends, peer quality, curriculum, and value added on passing the baccalaureate exam.[44] The second specification—"With measured attributes"—adds measured values of track characteristics. The third specification—"Update on all VA dimensions"—supposes that households may update their beliefs on all the value added dimensions that we asked about in the survey, not just on value added with respect to passing the baccalaureate exam. The preference model for this specification is similar to that in "With measured attributes"; however, it includes quality scores for each of the four value added dimensions. Further, when calculating $\mathrm{V}_{i,\mathrm{AS}}$ for this specification, we "correct" the scores for all of these dimensions.[45] Finally, the fourth specification—"Adjust for measurement error"—accounts for the fact that the preference estimate for value added may be attenuated by measurement error. It uses the same preference model as "With measured attributes" but inflates the value added coefficient by a factor of 1.5.

Before turning to results, we validate our approach by comparing our predictions with households' observed choices. In particular, *Inaccurate scores* is meant to represent households' beliefs in the absence of the experiment. As such, for households in the control group, $\mathrm{V}_{i,\mathrm{IS}}$ should match the value added of students' actual tracks.[46] We find that $\mathrm{V}_{i,\mathrm{IS}}$ has strong predictive power.[47]

Table 16 (page 31) presents the results of the simulation, indicating that inaccurate beliefs play a limited role in explaining why households leave value added unexploited. This finding holds for each of the four specifications that we use to make predictions. The column labeled "Potential increase in VA: $\mathrm{V}_{i,\mathrm{IS}}$" reveals how much value added households would leave unexploited, on average, under *Inaccurate scores*. Across specifications, values range from 0.73 to 0.84 s.d. for low-achieving students and 0.91 to 0.94 s.d. for high-achieving ones. "Potential increase in VA: $\mathrm{V}_{i,\mathrm{AS}}$" provides corresponding results for *Accurate scores*. These range from 0.55 to 0.67 s.d. for low-achieving students and 0.71 to 0.82 s.d. for high-achieving ones. "Change in VA" displays the mean difference between $\mathrm{V}_{i,\mathrm{AS}}$ and $\mathrm{V}_{i,\mathrm{IS}}$—or the predicted treatment effect of correcting households' scores. This column shows that under correct scores, low- (high-) achieving students would, on average, attend tracks with 0.13 to 0.18 (0.11 to 0.20) s.d. of additional value added. Finally, "Share of potential increase" divides the predicted treatment effects by the potential increase in value added under *Inaccurate scores*. It thus reveals the percent of the

44. In practice, the model differs in two ways from that in Table 15. First, we allow coefficients to vary based on whether a student is low- or high-achieving (as in Table A20). Second, we estimate the model after imputing missing quality scores with the predictions from the random forest (as in Table A21). Thus, the coefficients for this preference model are those in the third and sixth columns of Table A21. We present coefficients for all the preference models that we use in this section in Table A24.

45. We always replace them with the within-town quintile of value added on passing the baccalaureate exam.

46. The analogous exercise for households in the treatment group would compare $\mathrm{V}_{i,\mathrm{AS}}$ 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.

47. Figure A7 plots the value added of control students' tracks against $\mathrm{V}_{i,\mathrm{IS}}$. It also includes best fit lines from linear regressions. For the "Just quality scores" specification, the R-squared of the best-fit line is 0.61; for the other specifications, it is 0.67. The slope coefficients are all close to 1.

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

	Potential increase in VA		**Change**	**Share of**
	$V_{i,IS}$	$V_{i,AS}$	**in VA**	**pot. incr.**
Panel A: Just quality scores				
All students	0.882	0.691	0.191	0.216
Low-achieving	0.835	0.667	0.168	0.201
High-achieving	0.910	0.705	0.204	0.225
Panel B: With measured attributes				
All students	0.857	0.745	0.113	0.131
Low-achieving	0.736	0.611	0.125	0.170
High-achieving	0.929	0.824	0.105	0.113
Panel C: Update on all VA dimensions				
All students	0.862	0.702	0.160	0.185
Low-achieving	0.738	0.573	0.165	0.223
High-achieving	0.935	0.778	0.157	0.167
Panel D: Adjust for measurement error				
All students	0.853	0.691	0.162	0.190
Low-achieving	0.726	0.546	0.180	0.248
High-achieving	0.928	0.777	0.151	0.163

The table describes our predictions for how the value added of students' tracks would change under accurate beliefs. "Potential increase in VA" is the mean difference between (i) the maximum value added in students' feasible sets and (ii) the listed variables. "Change in VA" is the mean difference between $V_{i,AS}$ and $V_{i,IS}$. "Share of potential increase" is the ratio of "Change in VA" to the potential increase under inaccurate scores. See Section 4.2 for the definitions behind panels A-D. The sample includes 997 low-achieving and 1,680 high-achieving students. It is similar to that for the experimental treatment effects from Section 3. However, it excludes 15 students who did not score above the cutoff for any tracks that existed in both 2018 and 2019. (These students were assigned to tracks that were newly created in 2019 and for which we did not elicit beliefs.)

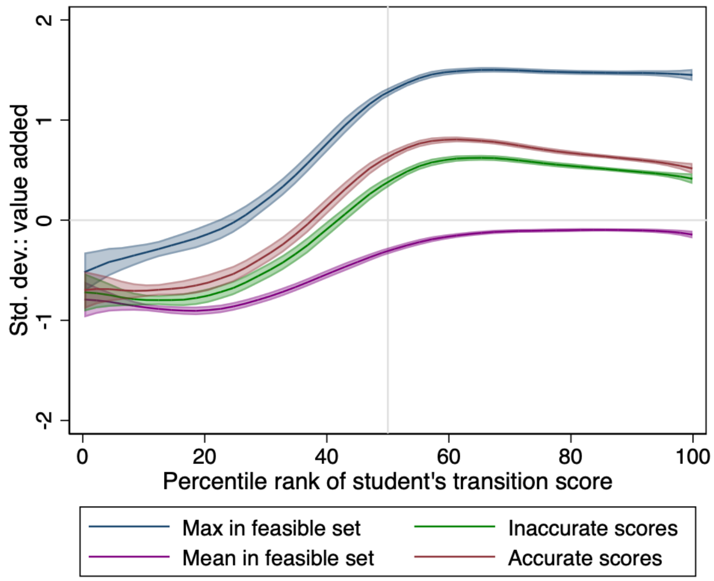
unexploited value added that is due to inaccurate beliefs. For low- (high-) achieving students, these percentages vary from 17% to 25% (11% to 23%).[48]

These results are in Figure 3 (page 32), which is similar to Figure 2 in Section 1. The figure shows how our predictions for the value added of students' tracks compare to students' feasible options; in addition, it reveals how these patterns vary based on a student's achievement. As in Table 16, we can see that correcting households' value added scores would cause students to attend tracks with higher value added; however, the effects would represent only a fraction of the potential increase. This holds throughout the achievement distribution.

The fact that households would continue leaving value added unexploited under accurate beliefs reflects that they are constrained by their preferences for other track characteristics—mostly by those for curriculum and peer quality. Table 17 shows the mean potential increase in value added (under accurate beliefs) if we were to set certain preference coefficients, $\hat{\beta}_q$, to 0 when calculating $V_{i,AS}$. If households cared only about value added and curriculum, they would leave an average of 0.46 s.d. of value added on the table (Column 1). If they cared only about value added, curriculum, and peer quality, they would leave 0.58 s.d. unexploited

48. Table 16 estimates the preference model using only the first two choices in a household's preference ranking. Further, it defines a choice set as the full set of tracks in the household's town. We explored alternative ways of estimating the model (Table A25). In some specifications we use different numbers of choices and in others we limit the choice set to tracks that households could have expected to be feasible. These changes do not alter our findings. For instance, across all specifications, the largest predicted treatment effect is 0.23 s.d.

Figure 3: The value added of students' tracks under accurate beliefs



The figure shows how the value added of students' tracks would change under accurate beliefs. It plots the relationship between the percentile rank of the student's transition score and multiple value added variables. The blue (purple) line is the maximum (mean) value added in the student's feasible set. The green and brown lines are the value added predictions, $V_{i,\mathrm{IS}}$ and $V_{i,\mathrm{AS}}$, respectively. All lines are calculated using local linear regressions. The predictions, $V_{i,\mathrm{IS}}$ and $V_{i,\mathrm{AS}}$, are from the "Adjust for measurement error" specification. The sample is as in Table 16.

(Column 3). Meanwhile, in our main simulation—in which households care about a variety of track characteristics—they are predicted to leave 0.69 s.d. unexploited (Column 5). Thus, 84% of the value added left on the table is due to preferences for curriculum and peer quality.[49]

Table 17: The effect of preferences on the amount of unexploited value added

	Potential increase in VA: $V_{i,\mathrm{AS}}$				
	(1)	(2)	(3)	(4)	(5)
All students	0.46	0.22	0.58	0.60	0.69
Low-achieving	0.24	0.02	0.25	0.28	0.55
High-achieving	0.59	0.35	0.77	0.78	0.78

The table shows how preferences for track attributes constrain households' choices with respect to value added. It presents the mean potential increase in value added, under *Accurate scores*, for different versions of the simulation. The versions differ in that they set $\hat{\beta}_q = 0$ for different quality dimensions in calculating $V_{i,\mathrm{AS}}$. All versions use the "Adjust for measurement error" specification. Column (1) is if households care only about value added and curriculum; (2) is if they care only about value added and peer quality; (3) is for value added, curriculum, and peer quality; (4) is for value added, curriculum, peer quality, siblings and friends, and location; (5) is for all the attributes in "Adjust for measurement error", including unexplained idiosyncratic factors. (5) is the same as in Panel D of the second column of Table 16. See Table 16 for details on the sample.

We conclude by comparing the results from our simulation with the experimental treatment effects from Section 3. We note that the simulation results may seem small given the size of the observed treatment effects. For instance, for the students for whom the treatment had an impact—low-achieving students who were not admitted to their two top baseline choices—the

49. There is some heterogeneity by achievement. For low-achieving students, choices are constrained largely by preferences for curriculum (Column 1) and unexplained factors (Column 5).

effect was 0.20 s.d. This value is larger than the largest predicted treatment effect for low-achieving students in Table 16 (0.18 s.d.).[50]

A possible explanation is that the treatment may have affected choices via channels other than the accuracy of households' beliefs. As discussed, expected utility may depend on the precision of beliefs in ways not captured by the quality scores. In addition, providing information may cause households to care more about value added. If so, then the predicted treatment effects in Table 16 do not represent an upper bound on the impact of providing information. Nonetheless, these other channels would have to be sizable in order to change our main finding. In other words, it appears that households are likely to leave substantial academic value added on the table, even under correct beliefs. This is due to their preferences for other track attributes.

5 Conclusion

Recent research studies how to allocate students to schools—for example, how to implement mechanisms that free households from strategizing as they apply to schools. A separate question concerns what incentives the resulting demand patterns generate for schools. If demand reflects households' desire for value added, then schools may feel pressure to raise their value added. By contrast, if demand reflects households' desire for peer quality, then schools might focus on becoming more selective, and so forth (Rothstein 2006; Abdulkadiroglu et al. 2020).

Why might households not always demand the schools that researchers deem most productive? What constraints or factors might lead them to leave value added "on the table"? This paper considers two possibilities. First, households may lack *information*. Value added is difficult to observe, even for researchers with access to ample data. 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 offer a short commute or desirable peers. In this case, households may willingly give up value added in exchange for other dimensions of quality (MacLeod and Urquiola 2019).

Our results suggest that both candidate explanations are relevant. We find that essentially all types of households make school choices that leave value added on the table. In addition, our experiment shows that distributing information can affect households' school rankings, placements, and value added—particularly for households with low-achieving students. That said, our simulations suggest that even correcting all informational shortfalls would leave households far from maximizing school value added.

We note that 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

50. In fact, for low-achieving who were not admitted to their two top baseline choices, the largest predicted treatment effect is only 0.15 s.d. Moreover, it seems unlikely that the treatment caused these households to have fully accurate beliefs about value added.

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 issues for further research. First, one of our robust findings is that households attach great weight to their top school choices—it is difficult to influence their decisions on these. This might generalize to other settings where, at least anecdotally, households tend to prioritize "favorite" schools (e.g., those used by previous generations of the family). Such behavior could reflect aspects related to costly attention (Arteaga et al. 2021). Second, the effects of information on value added 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. 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.

Abdulkadroglu, Atila, Joshua D. Angrist, Yusuke Narita, Parag A. Pathak, and Roman A. Zarate. 2017. "Regression Discontinuity in Serial Dictatorship: Achievement Effects at Chicago's Exam Schools." *American Economic Review* 107, no. 5 (May): 240–45.

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, Francisco 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.

Arteaga, Felipe, Adam J. Kapor, Christopher A. Neilson, and Seth D. Zimmerman. 2021. *Smart Matching Platforms and Heterogeneous Beliefs in Centralized School Choice.* Working Paper. Princeton University, June.

Artemov, Georgy, Yeon-Koo Che, and Yinghua He. 2020. *Strategic 'Mistakes': Implications for Market Design Research.* Working Paper. March.

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.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63 (4): 841–890.

———. 2004. "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market." *Journal of Political Economy* 112 (1): 68–105.

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, 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.

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.

Dobbie, Will, and Jr. Fryer Roland G. 2014. "The Impact of Attending a School with High-Achieving Peers: Evidence from the New York City Exam Schools." *American Economic Journal: Applied Economics* 6, no. 3 (July): 58–75.

Fack, Gabrielle, Julien Grenet, and Yinghua He. 2019. "Beyond Truth-Telling: Preference Estimation with Centralized School Choice and College Admissions." *American Economic Review* 109, no. 4 (April): 1486–1529.

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

Gilraine, Michael, Jiaying Gu, and Robert McMillan. 2020. *A New Method for Estimating Teacher Value-Added.* Working Paper 27094. National Bureau of Economic Research, May.

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.

Jacob, Brian A., and Lars Lefgren. 2008. "Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education." *Journal of Labor Economics* 26 (1): 101–136.

Kane, Thomas J, and Douglas O Staiger. 2008. *Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation.* Working Paper 14607. National Bureau of Economic Research, December.

Kapor, Adam J., Christopher A. Neilson, and Seth D. Zimmerman. 2020. "Heterogeneous Beliefs and School Choice Mechanisms." *American Economic Review* 110, no. 5 (May): 1274–1315.

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.

Kwon, Soonwoo. 2021. *Optimal Shrinkage Estimation of Fixed Effects in Linear Panel Data Models.* Working Paper. January.

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.

Morris, Carl N. 1983. “Parametric Empirical Bayes Inference: Theory and Applications.” *Journal of the American Statistical Association* 78 (381): 47–55.

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.

Petrin, Amil, and Kenneth Train. 2010. “A Control Function Approach to Endogeneity in Consumer Choice Models.” *Journal of Marketing Research* 47 (1): 3–13.

Riehl, Evan, Juan Saavedra, and Miguel Urquiola. 2019. “Learning and earning: An approximation to college value added in two dimensions.” In *Productivity in Higher Education,* edited by Caroline M. Hoxby and Kevin Stange. Chicago, IL: The University of Chicago Press.

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.

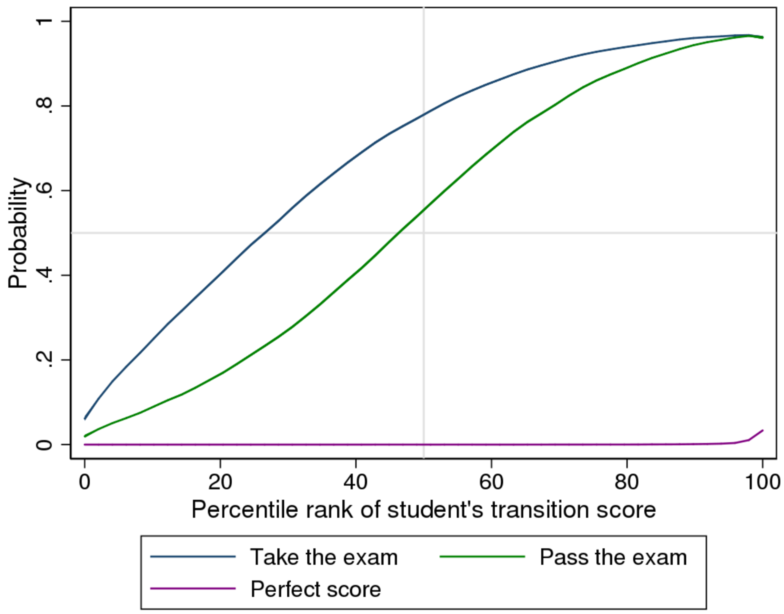
———. 2017. “Measuring the Impacts of Teachers: Comment.” *American Economic Review* 107, no. 6 (June): 1656–84.

