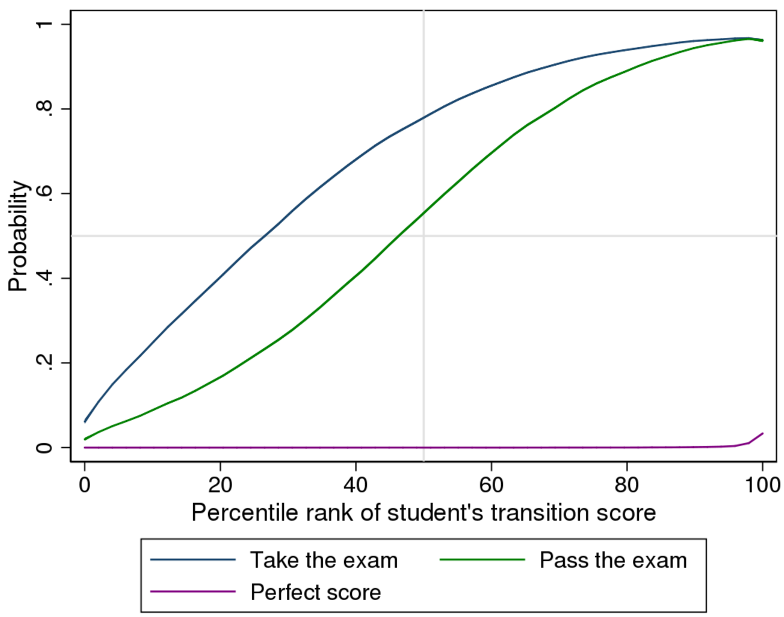
Online appendix for:

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

Additional figures

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

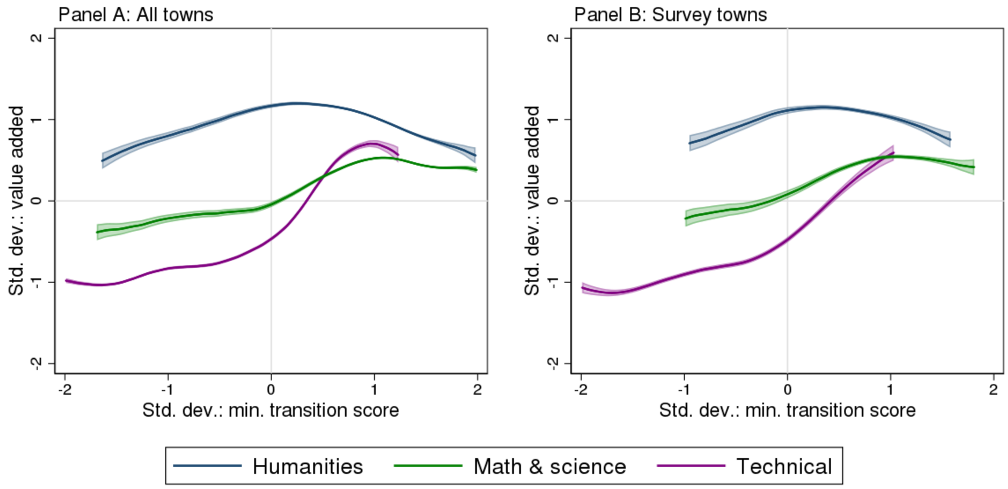
A team of economists at New York University has analyzed data in your hometown, Sebes Alba. They have calculated which tracks most effectively improve students' chances of passing the baccalaureate exam relative to their 9th grade starting points.

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.

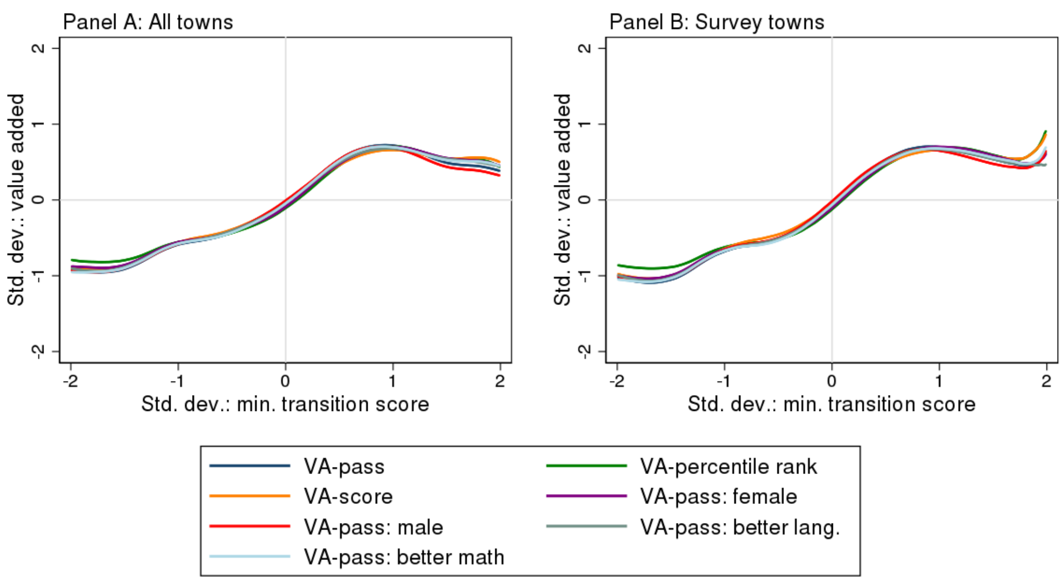
The figure displays the second page of the information sheet provided at the conclusion of the baseline survey. This information was provided only to treated households.