Wooldridge, Jeffrey M. 2014. “Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables.” *Journal of Econometrics* 182 (1): 226–234.

———. 2015. “Control Function Methods in Applied Econometrics.” *The Journal of Human Resources* 50 (2): 420–445.

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

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 page of the information sheet provided at the conclusion of the baseline survey. This information was provided to all households.

Figure A3: Page 2 of the information sheet

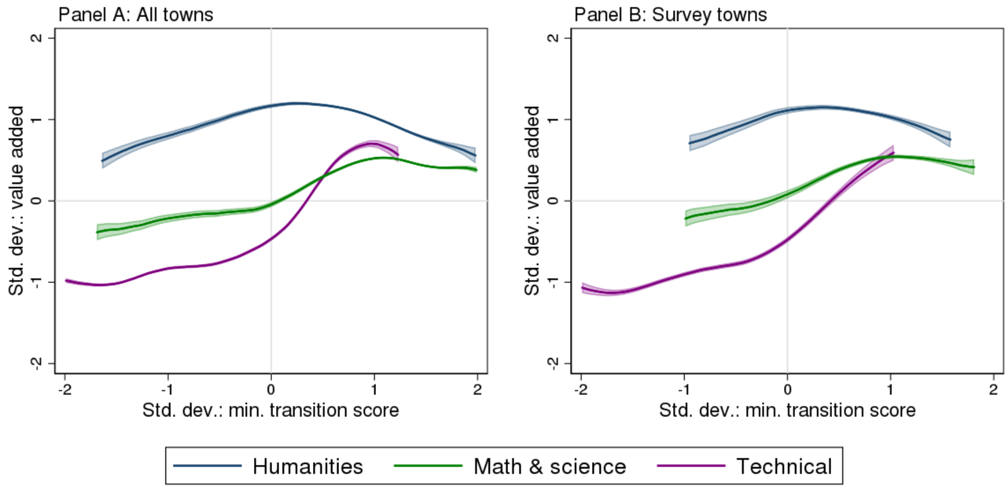
A team of economists at New York University has analyzed data in your hometown, Sebes Alba. They have calculated which tracks most effectively improve students' chances of passing the baccalaureate exam relative to their 9th grade starting points.

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.

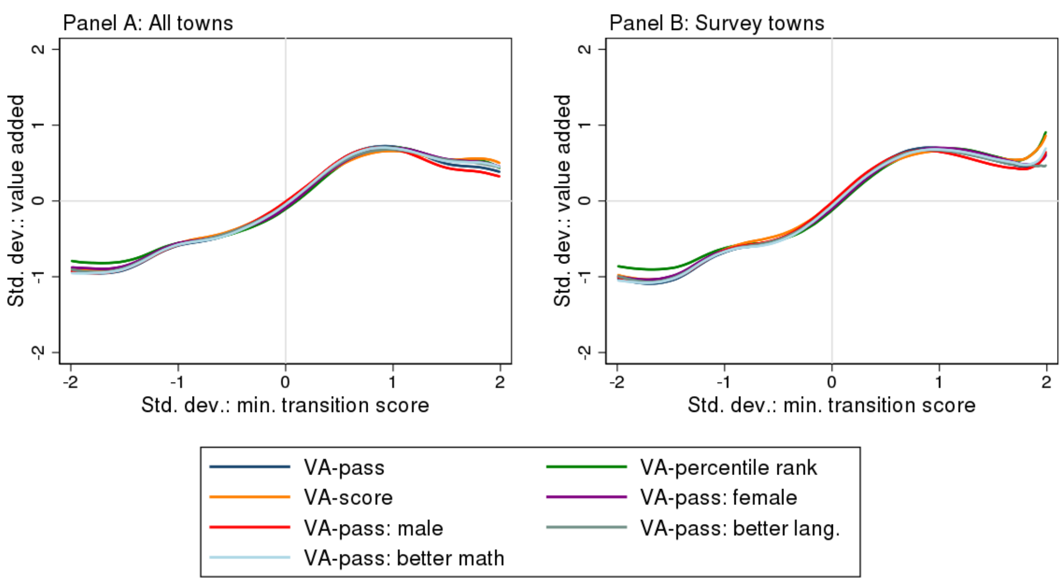
The figure displays the second page of the information sheet provided at the conclusion of the baseline survey. This information was provided only to treated households.

Figure A4: The relationship between value added and selectivity by curricular category



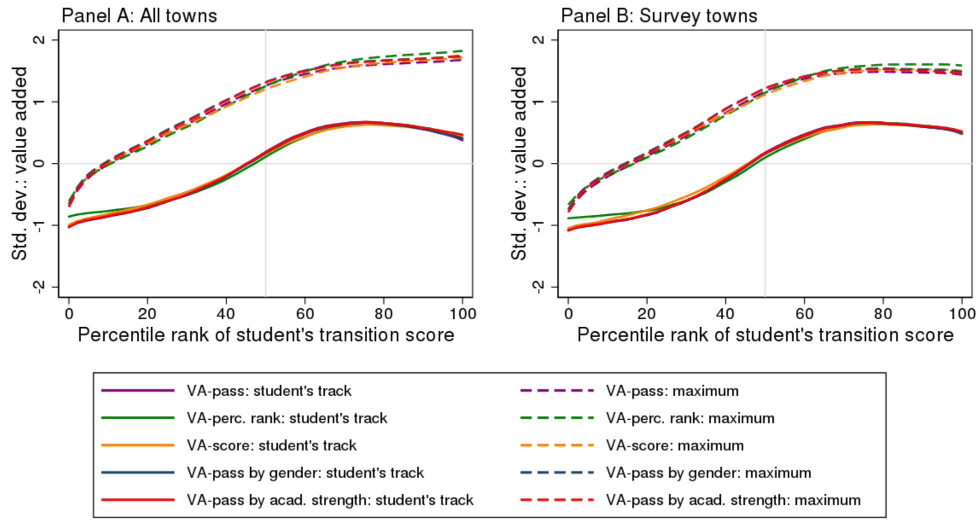
The figure presents the relationship between value added and selectivity for subsets of tracks by a track's curriculum. Specifically, it replicates the local linear regressions in Figure 1 separately for tracks with curricular focuses in humanities, math and science, or technical subjects. See Figure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

Figure A5: The rel. between VA and selectivity: robustness to alternative VA measures



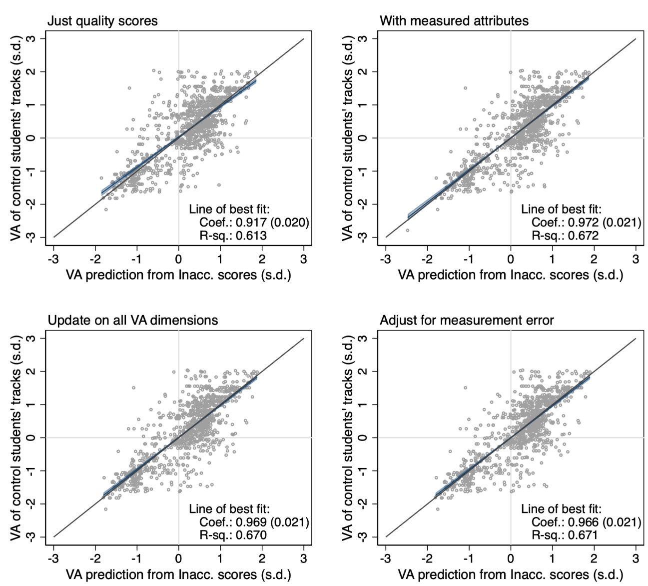
The figure replicates Figure 1 for alternative value added measures. Specifically, it presents local linear regressions of standardized values of various value added measures on standardized values of minimum transition score, MTS_{jt}. The value added measures are: (i) "VA-pass": a track-year effect 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; (iii) "VA-score": a track-year effect on the exam score, and (iv) 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 further details.

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 plot local linear regressions of the maximum value added in the student's feasible set vs. the within-year percentile rank of the student's transition score. The solid lines are local linear regressions of the value added in the track the student attends vs. 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, (iii) a track-year effect on the exam score, and (iv) track-year effects on the probability of passing that vary by a student's gender or relative academic strength. See Section 0.3 and Figure 2 for more details.

Figure A7: Comparing our predictions with the value added of students' actual tracks



The figure illustrates the quality of our predictions for the value added of students' tracks. The plots are for students in the control group. They show how our predictions under inaccurate scores, $V_{i,\mathrm{IS}}$, compare with the value added of the tracks students attend. Each plot presents a different specification. See Section 4.2 for definitions of "Just quality scores", "With measured attributes", "Update on all VA dimensions", and "Adjust for measurement error". 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 each case. The notes to Table 16 discuss the sample.

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
2019	312	1,015	3,038	105,230
Mean	398	1,187	3,835	144,182
Total	5,969	17,808	57,521	2,162,736
Distinct	512	1,401	13,405	2,162,736

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 2004-2017 and 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. We exclude the 2018 cohort due to a reporting issue in that year.

Table A2: Correlations of alternative VA 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
Exam score	0.931	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, (ii) a track-year effect on a student's imputed exam score, and (iii) 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: 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	0.885	16	504	15	476	132	6	2
Alba	Sebes	0.863	10	290	10	297	35	3	0
Arges	Campulung	0.834	13	423	11	420	67	4	0
Bacau	Moinesti	0.798	9	303	9	280	87	3	2
Bacau	Onesti	0.800	16	650	16	637	157	6	2
Bihor	Beius	0.620	11	307	10	322	72	2	2
Bistrita Nasaud	Bistrita	0.809	28	925	23	782	148	7	2
Brasov	Fagaras	0.925	10	323	9	273	117	3	2
Buzau	Ramnicu Sarat	0.731	12	476	13	445	113	4	2
Calarasi	Calarasi	0.918	24	666	20	709	161	8	2
Caras Severin	Resita	0.558	20	473	18	425	103	7	1
Cluj	Dej	0.869	10	300	10	299	80	4	1
Cluj	Gherla	0.659	10	261	10	265	37	2	0
Cluj	Turda	0.759	12	282	11	281	71	5	0
Constanta	Mangalia	0.627	10	336	9	252	145	5	2
Constanta	Medgidia	0.671	10	308	9	280	27	1	0
Covasna	Sfantul Gheorghe	0.831	20	396	20	437	43	2	1
Covasna	Tirgu Secuiesc	0.638	9	219	9	233	43	3	0
Dolj	Calafat	0.798	7	183	6	168	37	2	0
Galati	Tecuci	0.858	18	753	16	728	79	5	0
Giurgiu	Giurgiu	0.889	15	591	14	602	148	9	2
Gorj	Motru	0.582	11	362	9	308	53	3	0
Harghita	Gheorgheni	0.888	11	280	12	263	22	2	0
Harghita	Miercurea Ciuc	0.856	22	602	21	589	48	4	1
Harghita	Odorheiu Secuiesc	0.863	15	392	15	364	39	3	1
Harghita	Toplita	0.756	7	170	8	172	22	2	0
Hunedoara	Deva	0.906	19	369	19	353	102	5	1
Hunedoara	Hunedoara	0.721	11	364	10	308	91	6	0
Hunedoara	Petrosani	0.739	9	299	8	224	101	4	2
Ialomita	Slobozia	0.893	19	636	16	644	91	4	2
Ialomita	Urziceni	0.855	11	316	7	280	59	3	0
Iasi	Harlau	0.876	8	222	7	224	34	2	0
Iasi	Pascani	0.841	17	688	16	644	109	4	2
Iasi	Targu Frumos	0.771	7	222	6	196	49	3	0
Maramures	Sighetu Marmatiei	0.679	21	582	19	565	104	5	2
Mures	Sighisoara	0.818	14	307	14	301	55	4	0
Mures	Tarnaveni	0.567	8	231	8	194	46	3	0
Neamt	Roman	0.845	21	825	19	672	48	3	1
Prahova	Campina	0.741	16	530	16	554	76	4	0
Salaj	Zalau	0.899	22	759	21	741	125	7	1
Satu Mare	Carei	0.792	10	247	8	224	54	3	0
Suceava	Gura Humorului	0.431	8	304	9	289	48	3	0
Suceava	Radauti	0.798	16	672	18	672	114	4	2
Teleorman	Alexandria	0.685	15	699	16	746	88	4	2
Timis	Lugoj	0.543	14	427	12	373	131	6	0
Valcea	Dragasani	0.787	12	328	7	308	108	2	2
Vaslui	Birlad	0.849	20	758	18	694	158	8	2
Vrancea	Adjud	0.802	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	0.431	7	170	6	168	21	1	0
	Max	0.925	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 true value added explained by forecasted value added for the town during 2008-2014 (see the notes to Table A26). "Two-class schools" indicates the number of middle schools in which we visited two classrooms.

Table A4: The frequency with which households assign baseline quality scores of each value

	All students	**Scored all tracks**
Share of scores with a value of:		
1	0.14	0.16
2	0.12	0.14
3	0.16	0.18
4	0.22	0.23
5	0.36	0.29
Students	2,759	819
Student-tracks	20,482	10,501
Scores	142,692	82,070

The table reveals how often households assign quality scores of a given value. Specifically, it shows the share of households' baseline quality scores that are equal to each value from 1 to 5. The results are calculated using scores for all quality dimensions. The results in the column labeled "Scored all tracks" are calculated using a limited sample. Specifically, they are for survey respondents who provided quality scores for both peer quality and value added on passing the baccalaureate exam for all of the tracks in their towns. This sample restriction is useful because it eliminates variation in frequencies related to which tracks households choose to score.

Table A5: Summary statistics for households' baseline quality scores

	Mean	Std. dev.	Min	Max	Students	Student-tracks
Location	3.84	1.31	1	5	2,673	19,959
Siblings & friends	2.87	1.63	1	5	2,091	15,588
Peer quality	3.57	1.37	1	5	2,496	18,478
Curriculum	3.39	1.45	1	5	2,516	18,134
Teacher quality	3.83	1.29	1	5	2,478	17,940
VA: pass the bacc.	3.71	1.36	1	5	2,469	17,882
VA: college	3.51	1.44	1	5	2,406	17,451
VA: wages	3.49	1.39	1	5	2,343	17,260

The table describes households' baseline scores for track characteristics.

Table A6: Correlations between households' baseline quality scores

	Location	Siblings	Peers	Curriculum	Teachers	Pass bacc.	College	Wages
Location	1							
Siblings and friends	0.482	1						
Peer quality	0.571	0.599	1					
Curriculum	0.517	0.604	0.774	1				
Teacher quality	0.569	0.523	0.774	0.726	1			
VA: pass the bacc.	0.543	0.548	0.768	0.764	0.810	1		
VA: college	0.516	0.576	0.778	0.797	0.752	0.861	1	
VA: wages	0.507	0.561	0.726	0.744	0.735	0.809	0.846	1

The table shows correlations between respondents' scores for different track characteristics.

Table A7: 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		Y						
Transition exam–Math score: average	Y								
Transition exam–Romanian score: average	Y		Y						
Transition exam–Romanian score: std. dev.	Y								
Share female	Y		Y			Y	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								
Average transition exam–Rom. score in students' middle schools: average	Y								
Rank of track in school by average transition score	Y								
Rank of track in town by average transition score	Y								
Share of students who took the baccalaureate exam						Y	Y	Y	
VA-pass						Y	Y	Y	Y
Rank of track in town by VA-pass						Y	Y	Y	
VA-pass de-meaned by town-year						Y	Y		
Squared standard error of VA-pass						Y	Y		
VA-pass: female						Y	Y		
VA-pass: male						Y	Y		
VA-pass: better at language						Y	Y		
VA-pass: better at math						Y	Y		
VA-percentile rank						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 A8: 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 humanities tracks	Y								
Number of math or science tracks	Y								
Number of 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								
Transition exam–Romanian score: average	Y								
Share female	Y								
Average transition score in students' middle schools: average	Y								
Share of students who took the baccalaureate exam						Y	Y		
VA-pass: average						Y	Y		
VA-pass: 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 A9: 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 humanities tracks	Y	Y				Y			
Number of math or science tracks	Y	Y				Y			
Number of 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								
Transition exam–Romanian score: average	Y								
Transition exam–Romanian score: std. dev.	Y								
Share of students who took the baccalaureate exam						Y	Y		
VA-pass: average						Y	Y		
VA-pass: 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 A10: A comparison of the samples used in the paper

	Mean	**Std. dev.**	**Students**
Panel A: Administrative data			
Female	0.53	0.50	2,162,736
Transition score	7.70	1.35	2,162,736
Middle school GPA	8.65	0.97	2,162,736
Transition exam score	7.05	1.69	2,162,736
Panel B: Baseline sample			
Female	0.52	0.50	3,898
Transition score	7.72	1.41	3,746
Middle school GPA	9.05	0.85	3,769
Transition exam score	7.35	1.61	3,830
Panel C: Experimental sample			
Female	0.53	0.50	2,692
Transition score	7.87	1.31	2,692
Middle school GPA	9.20	0.68	2,692
Transition exam score	7.54	1.50	2,692
Panel D: Follow-up sample			
Female	0.52	0.50	1,533
Transition score	7.92	1.29	1,533
Middle school GPA	9.23	0.66	1,533
Transition exam score	7.59	1.48	1,533

The table provides summary statistics on the main samples used in the paper. See Section 0 for details on the samples.

Table A11: Year-specific correlations between value added and selectivity

Year	**Coefficient**	**Std. error**	**Towns**	**Tracks**	**Students**
2004	0.627	0.019	426	3,691	185,383
2005	0.468	0.026	405	3,500	146,712
2006	0.512	0.023	386	3,284	136,671
2007	0.558	0.018	383	3,259	134,692
2008	0.565	0.015	476	4,851	172,174
2009	0.635	0.011	438	4,470	170,087
2010	0.593	0.015	417	4,018	164,146
2011	0.607	0.013	437	4,506	187,442
2012	0.613	0.016	410	4,234	146,114
2013	0.574	0.017	420	4,269	141,934
2014	0.504	0.014	378	3,784	124,675
2015	0.529	0.016	368	3,649	121,880
2016	0.524	0.017	362	3,541	115,902
2017	0.501	0.019	351	3,427	109,694
2019	0.486	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, V_{jt}, on standardized values of minimum transition score, MTS_{jt}. The sample includes the full set of towns. See Figure 1 and Table 5 for additional details.

Table A12: 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, V_{jt}	64.2	61.4	66.7	66.1	61.7	70.2
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, V_{jt}	0.55	0.58	0.53	0.45	0.49	0.40
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 6 using a different choice set. The choice set in Table 6 is the set of tracks that a student is eligible to attend (i.e., the student's "feasible set"). 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. See Table 6 for additional details.

Table A13: Effects on the value added of students' tracks: 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	-0.22	0.41	0.62	0.58	0.76	0.65
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 various versions of regression (1). In the first column, the outcome is the value added of the feasible track that the student ranked highest in the baseline survey. This regression is thus a balance test. In the columns labeled "Final," the outcome is the value added of the track the student attends. The results in Column 4 correspond to those in the first column of Table 9. 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 highest at baseline. 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 A14: Effects on the value added of students' tracks, as measured by households' baseline beliefs

	VA: pass the bacc.	**VA: college**	**VA: wages**
Panel A: Eligible for at least one of two top baseline choices			
Treated	0.009	0.009	0.003
	(0.012)	(0.013)	(0.012)
Clusters	78	78	78
Students	1,990	1,990	1,990
Panel B: Ineligible for two top baseline choices			
Treated	0.107*	0.132**	0.117**
	(0.054)	(0.059)	(0.056)
Clusters	76	76	76
Students	515	515	515

The table presents results similar to those in Table 12, but for additional outcome variables. The outcome variables are track-level means for the listed baseline quality scores. As in Table 12, regressions control for the value of the outcome variable for the feasible track that the household ranked the highest in the baseline survey. This is the track to which the student would have been assigned based on the baseline preference ranking. The regressions also include indicators for students who did not rank any feasible tracks in the baseline. The sample is slightly smaller than that in Table 12 because we omit students who attend tracks in different towns from where they attended middle school. We do this because we can't observe households' quality scores for these tracks. Standard errors are clustered by the middle school treatment-control pairs within which we conducted the randomization.

Table A15: 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	≥ 3rd-most-preferred	≥ 4th-most-preferred	≥ 5th-most-preferred	≥ 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 A16: 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	≥ 3rd-most-preferred	≥ 4th-most-preferred	≥ 5th-most-preferred	≥ 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 A17: 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 A18: 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	$\geq$ 3rd-most-preferred	$\geq$ 4th-most-preferred	$\geq$ 5th-most-preferred	$\geq$ 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 A19: 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 (2), as in Table 13. Panel B presents results from regression (3), 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 A20: Households' preferences for track attributes: heterogeneity by achievement

	Low-achieving students					High-achieving students				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Location	0.234**	0.259**	0.222*	0.248**	0.254**	0.333***	0.343***	0.348***	0.340***	0.337***
	(0.116)	(0.112)	(0.113)	(0.111)	(0.113)	(0.0861)	(0.0897)	(0.0871)	(0.0821)	(0.0893)
Siblings and friends	0.335***	0.305***	0.316***	0.351***	0.289***	0.322***	0.328***	0.307***	0.325***	0.315***
	(0.0802)	(0.0756)	(0.0809)	(0.0804)	(0.0772)	(0.0633)	(0.0674)	(0.0639)	(0.0619)	(0.0673)
Peer quality	0.0642	0.103	0.0761	0.126	0.109	0.542***	0.473***	0.508***	0.562***	0.429***
	(0.0931)	(0.0878)	(0.0941)	(0.0903)	(0.106)	(0.0779)	(0.0815)	(0.0820)	(0.0805)	(0.0833)
Curriculum	0.711***	0.611***	0.717***	0.774***	0.616***	1.085***	0.937***	1.009***	1.147***	0.886***
	(0.117)	(0.109)	(0.113)	(0.114)	(0.107)	(0.0917)	(0.0896)	(0.0830)	(0.0923)	(0.0917)
VA: pass the bacc.	0.265**				0.123	0.430***				-0.0365
	(0.111)				(0.101)	(0.109)				(0.114)
VA: college		0.340***			0.170		0.626***			0.447***
		(0.111)			(0.121)		(0.0923)			(0.103)
VA: wages			0.317***		0.227**			0.572***		0.342***
			(0.0969)		(0.0951)			(0.0776)		(0.0892)
Teacher quality				0.0406	-0.103				0.296***	0.0602
				(0.132)	(0.116)				(0.0903)	(0.0957)
R-sq.	0.21	0.21	0.21	0.20	0.22	0.40	0.41	0.40	0.40	0.41
Clusters	119	118	117	119	117	136	135	135	136	135
Students	394	382	387	394	376	776	775	764	774	761
Student-tracks	3,966	3,806	3,889	3,971	3,756	7,609	7,589	7,493	7,602	7,464

The table presents results analogous to those in Table 15, but separately for low- and high-achieving students. See the notes to Table 15 for additional details.

Table A21: Households' preferences for track attributes: robustness to missing baseline quality scores

	Low-achieving students			High-achieving students		
	No imputations	Scored all tracks	Imputations	No imputations	Scored all tracks	Imputations
Location	0.234**	0.138	0.340***	0.333***	0.238***	0.374***
	(0.116)	(0.144)	(0.0913)	(0.0861)	(0.0849)	(0.102)
Siblings and friends	0.335***	0.405***	0.266***	0.322***	0.427***	0.365***
	(0.0802)	(0.118)	(0.0761)	(0.0633)	(0.0693)	(0.0665)
Peer quality	0.0642	-0.0405	-0.153**	0.542***	0.623***	0.806***
	(0.0931)	(0.116)	(0.0679)	(0.0779)	(0.101)	(0.0690)
Curriculum	0.711***	0.901***	1.105***	1.085***	1.106***	0.827***
	(0.117)	(0.133)	(0.0774)	(0.0917)	(0.102)	(0.0732)
VA: pass the bacc.	0.265**	0.255**	0.353***	0.430***	0.330**	0.581***
	(0.111)	(0.125)	(0.114)	(0.109)	(0.137)	(0.0920)
R-sq.	0.21	0.25	0.21	0.40	0.44	0.39
Clusters	119	72	163	136	83	157
Students	394	199	993	776	354	1,671
Student-tracks	3,966	2,663	12,649	7,609	4,575	22,271

The table shows whether the results from the preference model, equation (4), are sensitive to missing values for households' baseline quality scores. "No imputations" are specifications that ignore missing scores. They correspond to the columns labeled (1) in Table A20. "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 15 for additional details on estimating the preference model.

Table A22: Households' preferences for track attributes: robustness to different definitions of the choice set

	Low-achieving students			**High-achieving students**		
	All tracks	Plausible	Feasible	All tracks	Plausible	Feasible
Location	0.340***	0.319***	0.234**	0.374***	0.373***	0.349***
	(0.0913)	(0.0829)	(0.104)	(0.102)	(0.102)	(0.0945)
Siblings and friends	0.266***	0.271***	0.275***	0.365***	0.366***	0.356***
	(0.0761)	(0.0813)	(0.0812)	(0.0665)	(0.0665)	(0.0631)
Peer quality	-0.153**	-0.0550	-0.0546	0.806***	0.807***	0.809***
	(0.0679)	(0.0677)	(0.0913)	(0.0690)	(0.0690)	(0.0680)
Curriculum	1.105***	1.020***	0.954***	0.827***	0.826***	0.802***
	(0.0774)	(0.0788)	(0.104)	(0.0732)	(0.0732)	(0.0704)
VA: pass the bacc.	0.353***	0.388***	0.259**	0.581***	0.581***	0.670***
	(0.114)	(0.110)	(0.114)	(0.0920)	(0.0919)	(0.0956)
R-sq.	0.21	0.22	0.22	0.39	0.39	0.40
Clusters	163	163	152	157	157	157
Students	993	973	788	1,671	1,671	1,649
Student-tracks	12,649	10,614	5,781	22,271	22,267	20,797

The table presents results from equation (4) for different definitions of a household's choice set. The columns labeled "All tracks" use all the tracks in a household's town. They correspond to the columns labeled "Imputations" in Table A21. The columns labeled "Plausible" and "Feasible" exclude tracks that households may have considered out of reach. "Plausible" uses only the tracks for which the prior-year minimum transition score, MTS_{jt-1}, is no more than 1.5 points above a student's transition score, TS_i. "Feasible" uses only the tracks that the student would have been eligible for in the prior year—the tracks for which $\text{TS}_i \geq \text{MTS}_{jt-1}$. All columns impute missing quality scores using a random forest (Section 4.1). See the notes to Table 15 for additional details on estimation.

Table A23: Households' preferences for track attributes: including measured values of track characteristics

	Low-achieving students					High-achieving students				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Households' baseline quality scores:										
Location	0.340***	0.297***	0.329***	0.310***	0.288***	0.374***	0.354***	0.333***	0.395***	0.364***
	(0.0913)	(0.0928)	(0.0930)	(0.0899)	(0.0895)	(0.102)	(0.103)	(0.0984)	(0.0944)	(0.0968)
Siblings and friends	0.266***	0.258***	0.266***	0.263***	0.247***	0.365***	0.353***	0.349***	0.380***	0.362***
	(0.0761)	(0.0767)	(0.0762)	(0.0773)	(0.0776)	(0.0665)	(0.0655)	(0.0652)	(0.0635)	(0.0633)
Peer quality	-0.153**	-0.158**	-0.189***	0.0513	0.0129	0.806***	0.845***	0.462***	0.416***	0.241***
	(0.0679)	(0.0665)	(0.0688)	(0.0706)	(0.0687)	(0.0690)	(0.0697)	(0.0732)	(0.0670)	(0.0666)
Curriculum	1.105***	1.161***	1.131***	1.019***	1.039***	0.827***	0.824***	0.998***	1.061***	1.123***
	(0.0774)	(0.0787)	(0.0782)	(0.0806)	(0.0802)	(0.0732)	(0.0718)	(0.0666)	(0.0674)	(0.0678)
VA: pass the bacc.	0.353***	0.204*	0.323***	0.281**	0.266**	0.581***	0.515***	0.383***	0.404***	0.324***
	(0.114)	(0.118)	(0.113)	(0.115)	(0.115)	(0.0920)	(0.0904)	(0.0894)	(0.0911)	(0.0894)
Measured track characteristics:										
Value added,		0.267***			0.152**		0.159***			0.287***
V_{jt} (s.d.)		(0.0394)			(0.0672)		(0.0524)			(0.0812)
Selectivity,			0.105		0.126			1.200***		0.757***
MTS_{jt} (s.d.)			(0.0655)		(0.0826)			(0.137)		(0.139)
Humanities				0.286***	-0.0941				1.214***	0.383**
				(0.108)	(0.149)				(0.199)	(0.185)
Math or science				-0.430***	-0.655***				1.851***	1.208***
				(0.118)	(0.126)				(0.212)	(0.206)
R-sq.	0.21	0.21	0.21	0.22	0.22	0.39	0.39	0.42	0.41	0.43
Clusters	163	163	163	163	163	157	157	157	157	157
Students	993	993	993	993	993	1,671	1,671	1,671	1,671	1,671
Student-tracks	12,649	12,649	12,649	12,649	12,649	22,271	22,271	22,271	22,271	22,271

The table presents versions of the preference model, equation (4), that control for measured values of track characteristics. "Humanities" and "Math or science" are indicators for a track's curricular focus. The omitted category is for technical tracks. All columns impute missing quality scores using a random forest, as described in Section 4.1. See the notes to Table 15 for additional details on estimating the preference model.

Table A24: The preference models used in the simulation

	Low-achieving students			High-achieving students		
	(1)	(2)	(3)	(1)	(2)	(3)
Households' baseline quality scores:						
Location	0.340***	0.288***	0.293***	0.374***	0.364***	0.390***
	(0.0913)	(0.0895)	(0.0891)	(0.102)	(0.0968)	(0.0981)
Siblings and friends	0.266***	0.247***	0.244***	0.365***	0.362***	0.355***
	(0.0761)	(0.0776)	(0.0787)	(0.0665)	(0.0633)	(0.0657)
Peer quality	-0.153**	0.0129	-0.0113	0.806***	0.241***	0.189***
	(0.0679)	(0.0687)	(0.0702)	(0.0690)	(0.0666)	(0.0669)
Curriculum	1.105***	1.039***	0.997***	0.827***	1.123***	1.023***
	(0.0774)	(0.0802)	(0.0869)	(0.0732)	(0.0678)	(0.0743)
VA: pass the bacc.	0.353***	0.266**	0.181	0.581***	0.324***	0.0676
	(0.114)	(0.115)	(0.114)	(0.0920)	(0.0894)	(0.0935)
VA: college			0.130			0.347***
			(0.100)			(0.0829)
VA: wages			0.0806			0.104
			(0.105)			(0.0803)
Teacher quality			-0.0525			-0.0661
			(0.105)			(0.105)
Measured track characteristics:						
Value added, V_{jt} (s.d.)		0.152**	0.158**		0.287***	0.275***
		(0.0672)	(0.0674)		(0.0812)	(0.0818)
Selectivity, MTS_{jt} (s.d.)		0.126	0.123		0.757***	0.737***
		(0.0826)	(0.0822)		(0.139)	(0.133)
Humanities		-0.0941	-0.0715		0.383**	0.453**
		(0.149)	(0.150)		(0.185)	(0.186)
Math or science		-0.655***	-0.652***		1.208***	1.221***
		(0.126)	(0.128)		(0.206)	(0.207)
R-sq.	0.21	0.22	0.22	0.39	0.43	0.44
Clusters	163	163	163	157	157	157
Students	993	993	993	1,671	1,671	1,671
Student-tracks	12,649	12,649	12,649	22,271	22,271	22,271

The table presents results for the preference models used in the simulations in Section 4.2. The columns labeled 1 are for the specification titled "Just quality scores". Columns 2 are for the "With measured attributes" specification. Columns 3 are for the "Update on all value added dimensions" specification. Finally, the "Adjust for measurement error" specification uses the same coefficients as in Columns 2. However, it inflates the coefficient on "VA: pass the bacc." by a factor of 1.5. All columns impute missing quality scores using a random forest, as described in Section 4.1. See the notes to Table 15 for additional details on estimating the preference model.