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



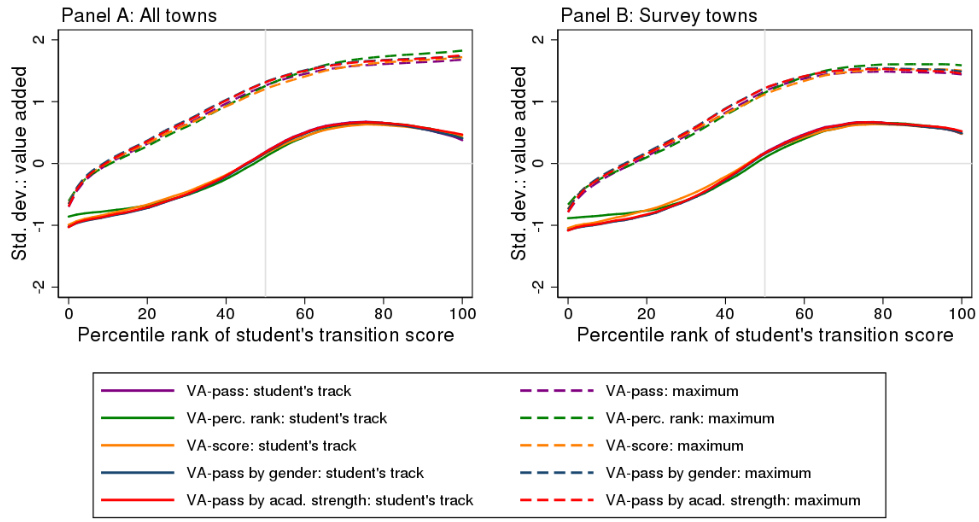
The figure presents the relationship between value added and selectivity for subsets of tracks by a track's curricular focus. Specifically, it replicates the local linear regressions in Figure 1 separately for tracks with curricular focuses in humanities, math and science, or technical subjects. See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

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



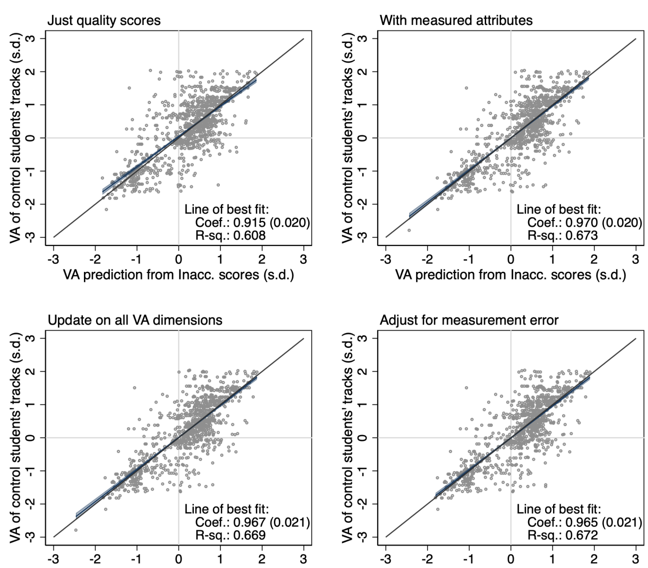
The figure replicates Figure 1 for alternative value added measures. Specifically, it presents local linear regressions of standardized values of various value added measures on standardized values of minimum transition score, MTS_{jt}. The value added measures are: (i) "VA-pass": a track-year effect 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 I.C 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 I.C 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 V.B 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 14 discuss the sample.

Additional tables

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 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 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 I.C 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	0.893	16	504	15	476	132	6	2
Alba	Sebes	0.844	10	290	10	297	35	3	0
Arges	Campulung	0.793	13	423	11	420	67	4	0
Bacau	Moinesti	0.823	9	303	9	280	87	3	2
Bacau	Onesti	0.807	16	650	16	637	157	6	2
Bihor	Beius	0.601	11	307	10	322	72	2	2
Bistrita Nasaud	Bistrita	0.822	28	925	23	782	148	7	2
Brasov	Fagaras	0.896	10	323	9	273	117	3	2
Buzau	Ramnicu Sarat	0.736	12	476	13	445	113	4	2
Calarasi	Calarasi	0.911	24	666	20	709	161	8	2
Caras Severin	Resita	0.509	20	473	18	425	103	7	1
Cluj	Dej	0.864	10	300	10	299	80	4	1
Cluj	Gherla	0.637	10	261	10	265	37	2	0
Cluj	Turda	0.788	12	282	11	281	71	5	0
Constanta	Mangalia	0.617	10	336	9	252	145	5	2
Constanta	Medgidia	0.673	10	308	9	280	27	1	0
Covasna	Sfantul Gheorghe	0.849	20	396	20	437	43	2	1
Covasna	Tirgu Secuiesc	0.604	9	219	9	233	43	3	0
Dolj	Calafat	0.784	7	183	6	168	37	2	0
Galati	Tecuci	0.861	18	753	16	728	79	5	0
Giurgiu	Giurgiu	0.875	15	591	14	602	148	9	2
Gorj	Motru	0.617	11	362	9	308	53	3	0
Harghita	Gheorgheni	0.903	11	280	12	263	22	2	0
Harghita	Miercurea Ciuc	0.847	22	602	21	589	48	4	1
Harghita	Odorheiu Secuiesc	0.861	15	392	15	364	39	3	1
Harghita	Toplita	0.747	7	170	8	172	22	2	0
Hunedoara	Deva	0.922	19	369	19	353	102	5	1
Hunedoara	Hunedoara	0.744	11	364	10	308	91	6	0
Hunedoara	Petrosani	0.742	9	299	8	224	101	4	2
Ialomita	Slobozia	0.936	19	636	16	644	91	4	2
Ialomita	Urziceni	0.887	11	316	7	280	59	3	0
Iasi	Harlau	0.861	8	222	7	224	34	2	0
Iasi	Pascani	0.842	17	688	16	644	109	4	2
Iasi	Targu Frumos	0.792	7	222	6	196	49	3	0
Maramures	Sighetu Marmatiei	0.675	21	582	19	565	104	5	2
Mures	Sighisoara	0.805	14	307	14	301	55	4	0
Mures	Tarnaveni	0.576	8	231	8	194	46	3	0
Neamt	Roman	0.847	21	825	19	672	48	3	1
Prahova	Campina	0.733	16	530	16	554	76	4	0
Salaj	Zalau	0.869	22	759	21	741	125	7	1
Satu Mare	Carei	0.830	10	247	8	224	54	3	0
Suceava	Gura Humorului	0.479	8	304	9	289	48	3	0
Suceava	Radauti	0.775	16	672	18	672	114	4	2
Teleorman	Alexandria	0.708	15	699	16	746	88	4	2
Timis	Lugoj	0.567	14	427	12	373	131	6	0
Valcea	Dragasani	0.818	12	328	7	308	108	2	2
Vaslui	Birlad	0.856	20	758	18	694	158	8	2
Vrancea	Adjud	0.804	9	314	7	280	21	2	0
	Total	-	663	20,874	614	19,793	3,898	194	44
	Mean	0.776	13.8	435	12.8	412	81.2	4.0	0.9
	Min	0.479	7	170	6	168	21	1	0
	Max	0.936	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 during 2008-2014 (see the notes to Table A29). "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. They are for households 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
Curricular focus	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	Curricular focus	Teachers	Pass bacc.	College	Wages
Location	1							
Siblings and friends	0.482	1						
Peer quality	0.571	0.599	1					
Curricular focus	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 households' baseline scores for track characteristics.