Table A25: The effect of accurate beliefs on the value added of students' tracks: robustness

	Change in value added: $V_{i,AS} - V_{i,IS}$		
	All students	Low-achieving	High-achieving
Panel A: 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
Plausible: Top 2	0.198	0.186	0.204
Feasible: Top 2	0.193	0.130	0.230
Panel B: With measured attributes			
Top 1	0.118	0.151	0.099
Top 2	0.113	0.125	0.105
Top 3	0.109	0.125	0.099
Top 4	0.115	0.131	0.106
Plausible: Top 2	0.105	0.106	0.105
Feasible: Top 2	0.103	0.075	0.120
Panel C: Update on all VA dimensions			
Top 1	0.208	0.225	0.198
Top 2	0.160	0.165	0.157
Top 3	0.141	0.143	0.140
Top 4	0.137	0.130	0.141
Plausible: Top 2	0.151	0.140	0.157
Feasible: Top 2	0.178	0.192	0.169
Panel D: Adjust for measurement error			
Top 1	0.169	0.214	0.142
Top 2	0.162	0.180	0.151
Top 3	0.156	0.180	0.142
Top 4	0.165	0.189	0.151
Plausible: Top 2	0.152	0.153	0.151
Feasible: Top 2	0.148	0.110	0.171

The table presents the mean difference between $V_{i,AS}$ and $V_{i,IS}$ for alternative specifications of the preference model. See Section 4.2 for definitions of "Just quality scores", "With measured attributes", "Update on all VA dimensions", and "Adjust for measurement error". The rows within each panel represent specifications in which we estimate the preference model in different ways. "Top 1" fits the rank-ordered logit using just a household's top choice, "Top 2" uses the household's two top choices, and analogously for "Top 3" and "Top 4". Each of these specifications defines a choice set using all the tracks in a household's town. "Plausible: Top 2" and "Feasible: Top 2" fit the rank-ordered logit using a household's two top choices among tracks that the household could reasonably have expected to be feasible at the time of the baseline survey. Specifically, "Plausible" defines a choice set as the tracks for which the prior-year minimum transition score, MTS_{jt-1}, is no more than 1.5 points above the student's transition score, TS_i. "Feasible" uses only the tracks that the student would have been eligible for in the prior year. These are the tracks for which $TS_i \geq MTS_{jt-1}$. See Footnote 38 for additional details on fitting the rank-ordered logit. See the notes to Table 16 for details on the sample.

A1 Value added

This appendix presents an overview of our methodology for calculating value added. We provide additional details in Appendices A2 and A3.

A1.1 Estimating value added

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 other outcomes, p_i is defined as in Section 0.3.

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, such as gender and the components of the transition score. We estimate value added by regressing p_i on a set of track attendance dummies, d_{ij}, and on flexible controls for covariates, $f(X_i)$.[51] We allow both value added and the effects of controls to vary by year. Thus, for each cohort, we fit the model:

$$p_i = {\gamma_t}' \cdot f(X_i) + \sum_j \mathrm{V}^*_{jt} \cdot d_{ij} + e_i, \quad i \in \mathcal{I}_t. \tag{6}$$

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

Equation (6) 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.[52] Specifically, let g index the group that a student falls into, either by gender or relative academic strength. We fit:

$$p_i = {\gamma_{gt}}' \cdot f(X_i) + \sum_j \mathrm{V}^*_{jgt} \cdot d_{ij} + e_i, \quad i \in \mathcal{I}_{gt}. \tag{7}$$

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.[53]

51. 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.

52. 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.

53. There is a complication when calculating value added on passing the exam. For this measure, p_i is a binary variable, and equations (6) and (7) are linear probability models. As such, they 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 is reasonable for students with a moderate chance of passing; however, it is less plausible for students with either a very high or very low chance. To test the impact of the assumption, we have fit versions of (6) and (7) using a logit. This alternative specification assumes that a track exerts a constant effect on the index function for the probability of passing, not on the raw probability itself. The results are hardly changed.

A1.2 Validating value added

Value added measures that rely on the selection-on-observables assumption may suffer from bias. Notably, they 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. Prior work finds that selection-on-observables value added measures often closely approximate causal effects (Rothstein 2010, 2017; Chetty, Friedman, and Rockoff 2014; Deming 2014; Angrist et al. 2017). Nonetheless, 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 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, the measures that rely on a single track-year effect perform as well as those that allow for treatment effect heterogeneity by gender or by relative academic strength. Nonetheless, due to our large sample size, we are able to reject an over-identification test that, for each cutoff, the RDs on the baccalaureate and value added outcomes are always the same.[54]

A1.3 Empirical Bayes posteriors and machine learning forecasts

We face two challenges in working with value added. First, value added estimates, $\hat{\mathrm{V}}_{jt}$, contain measurement error. Second, in our experimental intervention, we need to predict value added for

54. One might wonder why we use value added in our analysis, rather than working directly with RD effects. There are two reasons. First, the RD effects are much noisier. The RD treatment effect of attending track c is an IV quantity which is equal to the ratio of the reduced-form RD effect of being eligible to attend the track by the first-stage RD effect on the probability of attending the track. Calculating this quantity involves dividing one noisy estimate by another, which leads to substantial imprecision. Second, the RD 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 both on tracks' causal effects and on the share of students who "fall back" to each of the less selective tracks if not admitted to track c. Thus, the RDs do not allow us to easily compare tracks in the way that we can with value added.

cohorts of students for whom we cannot yet observe baccalaureate outcomes.[55] For these students, we cannot directly estimate value added using equations (6) or (7).

We deal with the first issue by calculating Empirical Bayes (EB) posterior means, $\text{V}_{jt}^{\text{EB}}$. We calculate these for the 2004-2014 cohorts, the years for which we can observe baccalaureate outcomes and thus estimate value added. Empirical Bayes strategies have been widely used in the value added literature (Kane and Staiger 2008; Jacob and Lefgren 2008; Chetty, Friedman, and Rockoff 2014; Angrist et al. 2017; Abdulkadiroglu et al. 2020). They account for measurement error in noisy estimates via shrinkage. In our implementation, we use the procedure of Morris (1983), which we discuss in Appendix A3.2.

We deal with the second issue by using machine learning to forecast value added four years into the future. Specifically, we predict a track's value added for a given cohort using only the information available at the time of track choice. We obtain these predictions using a local linear forest (Athey et al. 2019).[56] Our model incorporates current and lagged values of a large number of track covariates. Notably, this includes a track's prior value added, its curriculum, and its past and current selectivity and demographics. Tables A7-A9 list the full set of covariates and lags. The first presents the covariates that relate to the track itself, the second displays covariates that relate to the track's high school, and the third shows covariates of the track's town.

We make forecasts, V_{jt}^{P}, for the 2008-2017 and 2019 cohorts.[57] 2015-2017 and 2019 are the years in which we cannot observe baccalaureate outcomes. 2008-2014 allow us to gauge the degree of forecast error. As explained in Appendix A3.3, in these years we calculate out-of-sample R-squared in predicting true value added, V_{jt}^{*}, using the forecasts, V_{jt}^{P}. The results are presented in the "R-sq" column of Table A26. They show that our model has substantial predictive power: in each year, the forecasts predict about 80% of the variation in true value added.

In the analysis, we use a variable which we label V_{jt}. For 2004-2014, V_{jt} is equal to the Empirical Bayes posteriors, $\text{V}_{jt}^{\text{EB}}$. For 2015-2017 and 2019, it is equal to the machine learning forecasts, V_{jt}^{P}.

A1.4 The magnitude of value added

Table A26 describes the magnitude of value added. The results are for our main measure—track-year effects on passing the baccalaureate exam. Specifically, the column labeled V_{jt} presents year-specific standard deviations for the value added variable that we use in analysis. The column titled V_{jt}^{*} displays standard deviations for the unobservable "true effects". For 2004-2014, these values are calculated by adjusting the standard deviations of $\hat{\text{V}}_{jt}$ for measurement error. For 2015-2017 and 2019, they are calculated by adjusting the standard deviations of V_{jt}^{P} for forecast error.[58] Finally, as a point of comparison, 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 A26 reveals that tracks vary widely in both pass rates and value added.[59] For the 2008-

55. Specifically, 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 high school choices—it cannot be observed until students take the baccalaureate exam, at the end of their high school careers.

56. 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.

57. For 2004-2007, we lack sufficient prior data to compute lagged values of covariates.

58. The procedures used in making these adjustments are described in Appendices A3.1 and A3.4.

59. In the results, we distinguish between three groups of cohorts: 2004-2007, 2008-2014, and 2015-2019. The 2004-2007 cohorts featured frequent instances of cheating. Beginning with the 2008 cohort, the government cracked down on cheating by installing video surveillance in exam centers, and by drastically increasing punishments. These

Table A26: Summary statistics for value added on passing the baccalaureate exam

Years	Standard deviation					R-sq.	Towns	Tracks	Students
	p_{jt}	V^*_{jt}	V_{jt}	V^{EB}_{jt}	V^P_{jt}				
2004	0.318	0.206	0.200	0.200	-	-	426	3,691	185,383
2005	0.256	0.163	0.151	0.151	-	-	405	3,500	146,712
2006	0.289	0.194	0.185	0.185	-	-	386	3,284	136,671
2007	0.350	0.214	0.208	0.208	-	-	383	3,259	134,692
2008	0.365	0.187	0.183	0.183	0.167	0.822	476	4,851	172,174
2009	0.369	0.153	0.146	0.146	0.134	0.794	438	4,470	170,087
2010	0.365	0.137	0.130	0.130	0.118	0.751	417	4,018	164,146
2011	0.364	0.130	0.123	0.123	0.114	0.765	437	4,506	187,442
2012	0.374	0.123	0.115	0.115	0.112	0.801	410	4,234	146,114
2013	0.372	0.114	0.105	0.105	0.105	0.786	420	4,269	141,934
2014	0.356	0.125	0.116	0.116	0.111	0.790	378	3,784	124,675
2015	-	0.123	0.110	-	0.110	-	368	3,649	121,880
2016	-	0.120	0.107	-	0.107	-	362	3,541	115,902
2017	-	0.120	0.107	-	0.107	-	351	3,427	109,694
2019	-	0.120	0.107	-	0.107	-	312	3,038	105,230
2004-2007	0.314	0.196	0.188	0.188	-	-	1,600	13,734	603,458
2008-2014	0.371	0.142	0.135	0.135	0.126	0.791	2,976	30,132	1,106,572
2015-2019	-	0.121	0.108	-	0.108	-	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, V^*_{jt} is the track's (unobserved) true value added, and V_{jt} is the value added variable used in the analysis. For 2004-2014, V_{jt} is equal to the EB posteriors, V^{EB}_{jt}. For 2015-2017 and 2019, it is equal to the machine learning forecasts, V^P_{jt}. See Sections A1.3 and A3.2 for details. See Appendices A3.1 and A3.4 for how we calculate the standard deviation of V^*_{jt}. "R-sq." is the fraction of the variation in V^*_{jt} that is predicted by V^P_{jt}. It is an out-of-sample measure of prediction quality. This is because the forecasts are calculated using trees in the random forest that do not include the track-year being predicted. For details on the calculation of R-squared, see Appendix A3.3. All values are weighted by student.

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 these same cohorts, a one standard deviation increase in true value added, V^*_{jt}, is equivalent to a 14 percentage point increase in the probability of passing. Thus, in these years, value added explains 15% of the variation in pass rates. For the 2004-2007 cohorts, variation in pass rates is smaller and that in true value added is larger. In this early period, value added explains 39% of the variation in pass rates.[60] For the 2015-2019 cohorts, the variation in value added is slightly smaller than it is for 2008-2014. For these more recent years, a one standard deviation increase in true value added is equivalent to a 12 percentage point increase in the probability of passing. Finally, the results for V_{jt} show that standard deviations for this variable are similar to those for the true effects, V^*_{jt}.

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.

measures were highly successful (Borcan, Lindahl, and Mitrut 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. The 2015-2019 cohorts are the students for whom we must forecast value added.

60. Again, this large percentage could be partly due to cheating.

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 in cohort t. Let $\mathcal{F}_t^c$ be the set of "fallback" tracks to track c in cohort t. These are tracks in town l with admissions cutoffs (or minimum transition scores) that in cohort t are lower than that of track c: $\text{MTS}_{ft} < \text{MTS}_{ct}\ \forall\ f \in \mathcal{F}_t^c$. Calculate a running variable, m_i^c, for student i as the difference between the student's transition score, TS_i, and the track's minimum transition score: $m_i^c \equiv \text{TS}_i - \text{MTS}_{ct}$. Next, let $z_i^c \in \{0,1\}$ be an offer to attend track c, which the student receives if his or her value of the running variable is positive, $m_i^c > 0$.[61] Finally, let $d_{ij}^c(z)$ denote whether student i would attend track j under $z_i^c = z$.

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: $d_{if}^c(0) = d_{ic}^c(1) = 1$. 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: $d_{if}^c(0) = d_{if}^c(1) = 1$.

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 group g. This quantity is:

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

Here, y represents a generic outcome and $\mathcal{I}_{lgt}$ is the set of students in town l 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 $\mathcal{I}_{lgt}$. 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,

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

61. 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.

are equal to the share of students at the cutoff who are type-f compliers.[62]

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 that the student attends: $\hat{\mathrm{V}}_i$. These RDs capture the following quantities:

$$\mathrm{RD}^p_{cgt} = \sum_f \mathrm{E}[p_{ic} - p_{if} | d^c_{if}(0) = d^c_{ic}(1) = 1, m^c_i = 0, i \in \mathcal{I}_{lgt}] \cdot \omega^c_{fgt}$$
$$\mathrm{RD}^{\mathrm{V}}_{cgt} = \sum_f (\hat{\mathrm{V}}_{cgt} - \hat{\mathrm{V}}_{fgt}) \cdot \omega^c_{fgt}.$$

Here, p_{ij} is the potential baccalaureate outcome from attending track j, $\hat{\mathrm{V}}_{jgt}$ is track j's value added for students in group g in cohort t, and ω^c_{fgt} 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 $\mathrm{E}[p_{ic} - p_{if} | d^c_{if}(0) = d^c_{ic}(1) = 1, m^c_i = 0, i \in \mathcal{I}_{lgt}] = \hat{\mathrm{V}}_{cgt} - \hat{\mathrm{V}}_{fgt}$. Thus, under these conditions—and with infinite data—RDs calculated on the two outcomes would be the same.