Table A7: 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.

Table A8: 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	0.845	0.362	0.025	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 A9: 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 A10: 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 A11: 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 A12: 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 I for details on the samples.

Table A13: 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	0.532	0.016	368	3,649	121,880
2016	0.530	0.017	362	3,541	115,902
2017	0.506	0.019	351	3,427	109,694
2019	0.502	0.020	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 3 for additional details.

Table A14: Summary statistics on households' track choices:
Feasible tracks with the same curricular focus 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}	64.1	61.3	66.7	66.1	61.8	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	0.44	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 4 using a different choice set. The choice set in Table 4 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 focus as that of the student's track. See Table 4 for additional details.

Table A15: 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 for 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 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 7. 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 represent difference-in-difference regressions. The covariates in each specification are listed under "Controls." Standard errors are clustered by the middle school treatment-control pairs within which we conducted the randomization.

Table A16: Effects on the 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 10, but for additional outcome variables. The outcome variables are track-level means for the listed baseline quality scores. As in Table 10, 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 10 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 A17: 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 11. However, the sample is limited to students with transition scores in the bottom half of the national distribution. See the notes to Table 11 for additional details.

Table A18: 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 11. However, the sample is limited to students with transition scores in the top half of the national distribution. See the notes to Table 11 for additional details.

Table A19: 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 12. However, the sample is limited to students with transition scores in the bottom half of the national distribution. See the notes to Table 12 for additional details.

Table A20: 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 12. However, the sample is limited to students with transition scores in the top half of the national distribution. See the notes to Table 12 for additional details.

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

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

The table presents treatment effects on beliefs and preference rankings, distinguishing by a household's degree of certainty in their preference ranking at the time of the baseline survey. "Uncert. or somewhat certain" are households who reported being uncertain or somewhat certain of their preference rankings during this survey. "Very certain" are households who reported already being very certain. Panel A presents results from regression (2), as in Table 11. Panel B presents results from regression (3), as in Table 12. The sample is for tracks other than a household's two top baseline choices. See the notes to Tables 11 and 12 for additional details.

Table A22: 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)	(0.086)	(0.090)	(0.087)	(0.082)	(0.089)
Siblings and friends	0.335***	0.305***	0.316***	0.351***	0.289***	0.322***	0.328***	0.307***	0.325***	0.315***
	(0.080)	(0.076)	(0.081)	(0.080)	(0.077)	(0.063)	(0.067)	(0.064)	(0.062)	(0.067)
Peer quality	0.064	0.103	0.076	0.126	0.109	0.542***	0.473***	0.508***	0.562***	0.429***
	(0.093)	(0.088)	(0.094)	(0.090)	(0.106)	(0.078)	(0.082)	(0.082)	(0.081)	(0.083)
Curricular focus	0.711***	0.611***	0.717***	0.774***	0.616***	1.09***	0.937***	1.01***	1.15***	0.886***
	(0.117)	(0.109)	(0.113)	(0.114)	(0.107)	(0.092)	(0.090)	(0.083)	(0.093)	(0.092)
VA: pass the bacc.	0.265**				0.123	0.430***				-0.037
	(0.111)				(0.101)	(0.109)				(0.114)
VA: college		0.340***			0.170		0.626***			0.447***
		(0.111)			(0.121)		(0.093)			(0.103)
VA: wages			0.317***		0.227**			0.572***		0.342***
			(0.097)		(0.095)			(0.078)		(0.089)
Teacher quality				0.041	-0.103				0.296***	0.060
				(0.132)	(0.116)				(0.090)	(0.096)
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 13, but separately for low- and high-achieving students. See the notes to Table 13 for additional details.

Table A23: 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***	0.379***
	(0.116)	(0.144)	(0.091)	(0.086)	(0.085)	(0.099)
Siblings and friends	0.335***	0.405***	0.264***	0.322***	0.427***	0.302***
	(0.080)	(0.118)	(0.076)	(0.063)	(0.069)	(0.063)
Peer quality	0.064	-0.041	-0.200***	0.542***	0.623***	0.753***
	(0.093)	(0.116)	(0.064)	(0.078)	(0.101)	(0.064)
Curricular focus	0.711***	0.901***	1.06***	1.09***	1.11***	0.832***
	(0.117)	(0.133)	(0.076)	(0.092)	(0.102)	(0.073)
VA: pass the bacc.	0.265**	0.255**	0.424***	0.430***	0.330**	0.634***
	(0.111)	(0.125)	(0.114)	(0.109)	(0.137)	(0.095)
R-sq.	0.21	0.25	0.20	0.40	0.44	0.38
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 A22. "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 V.A. See the notes to Table 13 for additional details on estimating the preference model.

Table A24: 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***	0.320***	0.254**	0.379***	0.378***	0.359***
	(0.091)	(0.082)	(0.104)	(0.099)	(0.099)	(0.093)
Siblings and friends	0.264***	0.247***	0.263***	0.302***	0.303***	0.285***
	(0.076)	(0.080)	(0.079)	(0.063)	(0.063)	(0.061)
Peer quality	-0.200***	-0.106	-0.085	0.753***	0.754***	0.752***
	(0.064)	(0.066)	(0.092)	(0.064)	(0.064)	(0.063)
Curricular focus	1.06***	0.981***	0.917***	0.832***	0.831***	0.800***
	(0.076)	(0.077)	(0.102)	(0.073)	(0.073)	(0.070)
VA: pass the bacc.	0.424***	0.474***	0.307***	0.634***	0.634***	0.740***
	(0.114)	(0.110)	(0.113)	(0.095)	(0.095)	(0.099)
R-sq.	0.20	0.22	0.22	0.38	0.38	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 A23. 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 V.A). See the notes to Table 13 for additional details on estimation.

Table A25: 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***	0.293***	0.327***	0.305***	0.280***	0.379***	0.352***	0.340***	0.407***	0.371***
	(0.091)	(0.092)	(0.092)	(0.090)	(0.088)	(0.099)	(0.101)	(0.098)	(0.094)	(0.097)
Siblings and friends	0.264***	0.258***	0.265***	0.251***	0.239***	0.302***	0.285***	0.299***	0.320***	0.316***
	(0.076)	(0.076)	(0.076)	(0.077)	(0.077)	(0.063)	(0.062)	(0.063)	(0.060)	(0.061)
Peer quality	-0.200***	-0.162***	-0.235***	0.063	0.027	0.753***	0.830***	0.441***	0.396***	0.244***
	(0.064)	(0.061)	(0.065)	(0.065)	(0.063)	(0.064)	(0.065)	(0.067)	(0.063)	(0.062)
Curricular focus	1.06***	1.13***	1.09***	0.988***	1.01***	0.832***	0.840***	1.01***	1.05***	1.11***
	(0.076)	(0.076)	(0.076)	(0.078)	(0.078)	(0.073)	(0.071)	(0.065)	(0.066)	(0.066)
VA: pass the bacc.	0.424***	0.216*	0.381***	0.294***	0.271**	0.634***	0.502***	0.378***	0.417***	0.310***
	(0.114)	(0.118)	(0.113)	(0.114)	(0.115)	(0.095)	(0.091)	(0.090)	(0.093)	(0.091)
Measured track characteristics:										
Value added,		0.286***			0.171**		0.235***			0.302***
V_{jt} (s.d.)		(0.041)			(0.067)		(0.054)			(0.082)
Selectivity,			0.124*		0.139*			1.22***		0.786***
MTS_{jt} (s.d.)			(0.066)		(0.084)			(0.133)		(0.141)
Humanities				0.340***	-0.093				1.30***	0.402**
				(0.109)	(0.149)				(0.202)	(0.193)
Math or science				-0.428***	-0.685***				1.86***	1.16***
				(0.119)	(0.127)				(0.213)	(0.210)
R-sq.	0.20	0.21	0.20	0.22	0.22	0.38	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 technical tracks. All columns impute missing quality scores using a random forest, as described in Section V.A. See the notes to Table 13 for additional details on estimating the preference model.

Table A26: 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***	0.280***	0.282***	0.379***	0.371***	0.386***
	(0.091)	(0.088)	(0.088)	(0.099)	(0.097)	(0.097)
Siblings and friends	0.264***	0.239***	0.236***	0.302***	0.316***	0.313***
	(0.076)	(0.077)	(0.078)	(0.063)	(0.061)	(0.063)
Peer quality	-0.200***	0.027	-0.002	0.753***	0.244***	0.164***
	(0.064)	(0.063)	(0.064)	(0.064)	(0.062)	(0.063)
Curricular focus	1.06***	1.01***	0.972***	0.832***	1.11***	1.00***
	(0.076)	(0.078)	(0.084)	(0.073)	(0.066)	(0.073)
VA: pass the bacc.	0.424***	0.271**	0.202*	0.634***	0.310***	0.068
	(0.114)	(0.115)	(0.113)	(0.095)	(0.091)	(0.095)
VA: college			0.112			0.392***
			(0.099)			(0.086)
VA: wages			0.066			0.050
			(0.105)			(0.080)
Teacher quality			-0.034			-0.041
			(0.104)			(0.104)
Measured track characteristics:						
Value added, V_{jt} (s.d.)		0.171**	0.174**		0.302***	0.286***
		(0.067)	(0.068)		(0.082)	(0.083)
Selectivity, MTS_{jt} (s.d.)		0.139*	0.137		0.786***	0.769***
		(0.084)	(0.084)		(0.141)	(0.135)
Humanities		-0.093	-0.059		0.402**	0.512***
		(0.149)	(0.147)		(0.193)	(0.193)
Math or science		-0.685***	-0.677***		1.16***	1.19***
		(0.127)	(0.127)		(0.210)	(0.210)
R-sq.	0.20	0.22	0.22	0.38	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 V.B. 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 V.A. See the notes to Table 13 for additional details on estimating the preference model.

Table A27: The effect of accurate beliefs on the value added of students' tracks: results in levels of value added

	Potential increase in VA		**Change**	**Share of**
	$V_{i,\text{IS}}$	$V_{i,\text{AS}}$	**in VA**	**pot. incr.**
Panel A: Just quality scores				
All students	10.7	8.15	2.58	0.240
Low-achieving	10.2	7.79	2.37	0.234
High-achieving	11.1	8.36	2.70	0.244
Panel B: With measured attributes				
All students	10.2	8.92	1.32	0.129
Low-achieving	8.80	7.29	1.51	0.172
High-achieving	11.1	9.89	1.21	0.109
Panel C: Update on all VA dimensions				
All students	10.3	8.30	2.02	0.196
Low-achieving	8.83	6.83	2.00	0.227
High-achieving	11.2	9.17	2.04	0.182
Panel D: Adjust for measurement error				
All students	10.2	8.28	1.90	0.187
Low-achieving	8.67	6.50	2.17	0.251
High-achieving	11.1	9.34	1.73	0.157

The table is analogous to Table 14. However, it reports results in levels of value added, rather than in standard deviation units. Specifically, "Potential increase in VA" and "Change in VA" are in terms of percentage points of passing the baccalaureate exam.

Table A28: 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	0.207	0.214	0.203
Top 2	0.215	0.198	0.225
Top 3	0.225	0.190	0.246
Top 4	0.238	0.194	0.265
Plausible: Top 2	0.224	0.223	0.225
Feasible: Top 2	0.217	0.151	0.256
Panel B: With measured attributes			
Top 1	0.112	0.146	0.092
Top 2	0.110	0.126	0.101
Top 3	0.108	0.124	0.098
Top 4	0.115	0.129	0.107
Plausible: Top 2	0.103	0.106	0.101
Feasible: Top 2	0.102	0.074	0.118
Panel C: Update on all VA dimensions			
Top 1	0.223	0.232	0.218
Top 2	0.169	0.167	0.170
Top 3	0.145	0.140	0.148
Top 4	0.142	0.127	0.150
Plausible: Top 2	0.161	0.146	0.170
Feasible: Top 2	0.189	0.194	0.186
Panel D: Adjust for measurement error			
Top 1	0.160	0.207	0.133
Top 2	0.158	0.181	0.145
Top 3	0.155	0.179	0.141
Top 4	0.166	0.186	0.153
Plausible: Top 2	0.148	0.153	0.144
Feasible: Top 2	0.146	0.109	0.168

The table presents the mean difference between $V_{i,AS}$ and $V_{i,IS}$ for alternative specifications of the preference model. See Section V.B 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 39 for additional details on fitting the rank-ordered logit. See the notes to Table 14 for details on the sample.