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:

$$\mathrm{RD}^y_c = \sum_t \sum_g \mathrm{RD}^y_{cgt} \cdot \Pr[i \in \mathcal{I}_{lgt} | m^c_i = 0, i \in \mathcal{I}_l]$$

for $y \in \{p, \mathrm{V}\}$. 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_i and $\hat{\mathrm{V}}_i$) 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:

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

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

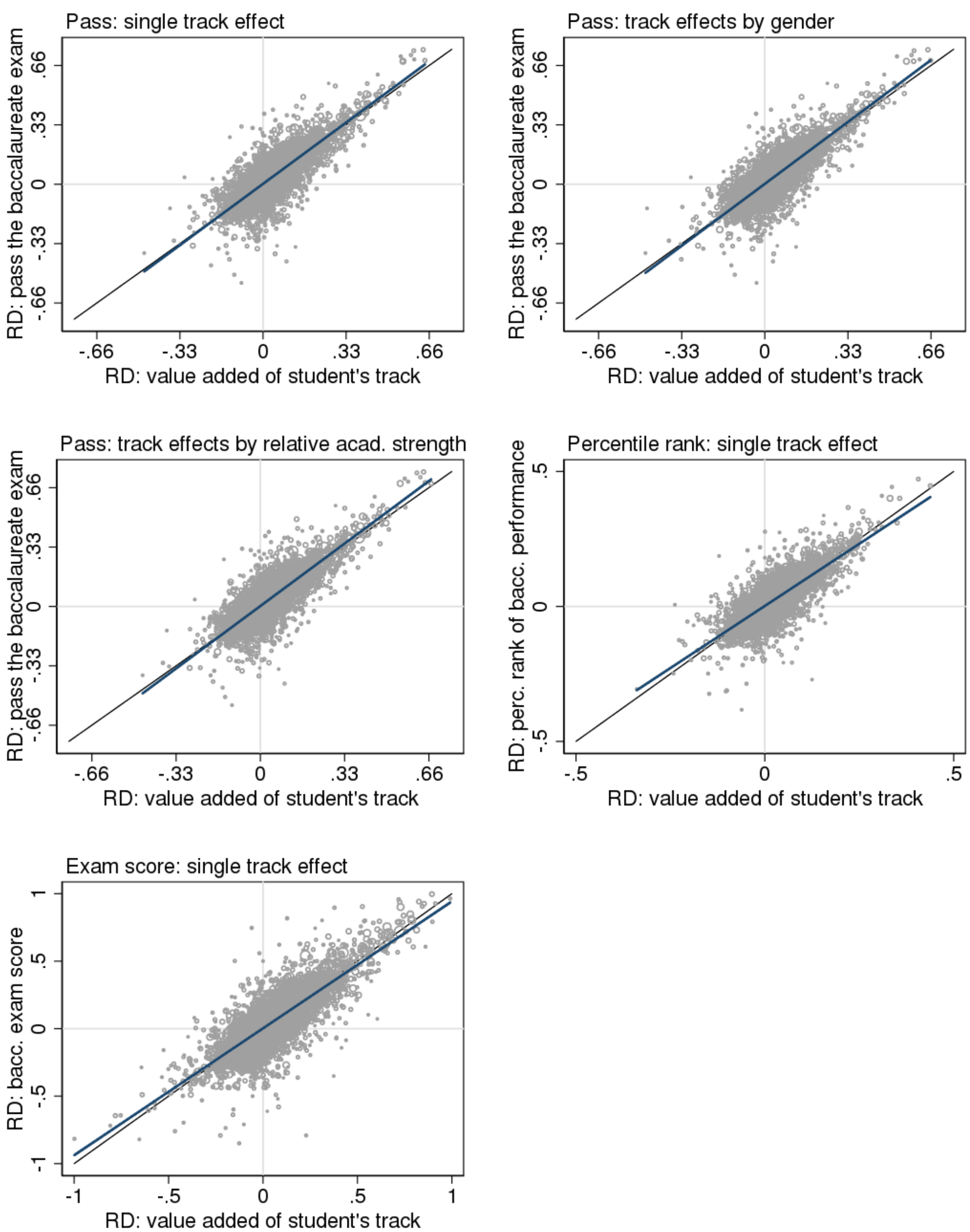
A2.5 Comparing RDs

We then compare the RDs for the two outcomes. Figure A8 plots estimated RDs for baccalaureate outcomes, $\hat{\mathrm{RD}}^p_c$, versus those for the value added of students' tracks, $\hat{\mathrm{RD}}^{\mathrm{V}}_c$. The figure includes plots for a variety of combinations of baccalaureate outcomes and value added measures. The

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

top-left, top-right, and middle-left plots are for the baccalaureate outcome of whether the student passes the exam. These plots use value added measures of, respectively, a single track-year effect on the probability of passing, track-year effects on this probability that vary by gender, and those that vary by relative academic strength. The middle-right plot is for the percentile rank of a student's baccalaureate performance; it uses a value added measure of a single track-year effect on this alternative baccalaureate outcome. Finally, the bottom-left plot is for a student's imputed exam score; it again uses a value added measure of a single track-year effect on this outcome.

Figure A8: Admissions-cutoff RDs



The figure plots estimates of admissions-cutoff RDs on baccalaureate outcomes, $\hat{\mathrm{RD}}_c^p$, versus those on the value added of students' tracks, $\hat{\mathrm{RD}}_c^V$. 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.

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. One can see 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.

Table A27 assesses 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:

$$R^2_{\text{raw}} = 1 - \frac{\sum_c \frac{N_c}{N}(\hat{\text{RD}}^p_c - \hat{\text{RD}}^{\text{V}}_c)^2}{\sum_c \frac{N_c}{N}(\hat{\text{RD}}^p_c - \sum_c \frac{N_c}{N}\hat{\text{RD}}^p_c)^2}. \tag{8}$$

Here, N_c is the number of students in the estimation sample for cutoff c (i.e., $i \in \mathcal{I}_l$ with $|m_i^c| \leq 1$ and $m_i^c \neq 0$), and N is the sum of the number of students in all cutoffs' estimation samples. 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 $\hat{\text{RD}}^y_c = \text{RD}^y_c + \varepsilon^y_c$, where ε^y_c is measurement error. The true (or adjusted) R-squared is:

$$R^2_{\text{adj.}} = 1 - \frac{\sum_c \frac{N_c}{N}[(\hat{\text{RD}}^p_c - \hat{\text{RD}}^{\text{V}}_c)^2 - (\varepsilon^p_c)^2 + 2 \cdot \varepsilon^p_c \cdot \varepsilon^{\text{V}}_c - (\varepsilon^{\text{V}}_c)^2]}{\sum_c \frac{N_c}{N}[(\hat{\text{RD}}^p_c - \sum_c \frac{N_c}{N}\hat{\text{RD}}^p_c)^2 - (\varepsilon^p_c)^2]}, \tag{9}$$

To calculate (9), we replace $(\varepsilon^y_c)^2$ with the squared standard error for $\hat{\text{RD}}^y_c$. Following Appendix C.3.2 of Chandra et al. (2016), we recover $\varepsilon^p_c \cdot \varepsilon^{\text{V}}_c$ 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** Raw	**R-squared** Adjusted	**Cutoffs**	**Student-cutoffs**
Pass the exam:				
All	0.748	0.994	10,210	24,173,143
Gender	0.754	0.996	10,210	24,173,143
Relative academic strength	0.748	0.986	10,210	24,173,143
Percentile rank of exam performance	0.701	0.964	10,210	24,173,143
Exam score	0.713	0.970	10,210	24,173,143

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 (8). Adjusted R-squared is defined in equation (9).

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 in Figure A8 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-year effect on this outcome. For this specification, the estimated RDs on value added explain 75% 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 two rows are value added measures for outcomes related to a student's exam score. They are track-year effects on the percentile rank of the student's exam performance and on a version of the student's exam score that imputes missing values. For these outcomes, the adjusted R-squared remains extremely high (0.964 and 0.970, respectively).

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 $\hat{\mathrm{V}}_i$ (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 $\hat{\mathrm{V}}_i$ 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 $\hat{\mathrm{V}}_i$ 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: 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 A2.6. 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 $\hat{\mathrm{V}}_i$ in an instrumental variables regression of p_i on $\hat{\mathrm{V}}_i$, 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.

Table A28 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.

A3 Adjusting for measurement error

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

A3.1 The standard deviation of $\mathbf{V}^*_{jt}$ based on $\widehat{\mathbf{V}}_{jt}$

When fitting equations (6) and (7), we obtain value added estimates, $\hat{\mathrm{V}}_{jt}$, rather than the true values, V^*_{jt}. Nonetheless, we can use $\hat{\mathrm{V}}_{jt}$ to estimate the standard deviation of true value added.

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:

$$\mathrm{Var}[\hat{\mathrm{V}}_{jt}] = \mathrm{Var}[\mathrm{V}^*_{jt}] + \mathrm{Var}[\varepsilon_{jt}],$$

or alternatively:

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

$\mathrm{Var}[\varepsilon_{jt}]$ can be estimated as the average of the squared standard errors for the $\hat{\mathrm{V}}_{jt}$ estimates. Thus, we can estimate the standard deviation of true value added by simply subtracting the average squared standard error from the sample variance of estimated value added and taking the square root. For tracks in set $\mathcal{S}$, we use the finite-sample formula:

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

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 $\hat{\varepsilon}^2_{jt}$ is the squared standard error for $\hat{\mathrm{V}}_{jt}$.

A3.2 Empirical Bayes posterior means based on $\widehat{\mathbf{V}}_{jt}$

We adjust individual value added estimates for measurement error by calculating Empirical Bayes posterior means, $\mathrm{V}^{\mathrm{EB}}_{jt}$. To do so, we make slightly stronger assumptions than those in Section A3.1. First, we assume the measurement error, ε_{jt}, is not just independent but also has a normal distribution; i.e.,

$$\varepsilon_{jt} \sim N(0, \mathrm{Var}[\varepsilon_{jt}|jt]),$$

where $\mathrm{Var}[\varepsilon_{jt}|jt]$ is the variance of the measurement error for track jt. As a result, the value added estimates are independently normally distributed around the true effects: $\hat{\mathrm{V}}_{jt} \sim N(\mathrm{V}^*_{jt}, \mathrm{Var}[\varepsilon_{jt}|jt])$.[63]

63. Note that, by the asymptotic normality of OLS, this assumption holds whenever the sample size is large. This seems reasonable given we have both a large total number of students and a sizable number of students per track.

Next, we assume that, in each year, the true effects are also independently normally distributed.[64] Further, they have a common variance, and they are centered around a grand mean which we allow to vary by curricular category:

$$\mathrm{V}^*_{jt} \sim N(\mu_{c(j),t}, \sigma_t^2).$$

Here, $\mu_{c(j),t}$ is the mean of true value added in cohort t for tracks with curricular category c, and σ_t^2 is the variance of $\mathrm{V}^*_{jt} - \mu_{c(j),t}$.

Given these assumptions, the posterior distributions of the true effects are:

$$\mathrm{V}^*_{jt} \mid \{\hat{\mathrm{V}}_{jt}, \mathrm{Var}[\varepsilon_{jt}|jt], \mu_{c(j),t}, \sigma_t^2\} \sim N\left(\mathrm{V}^{*,\mathrm{EB}}_{jt}, (1-b_{jt}) \cdot \mathrm{Var}[\varepsilon_{jt}|jt]\right),$$

with

$$\mathrm{V}^{*,\mathrm{EB}}_{jt} = (1-b_{jt}) \cdot \hat{\mathrm{V}}_{jt} + b_{jt} \cdot \mu_{c(j),t} \quad \text{and} \quad b_{jt} = \frac{\mathrm{Var}[\varepsilon_{jt}|jt]}{\mathrm{Var}[\varepsilon_{jt}|jt] + \sigma_t^2}.$$

We estimate the posterior means, $\mathrm{V}^{*,\mathrm{EB}}_{jt}$, using the procedure in Section 5 of Morris (1983). First, we estimate $\mathrm{Var}[\varepsilon_{jt}|jt]$ with the squared standard error of $\hat{\mathrm{V}}_{jt}$, $\hat{\varepsilon}^2_{jt}$. Second, we estimate $\mu_{c(j),t}$ as the student-weighted average of $\hat{\mathrm{V}}_{jt}$ for tracks in cohort t with curricular category c:

$$\hat{\mu}_{c(j),t} = \sum_{jt \in \mathcal{S}_{ct}} \frac{N_{jt}}{N_{\mathcal{S}_{ct}}} \hat{\mathrm{V}}_{jt}.$$

Here, N_{jt} is the number of students in track j in cohort t, $\mathcal{S}_{ct}$ is the set of tracks in the cohort with curricular category c, and $N_{\mathcal{S}_{ct}}$ is the number of students in this set. Third, we estimate σ_t^2 as:

$$\hat{\sigma}_t^2 = \sum_{jt \in \mathcal{S}_t} \frac{N_{jt}}{N_{\mathcal{S}_t}} \left[\left(\frac{|\mathcal{S}_t|}{|\mathcal{S}_t| - C}\right) \cdot (\hat{\mathrm{V}}_{jt} - \hat{\mu}_{c(j),t})^2 - \hat{\varepsilon}^2_{jt}\right].$$

Here, $\mathcal{S}_t$ is the set of tracks in cohort t, $|\mathcal{S}_t|$ is the number of tracks in this set, and C is the number of curricular categories. Fourth, we estimate b_{jt} as:

$$\hat{b}_{jt} = \left(\frac{|\mathcal{S}_t| - C - 2}{|\mathcal{S}_t| - C}\right) \cdot \left(\frac{\hat{\varepsilon}^2_{jt}}{\hat{\varepsilon}^2_{jt} + \hat{\sigma}^2_t}\right).$$

Finally, we estimate the posterior mean, $\mathrm{V}^{*,\mathrm{EB}}_{jt}$, as

$$\mathrm{V}^{\mathrm{EB}}_{jt} = (1-\hat{b}_{jt}) \cdot \hat{\mathrm{V}}_{jt} + \hat{b}_{jt} \cdot \hat{\mu}_{c(j),t}.$$

A3.3 R-squared for predicting $\mathbf{V}^*_{jt}$ using $\mathbf{V}^P_{jt}$

We assess the quality of the machine learning forecasts, V^P_{jt}, by examining how well they predict value added in years in which we can estimate value added. In this exercise, we are interested in prediction quality for true value added, V^*_{jt}, not estimated value added, $\hat{\mathrm{V}}_{jt}$. Specifically, the metric that we want is R-squared in predicting true value added:

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

64. It is common in education and health applications to assume that the true effects follow a normal distribution (e.g., Kane and Staiger (2008), Jacob and Lefgren (2008), Chandra et al. (2016), Angrist et al. (2017), and Abdulkadiroglu et al. (2020)). See Gilraine, Gu, and McMillan (2020) or Kwon (2021) for discussions.

$\text{Var}[\text{V}^*_{jt}]$ can be estimated using the approach explained in Section A3.1. The other term can be obtained via the following derivation:

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

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

$$R^2_{\mathcal{S}} = 1 - \frac{\sum_{jt\in\mathcal{S}} \frac{N_{jt}}{N_{\mathcal{S}}}[(\hat{\text{V}}_{jt} - \text{V}^P_{jt})^2 - \hat{\varepsilon}^2_{jt}]}{\sum_{jt\in\mathcal{S}} \frac{N_{jt}}{N_{\mathcal{S}}}[(\hat{\text{V}}_{jt} - \sum_{jt\in\mathcal{S}} \frac{N_{jt}}{N_{\mathcal{S}}}\hat{\text{V}}_{jt})^2 - \hat{\varepsilon}^2_{jt}]}.$$

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 $\hat{\varepsilon}^2_{jt}$ is the squared standard error for $\hat{\text{V}}_{jt}$.

A3.4 The standard deviation of $\mathbf{V}^*_{jt}$ based on $\mathbf{V}^P_{jt}$

For the 2015-2019 admissions cohorts, we cannot estimate value added and instead only have machine learning forecasts, V^P_{jt}. We would like nonetheless to estimate the standard deviation of the true effects, V^*_{jt}, for these years. To do this, we assume that the true effects are equal to the forecasts plus independent forecast error:

$$\text{V}^*_{jt} = \text{V}^P_{jt} + \vartheta_{jt},$$

with $\vartheta_{jt} \perp \text{V}^P_{jt}$. We calculate the variance of V^*_{jt} by assuming that V^P_{jt} has an R-squared in predicting V^*_{jt} equal to that observed for the 2008-2014 cohorts (0.791, Table A26). Specifically, we use the following derivation:

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

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

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

where $R^2_{0814} = 0.791$.