A Value added

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

A.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 I.C.

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, $f(X_i)$.[1] 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.[2] 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.[3]

1. 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.

2. 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.

3. 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.

A.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 B 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 B) 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.[4]

A.3 Empirical Bayes posteriors and machine learning forecasts

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

4. 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.[5] 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, $\mathrm{V}_{jt}^{\mathrm{EB}}$. 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 C.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).[6] Our model incorporates current and lagged values of a large number of track covariates. Notably, this includes a track's prior value added, its curricular focus, and its past and current selectivity and demographics. Tables A9-A11 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.[7] 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 C.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 A29. 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, $\mathrm{V}_{jt}^{\mathrm{EB}}$. For 2015-2017 and 2019, it is equal to the machine learning forecasts, V_{jt}^{P}.

A.4 The magnitude of value added

Table A29 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 $\hat{\mathrm{V}}_{jt}$ for measurement error. For 2015-2017 and 2019, they are calculated by adjusting the standard deviations of V_{jt}^{P} for forecast error.[8] 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 A29 reveals that tracks vary widely in both pass rates and value added.[9] For the 2008-

5. 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.

6. 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.

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

8. The procedures used in making these adjustments are described in Appendices C.1 and C.4.

9. 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 A29: 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	0.824	476	4,851	172,174
2009	0.369	0.153	0.146	0.146	0.133	0.796	438	4,470	170,087
2010	0.365	0.137	0.130	0.130	0.118	0.749	417	4,018	164,146
2011	0.364	0.130	0.123	0.123	0.114	0.762	437	4,506	187,442
2012	0.374	0.123	0.115	0.115	0.112	0.797	410	4,234	146,114
2013	0.372	0.114	0.105	0.105	0.105	0.779	420	4,269	141,934
2014	0.356	0.125	0.116	0.116	0.111	0.795	378	3,784	124,675
2015	-	0.124	0.110	-	0.110	-	368	3,649	121,880
2016	-	0.121	0.108	-	0.108	-	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	0.195	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	-	0.122	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 Appendices A.3 and C.2 for details. See Appendices C.1 and C.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 C.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.[10] 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}.

B 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.

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

B.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, $m_i^c > 0$.[11] 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:

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

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,

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

11. 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.[12]

B.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.

B.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.

B.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:

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

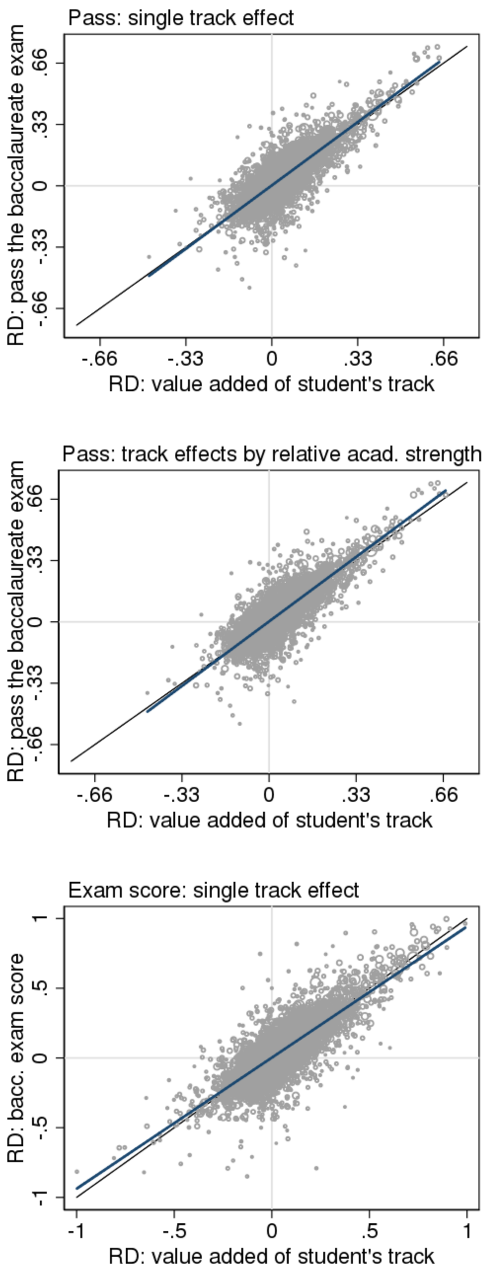
for $i \in \mathcal{I}_l$ with $|m^c_i| \leq 1$ and $m^c_i \neq 0$. Here, y_i is an outcome, λ_t is an intercept that varies by cohort, $\lambda_1 \cdot m^c_i + \mathbb{1}\{m^c_i \geq 0\} \cdot \phi_1 \cdot m^c_i$ is a linear spline in the running variable, and ϕ_0 is the RD treatment effect (RD^y_c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

B.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

12. 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.

Figure A8: Admissions-cutoff RDs



The figure plots estimates of admissions-cutoff RDs on baccalaureate outcomes, $\hat{\text{RD}}_c^p$, versus those on the value added of students' tracks, $\hat{\text{RD}}_c^V$. 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 Appendix B.5 for additional details.

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 exam score; it again uses a value added measure of a single track-year effect on this outcome.

In the plots, each dot represents a different cutoff. The grey diagonal line is a 45-degree line, and the blue line is a line of best fit from a linear regression. If the RDs on value added are an unbiased predictor of the RDs on baccalaureate outcomes, then the best fit line will equal the 45-degree line. If the RDs on the two outcomes are always equal, then all the dots will 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 A30 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 $|m^c_i| \leq 1$ and $m^c_i \neq 0$), 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 A30. 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 A30: Comparing admissions-cutoff RDs

Value added measure	**R-squared**		**Cutoffs**	**Student-cutoffs**
	Raw	Adjusted		
Pass the exam:				
All	0.746	0.994	10,210	24,173,143
Gender	0.753	0.996	10,210	24,173,143
Relative academic strength	0.746	0.987	10,210	24,173,143
Percentile rank of exam performance	0.698	0.964	10,210	24,173,143
Exam score	0.719	0.981	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 A30 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 0.987). The baccalaureate outcomes in the fourth and fifth rows are the percentile rank of the student's exam performance and the student's exam score (with imputations for missing values). The value added measures in these rows are single track-year effects on the given outcomes. It can be seen that the adjusted R-squared remains extremely high (0.964 and 0.981, respectively).

B.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 A31: Testing for bias using the Angrist et al. (2017) IV strategy

	Group									
	1	2	3	4	5	6	7	8	9	10
IV coefficient	1.02	1.05	1.04	1.02	1.05	1.07	1.00	1.09	0.97	1.04
	(0.019)	(0.019)	(0.020)	(0.019)	(0.020)	(0.019)	(0.021)	(0.021)	(0.021)	(0.021)
First-stage F statistic	29.2	28.2	26.6	27.0	24.8	26.5	23.9	24.4	25.5	24.7
Bias										
Wald statistic	1.58	7.06	4.88	1.38	5.39	12.2	0.05	16.8	2.04	3.94
p-value	0.208	0.008	0.027	0.240	0.020	0.000	0.830	0.000	0.153	0.047
Overidentification										
Hansen J statistic	1,330	1,429	1,437	1,556	1,404	1,447	1,456	1,410	1,418	1,497
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	2,446,084	2,428,844	2,428,808	2,384,319	2,500,574	2,444,417	2,381,536	2,494,635	2,423,694	2,195,957

The table presents results from the strategy of Angrist et al. (2017), described in Appendix B.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 A31 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.

C Adjusting for measurement error

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

C.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 $\hat{\mathrm{V}}_{jt}$.

C.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 Appendix C.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: $\hat{\mathrm{V}}_{jt} \sim N(\mathrm{V}^*_{jt}, \mathrm{Var}[\varepsilon_{jt}|jt])$.[13]

13. 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.[14] Further, they have a common variance, and they are centered around a grand mean which we allow to vary by curricular focus:

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

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

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

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

with

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

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 focus 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 focus c, and $N_{\mathcal{S}_{ct}}$ is the number of students in this set. Third, we estimate σ^2_t as:

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

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 focuses. 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

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

C.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:

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

14. 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 Appendix C.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 $\hat{\text{V}}_{jt}$.

C.4 The standard deviation of V^*_{jt} based on 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 A29). 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 $R^2_{0814} = 0.791$.

C.5 R-squared for beliefs about V^*_{jt}, proxied by V^P_{jt}

In Section III, 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 Appendix C.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.[15] 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.

15. 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.

D 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 focuses.

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".[16] 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.[17] 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 A32: 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 I.D, 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 A7). Table A32 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.[18] It shows that this distribution is more bimodal, with most households

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

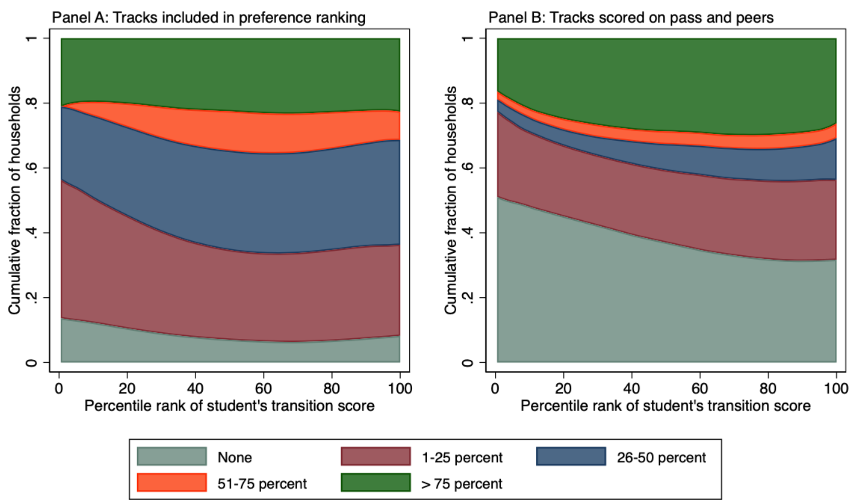
17. Omitting non-selective tracks is not an issue, as these are less preferred than the tracks that are ranked.

18. 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.

assigning scores 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 and the remaining columns of Table A32 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



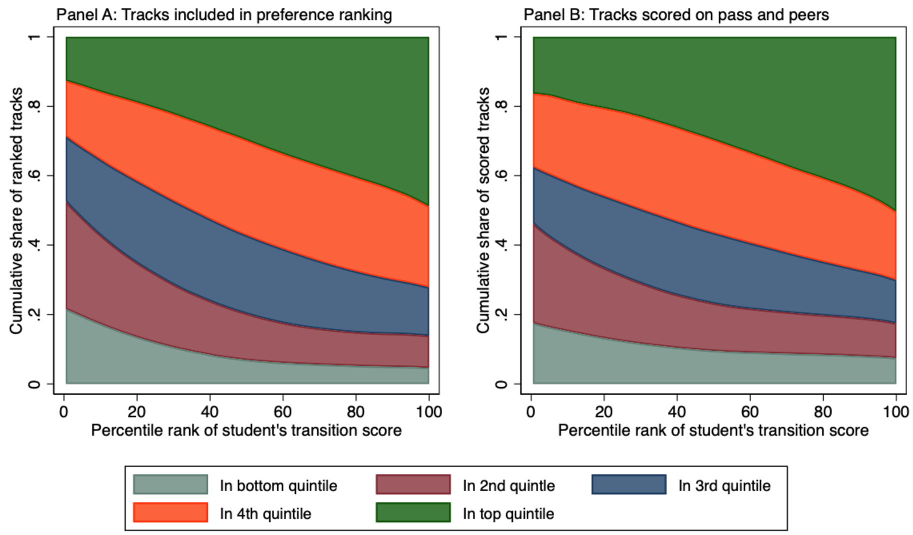
The figure shows how the share of tracks that a survey household ranks and/or scores varies with the student's transition score. Specifically, households are assigned to groups based on whether they ranked and/or scored none of the tracks in their towns, 1-25 percent of the tracks, 26-50 percent, 51-75 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.[19] 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

19. 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 focuses—or if, instead, they emphasize 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 curricular focus.