A3.5 R-squared for beliefs about $\mathbf{V}^*_{jt}$, proxied by $\mathbf{V}^P_{jt}$

In Section 2, we are interested in assessing how well households' beliefs about value added reflect a track's true value added. However, for the experimental (2019) cohort, we observe only a track's forecasted value added, V_{jt}^P, not its true value added, V_{jt}^*. Let p_{ij}^{V} be the fitted value from a regression of V_{jt}^P on s_{ij}^{V}. We estimate 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 estimated using the approach described in Section A3.4. 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 estimated from the data. $\mathrm{Var}[\vartheta_{jt}]$ can be written as:

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

Finally, we assume $\mathrm{E}[p_{ij}^{\mathrm{V}} \cdot \vartheta_{jt}] = 0$; that is, households' scores are not correlated with the unforecastable component of track value added.[65] Thus, R-squared is:

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

The finite-sample formula is:

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

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.

65. 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.

A4 How households rank and score tracks

In this appendix, we ask two questions related to how households rank and score tracks. First, we assess whether they consider all the tracks in their towns, or instead focus on a limited subset by selectivity. Second, we examine whether they rank tracks from multiple curricular categories.

The fact that the Romanian high school assignment mechanism is incentive compatible means that it is weakly dominant for a household to rank 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 admitted to the track. In practice, however, households may find it costly to evaluate tracks. As a result, they may focus only on the tracks that they believe their child is 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 rather a subset of them, with the particular subset depending on the student's achievement. For instance, a household with a low-achieving child may not rank and/or score selective tracks that it is sure will be "out of reach".[66] Similarly, a household with a high-achieving child may not rank and/or score non-selective tracks.

If households systematically omit certain tracks, there could be issues for our analysis. First, if households skip out-of-reach tracks, then their preference rankings would not reflect their true preferences. Households would leave tracks unranked that they actually prefer to those they do rank.[67] Second, if households consider only a subset of tracks, then their quality scores may pertain to the distribution of tracks within that subset, rather than among the town as a whole.

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	0.42	0.41	0.45	0.35	0.32	0.38
Fraction of households ranking / scoring:						
No tracks	0.09	0.09	0.06	0.38	0.43	0.32
1-25 percent	0.31	0.33	0.28	0.23	0.22	0.24
26-50 percent	0.29	0.28	0.31	0.08	0.06	0.10
51-75 percent	0.10	0.09	0.12	0.04	0.04	0.04
$>$ 75 percent	0.21	0.21	0.23	0.27	0.26	0.30
Number of students	3,898	1,554	2,192	3,898	1,554	2,192

The table describes the share of tracks that a survey household ranks and/or scores in the baseline survey. A household is said to rank a track if it includes the track in its preference ranking. A household scores a track if it assigns scores for both value added on passing the baccalaureate exam ("pass") and peer quality ("peers"). "Mean share of tracks ranked/scored" is the average share of tracks that a household ranks or scores. The remaining rows display the fraction of households that rank or score 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.

To assess these issues, we first provide additional detail on the share of tracks that households rank and score. As noted in Section 0.4, households, on average, rank 42% of the tracks in their towns, and they score 35% on academic value added and 36% on peer quality (Table 3). Table A29 summarizes how these shares vary across households. The first column shows that most households rank a significant share of tracks; 60% rank over a quarter, and 21% rank over three quarters. Meanwhile, 9% rank no tracks. The fourth column displays the share of tracks that a household scores.[68] It shows that this distribution is more bimodal, with most households assigning scores

66. This is the issue of "skipping" discussed by Fack, Grenet, and He (2019) and Artemov, Che, and He (2020).

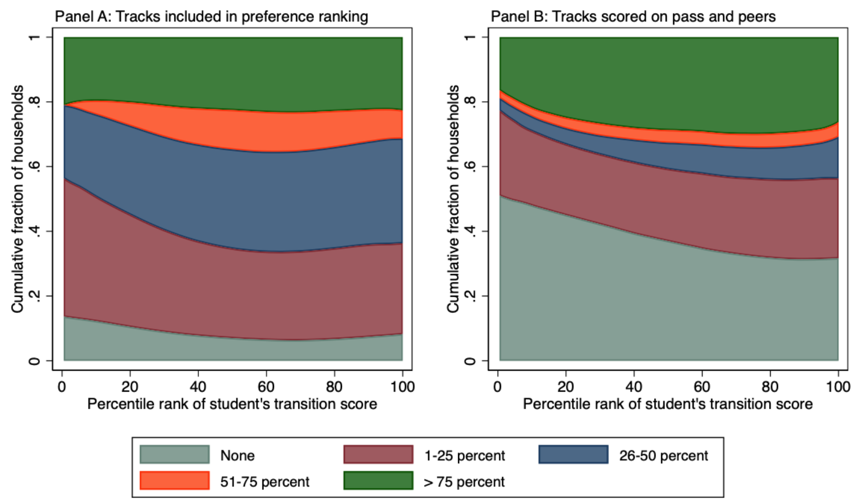
67. Omitting non-selective tracks is not an issue, as these are less preferred than the tracks that are ranked.

68. Here, we define a household as scoring a track if it assigns scores for both peer quality and value added on passing the baccalaureate exam.

to either a small or large share of the tracks in their towns. Specifically, 61% of households score a quarter of the tracks or fewer, with 38% scoring no tracks. On the other hand, 27% score over three quarters of the tracks.

Figure A9 (page 71) 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 achievement distribution.

Figure A9: The share of tracks that a household ranks and/or scores by student transition score



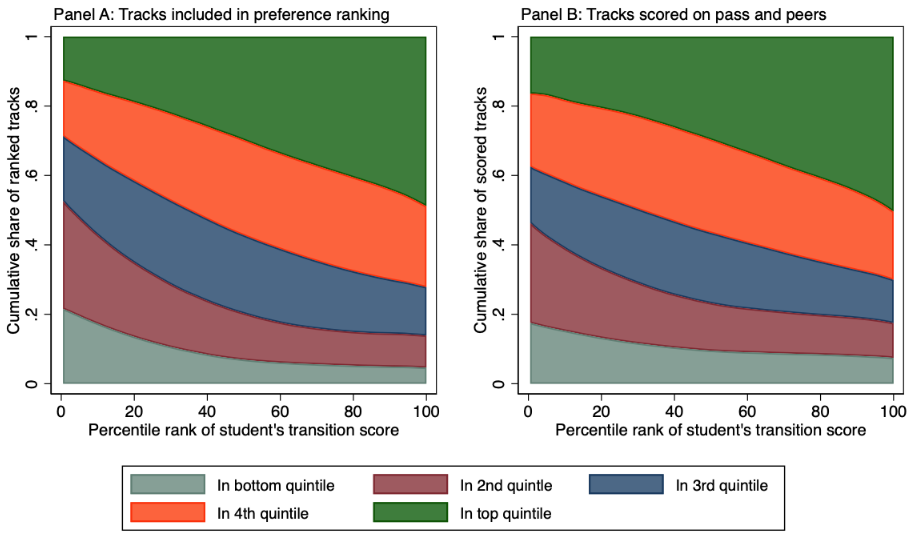
The figure shows how the share of tracks that a survey household ranks and/or scores varies with the student's transition score. Specifically, households are assigned to groups based on whether they ranked and/or scored none of the tracks in their towns, 1-25 percent of the tracks, 26-50 percent, 51-75 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.

Next, we inspect whether households with low- and high-achieving children differ in the selectivity of the tracks that they include in their preference rankings and sets of quality scores. We find that households include tracks from across the selectivity distribution. Figure A10 reveals the fraction of tracks that a household ranks and/or scores that come from each within-town quintile of selectivity.[69] The figure reveals a few notable facts. First, households with low-achieving children include tracks from each quintile at almost uniform rates. Second, households with high-achieving children are more likely to include selective tracks than non-selective ones. Among this group, about 40% of the tracks that a household ranks and/or scores fall into the top quintile. Nonetheless, these households still include significant fractions of non-selective tracks. About 20% of their ranked and/or scored tracks come from the two least-selective quintiles.

The evidence in this section thus counters the notion that households consider only a subset of tracks based on their child's achievement; instead, they rank and score tracks with a range of selectivities. Thus, the evidence broadly supports the assumptions that households' track

69. We define selectivity using a track's prior-year minimum transition score, MTS_{jt-1}. We use the prior-year (2018) value of this variable because it can be observed by households at the time of the information sessions. Prior-year MTS is published by the government just before these sessions—when the government announces the year's list of available tracks. In addition, households may be able to remember it from the previous allocation. As such, 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 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.

preference rankings reflect their true preferences and that their quality scores map to the full distribution of tracks within their towns. Nonetheless, in the main analysis we are careful to show that our results are not sensitive to these assumptions.

We next examine whether households rank tracks from multiple curricular categories—or if, instead, they focus on just one. The answer to this question reveals whether a student's choice set is best reflected by all its available options or by only those with its preferred curriculum.

Table A30: The number of curricular categories among a household's top choices

Among top:	**Curricular categories**		**Students**
	Rank ≥ 2	Mean	
2	0.36	1.36	3,227
3	0.68	1.73	2,783
4	0.82	1.97	2,365
5	0.93	2.26	1,868
6	0.98	2.39	1,452

The table provides summary statistics on the number of curricular categories that are included among a household's top choices in the baseline preference ranking. "Rank $\geq$ 2" is an indicator for whether the household ranks tracks from at least two categories. "Mean" is the mean number of categories from which a household ranks tracks. The sample in each row is restricted to households who ranked at least the listed number of tracks.

The results are presented in Table A30. The table shows the share of households who include tracks from multiple curricular categories among their top baseline choices. It also lists the mean number of categories that households rank. The results indicate that households consider tracks with differing curricula. For instance, among their top three choices, 68% of households include tracks from at least two categories. Among the top six choices, this value is 98%.

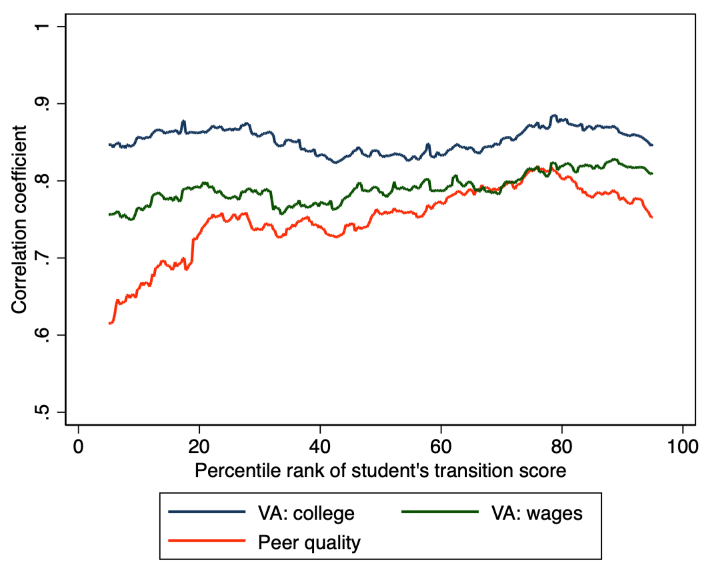
A5 Explaining households' beliefs about value added

This section considers two questions related to households' beliefs about value added. First, we investigate whether households believe that value added is multi-dimensionsal. For instance, do they think that some tracks have high value added in one dimension, while others have high value added in another dimension—or, instead, do they think the same tracks are good across the board? Second, we explore how households' beliefs about other track characteristics explain their beliefs about value added. In particular, do they believe value added is interchangeable with peer quality, or do they think it additionally depends on factors such as curriculum and teacher quality?

These questions have important implications for the paper's analysis. First, in the experiment, we provided information only on value added with respect to passing the baccalaureate exam. For students with a relatively even chance of passing, this outcome is of direct interest. However, it is less relevant for students with either high or low chances of passing. For these students, the information is of interest only to the extent that it illuminates tracks' effects on other outcomes, such as on wages or attending a high-quality college. If households believe that value added is correlated across dimensions, then they would interpret our information as a signal of tracks' value added on the outcomes they care about. If not, then they would find the information to be of little use. This would cause treatment effects to be smaller than if we had informed them about value added on the appropriate dimension.

Next, the second question has implications for the pathways through which our treatment effect operates. If households do not distinguish value added from peer quality (i.e., if they do not understand selection bias), then our intervention may teach households about the concept of value added. By contrast, if households already understand this distinction, then the impact of the intervention will operate mainly through revealing which tracks have high value added—although it may still serve to direct attention to value added vis-a-vis other track characteristics.

Figure A11: The correlation between households' scores for value added on passing the baccalaureate exam and their scores for other track attributes



The figure presents coefficients from correlations between households' baseline quality scores for value added on passing the baccalaureate exam and those for the listed track attributes. Each point in the figure is the coefficient from a different correlation. The correlations are calculated using the sample of students with transition scores within a 20-percentile-rank range. The coefficients are plotted against the median value of transition score percentile rank in the range.

The results for the first question are presented in Figure A11. The figure shows coefficients from correlations between households' baseline quality scores for value added on passing the baccalaureate exam and their scores for other value added dimensions. As a benchmark, it also includes correlations with scores for peer quality. Each point in the figure is the coefficient from a different correlation. The correlations are calculated using only students with transition scores within a 20-percentile-rank range. Thus, the figure reveals how the correlations vary in magnitude across the achievement distribution.

The figure indicates that households' believe value added is highly correlated across dimensions. Nonetheless, they also seem to see it as multi-dimensional. Scores for value added on passing the exam are most related to those for value added on college quality, with correlations that vary between 0.82 and 0.89 (blue line). For the full sample, the value is 0.86 (Table A6). Correlations with value added on wages (green line) are slightly lower, ranging from 0.75 to 0.83, with a full-sample value of 0.81. For neither set of correlations is there strong heterogeneity by student achievement. Next, the correlations with peer quality (orange line) are even lower than those with value added on wages. For peer quality, these vary between 0.62 and 0.82, with a full-sample value of 0.77. Further, for this track attribute there is more variation by achievement. Households with low-achieving children think there is a weaker relationship between value added and peer quality (correlation of 0.72) than do those with high-achieving children (correlation of 0.78).

The above results suggest that households likely interpreted our information as being informative about the value added dimensions they care about. However, treatment effects may be slightly smaller than if we had provided information on value added for additional outcomes.

Table A31: Regressions of scores for "VA: pass the bacc." on scores for other track attributes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Teacher quality	0.850***			0.465***	0.460***	0.438***	0.471***
	(0.012)			(0.033)	(0.031)	(0.046)	(0.036)
Curriculum		0.719***		0.281***	0.260***	0.225***	0.279***
		(0.017)		(0.024)	(0.023)	(0.025)	(0.033)
Peer quality			0.754***	0.194***	0.179***	0.185***	0.178***
			(0.016)	(0.023)	(0.025)	(0.042)	(0.026)
Location					0.037***	0.064***	0.024
					(0.013)	(0.023)	(0.015)
Siblings and friends					0.028***	0.046**	0.015
					(0.011)	(0.018)	(0.011)
All students	x	x	x	x	x		
Low-achieving						x	
High-achieving							x
R-sq.	0.65	0.58	0.58	0.73	0.73	0.68	0.76
Clusters	188	189	188	188	186	163	173
Students	2,382	2,390	2,370	2,333	1,957	706	1,251
Student-tracks	17,455	17,439	17,460	17,175	14,751	5,348	9,403

The table presents results for regressions of households' baseline quality scores for "VA: pass the bacc." on their scores for the listed track attributes. The sample for the 6^{th} (7^{th}) column is students with transition scores in the bottom- (top-) half of the national distribution. Standard errors are clustered by middle school.

Next, Table A31 reveals how households' beliefs about value added are explained by their beliefs about other track attributes. It shows results for regressions of scores for value added on passing the baccalaureate exam on scores for teacher quality, curriculum, and peer quality. In a few specifications, it also controls for scores for a track's location and for whether a student's siblings and friends attend the track.

The table reveals that households' value added scores are most closely related to their scores for teacher quality. However, they are also related to scores for curriculum and peer quality. A one unit increase in a score for teacher quality is associated with a 0.85 unit increase in the score for value added, with an R-squared of 0.65 (Column 1). Coefficients on scores for curriculum (Column 2) and peer quality (Column 3) are respectively 0.72 and 0.75, with R-squared in both cases equal to 0.58. Column 4 presents a horse race, showing that a one unit increase in a score for teacher quality is associated with a 0.47 unit increase in the score for value added, while scores for curriculum and peer quality are associated with increases of only 0.28 and 0.19. Next, Column 5 adds scores for location and siblings and friends. It indicates that these latter variables do not contribute additional explanatory power. Finally, Columns 6 and 7 reveal that there is little heterogeneity in the results by student achievement.

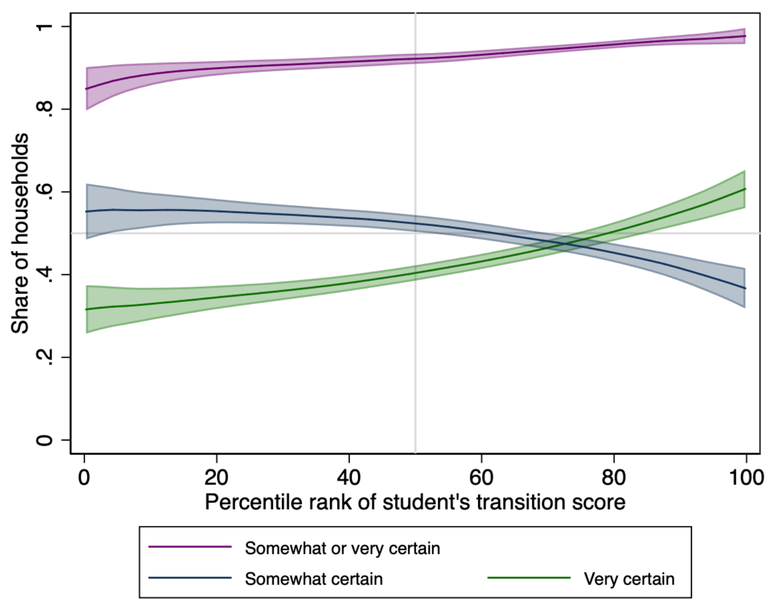
Thus, the evidence suggests that households did understand the concept of value added in advance of the experiment. Namely, they conceived value added as being related to the quality of a track's teachers, while also depending on the track's curriculum and peers.

A6 How certain were households at the time of the baseline?

In this section, we examine the timing of the baseline survey. Specifically, we investigate how carefully households had thought about their options at this point in time. Recall that the baseline survey is when we collected baseline preference rankings and quality scores and when we provided treated households with information on value added. It occurred about a month before households were required to submit their official rankings. Further, it occurred in school-organized information sessions that are used to explain the admissions process.

The question of timing is important for two reasons. First, it has implications for the magnitude of the treatment effects. If households had already settled on their track choices at the time of our intervention, then they may have been resistant to incorporating new information. This would cause the treatment effects to be smaller than if we had intervened earlier on. Second, the timing of the survey also has implications for the relevance of the baseline preference rankings and quality scores. Namely, if households hadn't yet begun to consider their options, then these would likely have little in common with households' beliefs and choices when they submit their official lists.

Figure A12: 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.

We explore the question of timing in two ways. First, we use self-reports from the baseline survey in which households were asked if they were already certain of their preference rankings. As mentioned in Section 0.4, on this measure, households appear to have differed in their degree of certainty. 39% report already being very certain, 46% report being somewhat certain, and 15% were uncertain (Table 3). Figure A12 reveals how these shares vary by student achievement. The figure plots the fraction of households who were somewhat certain, very certain, or either somewhat or very certain against the national percentile rank of the student's transition score. Table A32 presents corresponding summary statistics. The results indicate that households with low-achieving children were slightly less certain than those with high-achieving children. However, both groups exhibited a range of certainty. For households with children in the bottom half of the national distribution, 33% were very certain, 50% were somewhat certain, and 17% were uncertain. Meanwhile, for households with children in the top half of the national distribution, 45% were very certain, with 43% being somewhat certain and 12% being uncertain.

Table A32: Summary statistics on households' certainty in their baseline preference rankings

	All students	**Low-achieving**	**High-achieving**
Share who reported being:			
Very certain	0.39	0.33	0.45
Somewhat certain	0.46	0.50	0.43
Uncertain	0.15	0.17	0.12
Students	3,898	1,554	2,192

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

The second way we assess timing is by comparing baseline preference rankings with official track assignments for control households. Control households were not provided with information on value added. Thus, their behavior reveals the dynamics of decision-making in the absence of the experiment. If households were already settled on their choices at the time of the baseline survey, then track assignments for control households should match those implied by their baseline preference rankings. By contrast, if households hadn't yet thought through their options, then they would be likely to change their choices before submitting their official lists. As a consequence, their assignments would differ from those implied by their baseline preference rankings.

Table A33: The share of control households whose baseline preference rankings match their track assignments

	N	**Share**
All students	1,095	0.74
Very certain	481	0.78
Somewhat certain	544	0.72
Uncertain	70	0.64

The table reveals the fraction of households in the control group who were assigned to the feasible track that they ranked most highly in the baseline survey.

Table A33 presents the results. It displays the fraction of control households whose track assignment matches the track to which they would have been assigned based on their baseline preference ranking. Overall, it shows that 74% of control households fall into this group. Among those who reported being "very certain" in the baseline survey, the fraction is 78%. Meanwhile, for those who were somewhat certain or uncertain, the shares are 72% and 64%, respectively.

In total, the evidence suggests that we intervened at a reasonable time. Most households had already begun considering their options, but many were not yet fully settled on their choices. In addition, a quarter of control households meaningfully changed their choices after the baseline survey, while three quarters did not.

A7 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 as the 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.

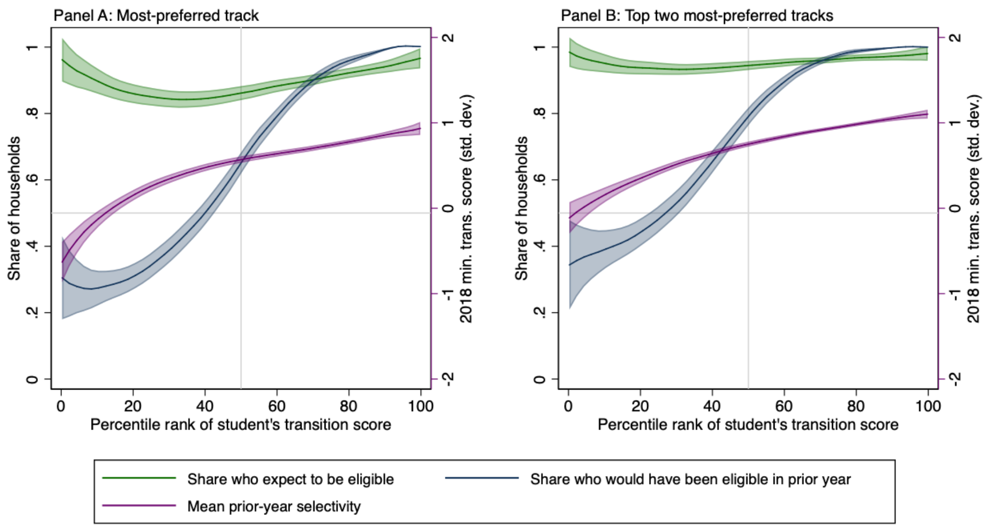
One complication for the matching algorithm was that some towns had an odd number of one-class schools. In these towns, we stopped the matching algorithm when there were three remaining one-class schools. We calculated the Mahalanobis distance of the covariates for each school in the triple to the average 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 its implementation. When there was an issue with one school in a matched pair, we dropped the entire pair. Thus, the final experimental sample included 170 middle schools in 45 towns, of which 86 middle schools were in the treatment group and 84 were in the control group.

A8 Households' top track choices

In this section, we examine households' behavior with respect to their 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 for.[70] We also assess the accuracy of households' expectations.

Figure A13: Summary statistics on a household's most-preferred tracks



The figure provides information on households' most-preferred tracks in the baseline survey. Panel A pertains to a household's top-ranked track; Panel B relates to its two highest-ranked tracks. The green lines display the shares of households that expect their child to be eligible for these tracks. The blue lines exhibit the shares 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 shows the mean 2018 selectivity of a household's most-preferred tracks (in standard deviation units).

Figure A13 provides the results. Panel A relates to a household's highest-ranked track; Panel B is for the 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, 84-94% of households believe their child will be admitted to their most-preferred track and 93-97% think their child will be admitted to at least one of their two most-preferred tracks.[71] Lending credence to these expectations, households with lower-performing children choose less selective tracks than do those with higher-performing children. However, households tend to be overly optimistic about track feasibility. For students with transition scores in the bottom half of the distribution, only 40% would be eligible for their top-ranked track based on the track's prior-year minimum transition score. Similarly, only 54% would be eligible for one of their two top 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.

Thus, the results in this section reveal that most households expect their child to attend one of their top choices. However, many households are over-optimistic in this regard.

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

71. Over the full sample, these values are 89.2% and 95.4%, respectively.

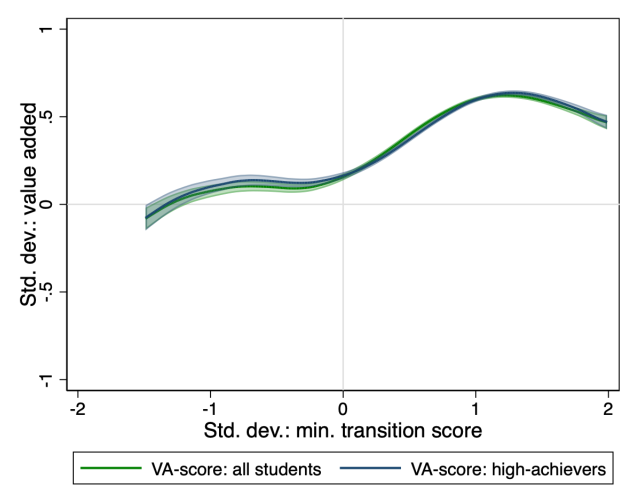
A9 Additional results for Section 1

In this appendix, we provide additional results for Section 1. First, we show additional robustness. Second, we use an additional approach to characterize households' track choices.

A9.1 Additional robustness

One concern with our value added measures is that they may be sensitive to effects on test taking. It is possible that there is a large reward—in terms of measured value added—from inducing students to switch their behavior on this dimension. If so, then a track's measured value added may depend on the share of the track's students who are marginal in terms of test taking. Notably, measured value added may be capped for tracks in which large shares of students take the exam. This may explain why the relationship between value added and selectivity flattens out among the most selective tracks (Figure 1).

Figure A14: The relationship between value added and selectivity: further robustness



The figure shows the relationship between value added and selectivity for two versions of value added on exam score. The first version, "VA-score: all students", is calculated using all students. The second version, "VA-score: high-achievers", is calculated using only high-achieving students, defined as those with transition scores in the top half of the within-year distribution. The sample is restricted to track-years with at least 5 high-achieving students. Variables are standardized by year using the mean and standard deviation among all tracks.

We do not believe this concern is a major issue, as we have shown that the pattern of results holds across multiple value added measures, some of which are more sensitive to test-taking than others. Nonetheless, we probe the concern using one further strategy. This strategy involves calculating tracks' value added for high-achieving students, defined as those with transition scores in the top half of the within-year distribution. 91% of these students take the baccalaureate exam; thus, effects on test taking can exert only a limited impact on this value added measure.

Specifically, among all track-years with at least five high-achieving students, we calculate value added on exam score using just the high-achieving students.[72] We then compare this measure with our main measure of value added on exam score, which is calculated using all available students. We find that the two measures are highly similar: their correlation is 0.95; in addition, they have the same relationship with selectivity, as seen in Figure A14. Importantly, for both measures, the relationship is flat or even negative among the most selective tracks. Thus, this finding is not due to differences in the share of students who are marginal.

72. We focus on exam score—rather than passing the exam—because a large fraction of high-achievers pass.

A9.2 Characterizing track choices using a discrete choice model

We next use an additional approach to characterize households' track choices. This approach involves explaining track choices using measured values of track characteristics, including value added, V_{jt}, and selectivity, MTS_{jt}. The approach is similar to the discrete choice analysis in Abdulkadiroglu et al. (2020) and Beuermann et al. (2019). It is useful because it allows us to precisely compare our results with those of the previous papers.

As discussed in Section 0, the administrative data reveals the track a student attends, j_i^*, as well as the set of tracks for which the student is eligible, $\mathcal{J}_i^e$. Following Fack, Grenet, and He (2019), we assume that the track the student attends is the household's most preferred among the feasible set. That is, expected utility from track j_i^*, $U_{ij_i^*}$, is at least as large as that from all the household's other options. This amounts to assuming that households—when submitting their preference rankings—did not believe that any feasible tracks were out of reach.

We write a household's expected utility for track j as a linear function of characteristics of the track, X_{jt}, in the household's cohort t:

$$U_{ij} = \omega' X_{jt} + \omega_{ij}. \tag{10}$$

We assume ω_{ij} is independent and has a Type I Extreme Value distribution. We then fit equation (10) to students' track assignments using a multinomial logit.

Table A34: How track utilities relate to measured values of track characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
Value added,	0.608***		-0.007	0.123***	0.035***	0.219***
V_{jt} (s.d.)	(0.008)		(0.006)	(0.007)	(0.012)	(0.010)
Selectivity,		2.05***	2.05***	2.03***	1.70***	2.26***
MTS_{jt} (s.d.)		(0.021)	(0.022)	(0.023)	(0.021)	(0.038)
Humanities				-0.473***	-0.416***	-0.390***
				(0.016)	(0.030)	(0.022)
Math or science				0.004	-0.415***	0.244***
				(0.014)	(0.020)	(0.022)
All students	x	x	x	x		
Low-achieving					x	
High-achieving						x
R-sq.	0.05	0.20	0.20	0.20	0.11	0.28
Clusters	5,969	5,969	5,969	5,969	5,906	5,728
Students	2,162,736	2,162,736	2,162,736	2,162,736	1,061,736	1,101,000
Student-tracks	47,298,148	47,298,148	47,298,148	47,298,148	13,429,626	33,868,524

The table presents results from equation (10). These rely on a multinomial logit to explain the track a student attends, j_i^*, among the options in its feasible choice set, $\mathcal{J}_i^e$. The sample is the administrative data for the 2004-2017 and 2019 cohorts. "Humanities" and "Math or science" are indicators for a track's curricular category (the omitted category is technical tracks). Colums 5 and 6 are for students with transition scores in the bottom- (top-) half of the within-year distribution. Standard errors are clustered by town-year.

Table A34 presents the results. The first column is for a specification that includes only value added, V_{jt}, while the second includes only selectivity, MTS_{jt}. The third and fourth columns include both variables, with the fourth also adding controls for a track's curriculum. Finally, the last two columns are the same as Column 4, but for either low- or high-achieving students.

The results suggest that both value added and selectivity explain utility; however, selectivity's explanatory power is much stronger.[73] A one standard deviation increase in value added (selectivity) is associated with a 0.61 (2.05) unit increase in utility. Both effects are significant at the

73. Note that the relationship between track utilities and selectivity is not mechanical. This is because tracks

1% confidence level. When the two variables are included together, in Column 3, the coefficient on value added falls to zero, while that on selectivity remains large. When we add controls for curricular focus, in Column 4, the coefficient on value added increases slightly, but is still only 6% as big as that on selectivity. Finally, the results in Columns 5 and 6 suggest that value added has some explanatory power, conditional on selectivity, for high-achieving students, but none for low-achieving ones.