Table A33: The number of curricular focuses among a household's top choices

Among top:	**Curricular focuses**		**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 focuses that are included among a household's top choices in the baseline preference ranking. "Rank ≥ 2" is an indicator for whether the household ranks tracks from at least two focuses. "Mean" is the mean number of focuses from which a household ranks tracks. The sample in each row is restricted to households who ranked at least the listed number of tracks.

Table A33 presents the results. The table shows the share of households who include tracks from multiple curricular focuses among their top baseline choices. It also lists the mean number of focuses 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 focuses. Among the top six choices, this value is 98%.

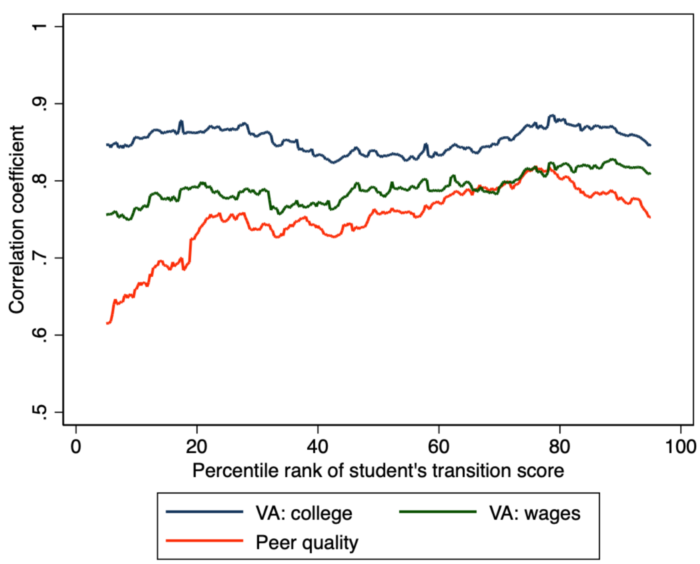
E Explaining households' beliefs about value added

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 curricular focus 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 A34: 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)
Curricular focus		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 A34 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, curricular focus, 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 curricular focus 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 curricular focus (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 curricular focus 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.

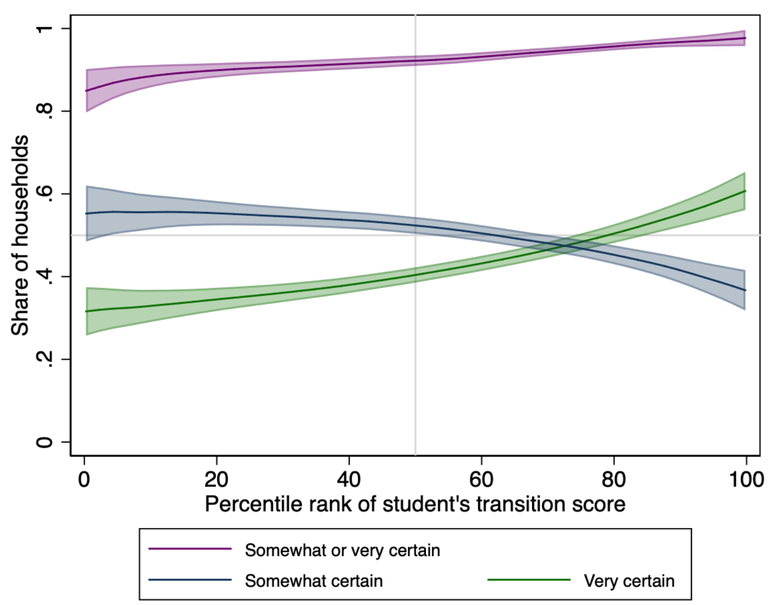
Thus, the evidence suggests that households did understand the concept of value added in advance of the experiment. Namely, they conceived value added as being related to the quality of a track's teachers, while also depending on the track's curricular focus and peers.

F 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 I.D, 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 A7). 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 A35 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 A35: 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 A36: 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 A36 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.

G 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.

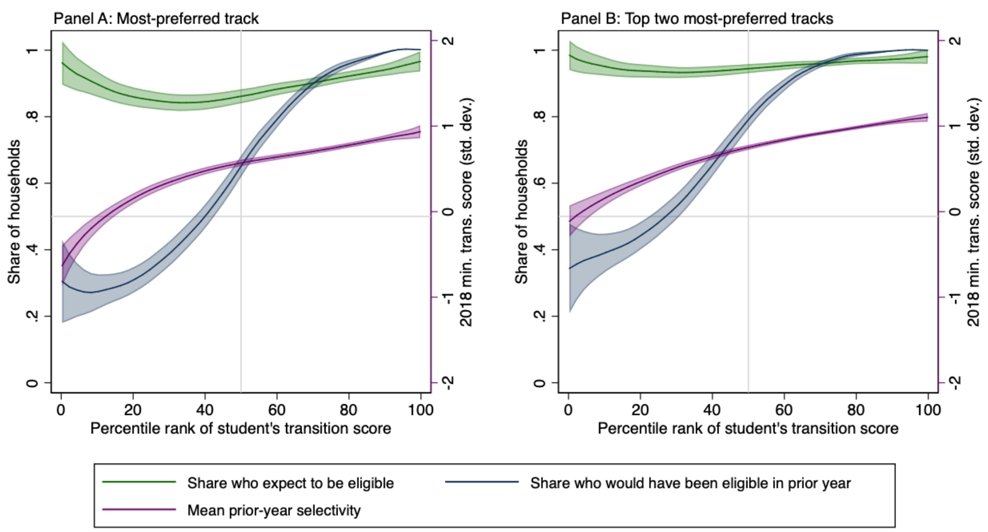
One complication for the matching algorithm was that some towns had an odd number of one-class schools. In these towns, we stopped the matching algorithm when there were three remaining one-class schools. We calculated the Mahalanobis distance of the covariates for each school in the triple to the average 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.

H 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.[20] 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.[21] 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.

20. 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.

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

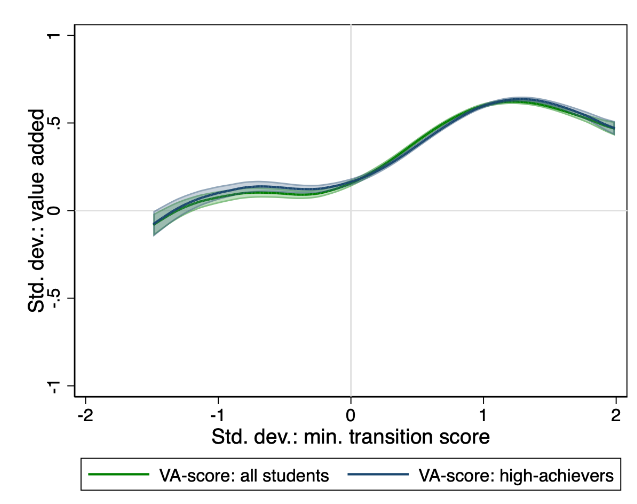
I Additional results for Section II

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

I.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.[22] 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.