These results are broadly similar to those of Abdulkadiroglu et al. (2020) and Beuermann et al. (2019). As in Abdulkadiroglu et al. (2020), we find that, over the full sample, value added doesn't explain utility after conditioning on a measure of peer quality. As in Beuermann et al. (2019), we find that it does—to some extent—for high-achieving students.

differ in size. In particular, it is possible for one track to be both more popular and less selective than another if it is larger. However, for two tracks that are the same size, the more selective one is necessarily more popular.

A10 Additional results for Section 2

In this appendix, we show that the results in Section 2 are highly robust.

Table A35: Explaining within-town quintiles of track attributes using households' quality scores: households who scored all tracks

	All students		Low-achieving		High-achieving	
	quint(V_{jt})	quint(MTS_{jt-1})	quint(V_{jt})	quint(MTS_{jt-1})	quint(V_{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.

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. In Table A35, we replicate Table 8 but restrict the sample to the 21% of households with no missing scores. Results are similar.

Table A36: 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_{jt})	quint(MTS_{jt-1})	quint(V_{jt})	quint(MTS_{jt-1})	quint(V_{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.

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. In Table A36 we replicate Table 8 but restrict the sample to tracks that a student would have been eligible to attend in the prior year. Results are again similar.

Third, it may be that households had not yet studied their options when the baseline survey took place. To investigate this, we replicate Table 8 while restricting the sample to the 39% of households that reported already being "very certain" of their preference rankings during the baseline survey. The results, in Table A37, are still similar.

Finally, the R-squared statistics that we provide in Table 8 may be misleading. These are R-squared in terms of explaining value added forecasts, $V_{jt} = V^P_{jt}$. However, we are ultimately interested in R-squared in terms of explaining true value added, V^*_{jt}. To investigate this distinction,

Table A37: 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_{jt})	quint(MTS_{jt-1})	quint(V_{jt})	quint(MTS_{jt-1})	quint(V_{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 A38: Explaining track attributes (in std. dev.) using households' quality scores

	All students		**Low-achieving**		**High-achieving**	
	V_{jt} (s.d.)	MTS_{jt-1} (s.d.)	V_{jt} (s.d.)	MTS_{jt-1} (s.d.)	V_{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 value added, V_{jt}, and prior-year selectivity, MTS_{jt-1}, on households' quality scores. The regressions are similar to those in Table 8. 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 forecast for value added, $V_{jt} = V^P_{jt}$, not the true value, V^*_{jt}. Appendix A3.3 explaings how we calculate "R-sq.: V^*_{jt}". See Table 8 for additional details.

we run regressions that use 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^P_{jt} for forecast error (Appendix A3.5 describes the procedure). Table A38 contains the results. It shows that R-squared for true value added, V^*_{jt}, is similar to, but slightly lower than, that for forecasted value added, V^P_{jt}.

A11 Testing for informational spillovers

This appendix investigates whether the experiment suffered from informational spillovers. In particular, it is 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, 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 with 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 traits have no obvious relationship with the magnitude of treatment effects, except via their effect on spillovers.

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 (1)) separately for these two groups.

Table A39: 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 (1) 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 9 for additional details on the regressions.

The results are in Table A39. The first three columns refer to the full sample of students, and the remaining to the sub-samples with 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 provide no evidence of spillovers. Instead, treatment effects are shown to be similar in magnitude for each group of towns.

A12 Estimating preferences using experimental variation

In this appendix, we provide additional results for Section 4. In Section 4, we fit the preference model, equation (4), using baseline quality scores and baseline preference rankings. We now fit the model using endline quality scores and endline preference rankings. In addition, we make use of experimental variation in beliefs about value added.

We first discuss why our main analysis relies on baseline data. We then explain our approach for using endline data and experimental variation. Finally, we present the results.

A12.1 Issues with the endline data and the experimental variation

There are two sets of reasons why we use baseline data in the main text. First, there are issues with the endline quality scores. Second, there are problems with the exclusion restriction that is needed to exploit the experimental variation.

There are a few issues with the endline quality scores. First, we have endline scores only for value added on passing the baccalaureate exam, not for other types of value added or for other track characteristics. Thus, to use endline data, we need to make assumptions about how households updated their beliefs about these other quality dimensions. Second, there is substantial missing data for the endline value added scores. To avoid restricting households' choice sets, we must impute these missing values. This requires us to make assumptions about how treated and control households updated their value added beliefs.[74] Third, there may be measurement error in the endline value added scores. As we discussed in Section 3.2, the follow-up survey was conducted a few weeks after households submitted their official track preference rankings. By this time, households may have forgotten some of what they knew when they settled on their rankings.[75] Finally, the endline data has a considerably smaller sample size than the baseline, due to non-response in the follow-up survey.

Next, the exclusion restriction required for identification may not be valid. Our goal is to use experimental variation in households' endline value added scores, $s^V_{ij,\text{fs}}$, to identify households' preference coefficient for value added, β_V. This requires us to assume that the experiment affected treated households' utilities from tracks only via its effects on their value added scores—not via any other channel. In reality, the experiment may have caused treated households to update their beliefs about tracks on multiple quality dimensions. If these changes in beliefs are correlated with the treatment effect on value added scores, then the approach of using experimental variation may overstate the preference for value added.[76] Similarly, the experiment may have directly impacted preferences. For instance, it may have caused treated households to care more about value added. If so, then the approach of using experimental variation would identify a special form of β_V. It would recover the value of β_V in a world in which policymakers signal the importance of value added. By contrast, it would not recover the value in the current institutional context, where policymakers are neutral about track characteristics.

74. Note that we did not impute missing scores in Section 3.2, when we calculated treatment effects on beliefs. There, we examined effects on the accuracy of the scores that households provided. Nonetheless, for estimating preferences and for running our simulation, it is important to have scores for all tracks in a choice set.

75. As we mentioned in Section 3.3, we do not believe that there is significant measurement error in the endline preference rankings. This is because we asked households to find their official submissions and read them to us.

76. We could model these other changes if we had endline scores for all quality dimensions; however, we do not. A related issue is that the experiment may have influenced the precision of value added beliefs in ways not captured by effects on the quality scores. If utility also depends on precision in ways not captured by the quality scores, then we may again overstate the preference for value added. That said, if this is the case, then using the baseline data likely understates the preference for value added.

Despite these various issues, we find that using endline data and experimental variation generates similar results as our main strategy, which relies on baseline data.

A12.2 Detailing our approach

Our approach for exploiting experimental variation is to augment the preference model with a control function. Control functions are commonly used to deal with endogeneity in discrete choice models, as discussed in Petrin and Train (2010) and Wooldridge (2014, 2015).

Our strategy proceeds in three steps. First, we impute missing endline value added scores, $s^V_{ij,\mathrm{fs}}$. Second, we run a first-stage regression of these scores on (i) the other variables that we want to include in the preference model, (ii) measured value added, (iii) the treatment indicator, and (iv) the interaction of measured value added and the treatment indicator. Third, we fit the preference model. As covariates, we include (i) and (ii) but not (iii) and (iv). Further, we add a flexible function of the first-stage residuals. This function is the "control function". It controls for the unexplained variation in endline value added scores, and it means that the only remaining variation in these scores is due to the treatment-induced increase in their association with measured value added. Thus, by adding the control function, the preference coefficient for value added scores, β_{V}, is identified using experimental variation.[77]

To probe robustness, we provide results under different assumptions about how to impute missing endline value added scores. For control households, we always replace missing values with baseline scores. For treated households, we use five alternative strategies. In the "Accurate" specification, we fill in accurate scores (i.e., the within-town quintile of measured value added). In the "Two-thirds accurate" and "Half accurate" specifications, we use averages of baseline scores and accurate scores that respectively place two-thirds and one-half weight on the accurate scores. In "Accurate other than top 2" and "Accurate other than top 4", we use baseline scores for either the two or four most-preferred tracks at baseline and accurate scores for the remainder.[78]

We also must construct variables to reflect households' beliefs about quality dimensions other than value added. To do this, we make the same assumptions as in Section 4.2. For "Location", "Siblings and friends", and "Curriculum", we use baseline scores. For "Peer quality", we use the within-town quintile of a track's selectivity.

A12.3 Results

We present results both for preference estimates and for the simulated impact of making households have accurate beliefs about value added.

The preference estimates are presented in Table A40. The columns of the table reflect the five different assumptions about how to impute missing endline value added scores for treated households. Due to sample size considerations, we fit the models using all students, rather than separately by achievement level. The variables in the models correspond to those in the "With measured attributes" specification in Section 4.2.[79] In addition, the models control for a cubic

77. Another commonly used approach for dealing with endogeneity in discrete choice models is that derived in Berry, Levinsohn, and Pakes (1995, 2004). The BLP approach can deal with unobservables that vary by track, but not by track and household. In our setting, if there is an unobservable that causes bias, it likely varies over households, given that households differ in their beliefs about track quality.

78. The logic for replacing missing values with accurate scores is two-fold. First, when a household does not score a track, it is often because the household is unfamiliar with the track. Second, in our experiment, treated households were allowed to keep the informational flyers. Thus, they may have referenced these flyers when deciding their track preferences and used them to shape their beliefs about unfamiliar tracks.

79. We use this specification for a few reasons. First, we must control for measured value added in the first-stage regression. Thus, we cannot use the "Just quality scores" specification. Second, we do not have endline scores for

function of first-stage residuals; we find that results are similar using alternative functional forms.

The preference estimates are similar to those calculated using baseline data (as shown in, e.g., Tables 15 and A24). The only differences are that the coefficient for location is smaller and the coefficient for value added scores is slightly larger. The small increase in the value added preference could reflect the violations of the exclusion restriction discussed in Section A12.1.

Table A41 provides the results of the simulation. It shows that these are quite close to the values based on baseline data, displayed in Tables 16 and A25. In particular, Table A41 reveals that correcting households' value added scores is predicted to, on average, cause low-achieving (high-achieving) students to attend tracks with between 0.10 and 0.24 (0.08 and 0.23) s.d. worth of additional value added. Table A25 shows that the corresponding ranges based on baseline data are 0.08 to 0.23 (0.10 to 0.23).

In sum, the results in this appendix show that the findings in Section 4 are robust to estimating preferences using endline data and experimental variation in beliefs about value added.

the alternative value added dimensions. As such, it would be difficult to use the "Update on all VA dimensions" specification. Third, the control function should mitigate measurement error issues. Consequently, it would be inappropriate to use the "Adjust for measurement error" specification.

Table A40: Households' preferences for track attributes: results calculated using experimental variation in value added scores

	(1)	(2)	(3)	(4)	(5)
Households' quality scores:					
Location	0.009	0.003	0.002	0.004	-0.008
	(0.061)	(0.065)	(0.068)	(0.062)	(0.061)
Siblings and friends	0.352***	0.356***	0.360***	0.347***	0.346***
	(0.066)	(0.068)	(0.069)	(0.067)	(0.068)
Peer quality	0.118*	0.119*	0.117*	0.112*	0.109*
	(0.066)	(0.066)	(0.065)	(0.066)	(0.065)
Curriculum	1.08***	1.08***	1.08***	1.06***	1.05***
	(0.075)	(0.083)	(0.095)	(0.074)	(0.076)
VA: pass the bacc.	0.391***	0.490**	0.554**	0.454***	0.532***
	(0.130)	(0.193)	(0.255)	(0.132)	(0.142)
Measured track characteristics:					
Value added, V_{jt}	0.173*	0.142	0.118	0.163*	0.146
(s.d.)	(0.092)	(0.094)	(0.096)	(0.091)	(0.090)
Selectivity, MTS_{jt}	0.339*	0.324*	0.314*	0.322*	0.300*
(s.d.)	(0.177)	(0.175)	(0.173)	(0.177)	(0.172)
Humanities	-0.317*	-0.276	-0.248	-0.346**	-0.353**
	(0.172)	(0.179)	(0.183)	(0.171)	(0.171)
Math or science	0.175	0.183	0.192	0.162	0.166
	(0.165)	(0.170)	(0.175)	(0.164)	(0.164)
Control function:					
Residuals	0.668***	0.740***	0.710***	0.539***	0.426***
	(0.150)	(0.214)	(0.275)	(0.149)	(0.149)
Squared residuals	0.451***	0.625***	0.688***	0.462***	0.448***
	(0.040)	(0.054)	(0.056)	(0.038)	(0.036)
Cubed residuals	-0.096***	-0.145***	-0.148***	-0.088***	-0.074***
	(0.017)	(0.028)	(0.032)	(0.018)	(0.018)
R-sq.	0.37	0.38	0.38	0.36	0.36
Clusters	76	76	76	76	76
Students	1,533	1,533	1,533	1,533	1,533
Student-tracks	20,029	20,029	20,029	20,029	20,029

The table presents results from versions of the preference model, equation (4), that are calculated using experimental variation in value added scores. The results are from rank-ordered logits that are fit using endline preference rankings and endline value added scores. The rank-ordered logits define the choice set as all tracks in a student's town and are estimated using a household's two top choices. "Residuals" are the residuals from a first-stage regression of endline value added scores on the other covariates included in the preference model, the treatment indicator, T_i, and the interaction of the treatment indicator and measured value added, $T_i \cdot \text{sd}(V_{jt})$. "Squared" and "Cubed" residuals are the square and cubic of these residuals. The columns provide results under different assumptions about how to impute missing endline value added scores for households in the treatment group. Column 1 is the "Accurate" specification. The specifications in the remaining columns are, respectively, "Two-thirds accurate", "Half accurate", "Accurate other than top 2", and "Accurate other than top 4". See Section A12.2 for details on these specifications. For households in the control group, we impute missing endline value added scores using baseline value added scores. See Section A12.2 for details on how we constructed the quality score variables for the other quality dimensions. The sample is students in the follow-up survey. Standard errors are clustered by the middle school treatment-control pairs within which we conducted the randomization.

Table A41: The effect of accurate beliefs on the value added of students' tracks: results based on experimental preference estimates

	Change in value added: $V_{i,\text{AS}} - V_{i,\text{IS}}$		
	All students	Low-achieving	High-achieving
Panel A: Accurate			
Top 1	0.162	0.166	0.160
Top 2	0.166	0.169	0.164
Top 3	0.167	0.167	0.166
Top 4	0.156	0.156	0.156
Plausible: Top 2	0.151	0.159	0.147
Feasible: Top 2	0.092	0.103	0.086
Panel B: Two-thirds accurate			
Top 1	0.199	0.205	0.195
Top 2	0.200	0.205	0.198
Top 3	0.196	0.198	0.196
Top 4	0.175	0.175	0.175
Plausible: Top 2	0.178	0.188	0.173
Feasible: Top 2	0.095	0.106	0.089
Panel C: Half accurate			
Top 1	0.231	0.239	0.226
Top 2	0.224	0.229	0.220
Top 3	0.215	0.217	0.215
Top 4	0.184	0.184	0.184
Plausible: Top 2	0.194	0.204	0.188
Feasible: Top 2	0.088	0.098	0.082
Panel D: Accurate other than top 2			
Top 1	0.192	0.197	0.189
Top 2	0.195	0.200	0.193
Top 3	0.194	0.196	0.194
Top 4	0.181	0.181	0.180
Plausible: Top 2	0.183	0.193	0.177
Feasible: Top 2	0.115	0.129	0.107
Panel E: Accurate other than top 4			
Top 1	0.221	0.227	0.217
Top 2	0.231	0.238	0.227
Top 3	0.232	0.235	0.230
Top 4	0.217	0.218	0.216
Plausible: Top 2	0.218	0.231	0.210
Feasible: Top 2	0.145	0.163	0.134

The table summarizes the difference between $V_{i,\text{AS}}$ and $V_{i,\text{IS}}$ for simulations in which the preference model is calculated using experimental variation in value added scores. The columns provide the means of this difference for the listed groups of students. The panels present results for versions of the preference model that are calculated under five different assumptions about how to impute missing endline value added scores for households in the treatment group. See Section A12.2 for details on these assumptions. The rows within each panel present results for versions of the preference model that are estimated using different numbers of choices and different choice sets; see the notes to Table A25 for more details. The sample is students in the follow-up survey.