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

I.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 I, 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 A37: How track utilities relate to measured values of track characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
Value added,	0.605***		-0.010	0.119***	0.033***	0.217***
V_{jt} (s.d.)	(0.008)		(0.006)	(0.007)	(0.012)	(0.011)
Selectivity,		2.05***	2.06***	2.03***	1.71***	2.26***
MTS_{jt} (s.d.)		(0.021)	(0.022)	(0.023)	(0.021)	(0.039)
Humanities				-0.468***	-0.405***	-0.384***
				(0.016)	(0.030)	(0.022)
Math or science				0.005	-0.406***	0.253***
				(0.014)	(0.019)	(0.023)
All students	x	x	x	x		
Low-achieving					x	
High-achieving						x
R-sq.	0.05	0.20	0.20	0.20	0.12	0.28
Clusters	5,969	5,969	5,969	5,969	5,911	5,714
Students	2,110,527	2,110,527	2,110,527	2,110,527	1,029,297	1,081,230
Student-tracks	47,298,149	47,298,149	47,298,149	47,298,149	13,812,463	33,485,686

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 focus (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 A37 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 curricular focus. 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.[23] 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

23. 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.

J Additional results for Section III

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

Table A38: 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 6. However, the sample is limited to households who provided quality scores for both value added and peer quality for all of the tracks in their towns. See Table 6 for additional details.

First, it is possible that the results in Table 6 are impacted by the fact that most households score only a subset of the tracks in their towns. In Table A38, we replicate Table 6 but restrict the sample to the 21% of households with no missing scores. Results are similar.

Table A39: 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 6. 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 6 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 6 would average over accurate scores for tracks that are plausibly feasible and inaccurate ones for tracks that are out of reach. In Table A39 we replicate Table 6 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 6 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 A40, are still similar.

Finally, the R-squared statistics that we provide in Table 6 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 A40: 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 6. However, the sample is limited to households who reported being "very certain" of their preference rankings in the baseline survey. See Table 6 for additional details.

Table A41: 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***	
	(0.013)		(0.023)		(0.011)	
Score: Peers		0.353***		0.322***		0.375***
		(0.021)		(0.031)		(0.021)
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 6. 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 C.5 explains how we calculate "R-sq.: V^*_{jt}". See Table 6 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 C.5 describes the procedure). Table A41 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}.

K 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 A42: 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 IV.A. "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 7 for additional details on the regressions.

The results are in Table A42. 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 IV.A. The columns labeled "All towns" replicate results from Section IV.A, 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.

L Estimating preferences using experimental variation

In this appendix, we provide additional results for Section V. In Section V, 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.

L.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.[24] Third, there may be measurement error in the endline value added scores. As we discussed in Section IV.B, 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.[25] 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.[26] 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.

24. Note that we did not impute missing scores in Section IV.B, 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.

25. As we mentioned in Section IV.C, 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.

26. 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.

L.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, $s^V_{ij,\text{fs}}$. 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.[27]

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.[28]

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 V.B. For "Location", "Siblings and friends", and "Curricular focus", we use baseline scores. For "Peer quality", we use the within-town quintile of a track's selectivity.

L.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 A43. 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 V.B.[29] In addition, the models control for a cubic

27. 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.

28. 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.

29. 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 13 and A26). 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 Appendix L.1.

Table A44 provides the results of the simulation. It shows that these are quite close to the values based on baseline data, displayed in Tables 14 and A28. In particular, Table A44 reveals that correcting households' value added scores is predicted to, on average, cause low-achieving (high-achieving) students to attend tracks with between 0.11 and 0.24 (0.09 and 0.23) s.d. worth of additional value added. Table A28 shows that the corresponding ranges based on baseline data are 0.07 to 0.23 (0.09 to 0.27).

In sum, the results in this appendix show that the findings in Section V 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 A43: Households' preferences for track attributes: results calculated using experimental variation in value added scores

	(1)	(2)	(3)	(4)	(5)
Households' quality scores:					
Location	0.008	0.002	0.001	0.003	-0.008
	(0.061)	(0.065)	(0.069)	(0.061)	(0.061)
Siblings and friends	0.368***	0.373***	0.376***	0.363***	0.363***
	(0.065)	(0.066)	(0.068)	(0.066)	(0.066)
Peer quality	0.119*	0.118*	0.115*	0.113*	0.109*
	(0.066)	(0.065)	(0.065)	(0.066)	(0.064)
Curricular focus	1.07***	1.06***	1.06***	1.04***	1.03***
	(0.074)	(0.083)	(0.095)	(0.074)	(0.075)
VA: pass the bacc.	0.399***	0.505***	0.575**	0.460***	0.536***
	(0.131)	(0.193)	(0.255)	(0.133)	(0.142)
Measured track characteristics:					
Value added, V_{jt}	0.180*	0.147	0.123	0.170*	0.154*
(s.d.)	(0.093)	(0.095)	(0.096)	(0.092)	(0.091)
Selectivity, MTS_{jt}	0.345*	0.331*	0.321*	0.329*	0.307*
(s.d.)	(0.177)	(0.175)	(0.174)	(0.177)	(0.173)
Humanities	-0.331*	-0.299	-0.276	-0.360**	-0.369**
	(0.176)	(0.182)	(0.187)	(0.174)	(0.175)
Math or science	0.143	0.148	0.155	0.130	0.133
	(0.163)	(0.169)	(0.173)	(0.163)	(0.163)
Control function:					
Residuals	0.640***	0.686***	0.639**	0.509***	0.391***
	(0.151)	(0.215)	(0.276)	(0.150)	(0.150)
Squared residuals	0.438***	0.596***	0.655***	0.448***	0.430***
	(0.038)	(0.050)	(0.052)	(0.037)	(0.035)
Cubed residuals	-0.090***	-0.128***	-0.125***	-0.081***	-0.064***
	(0.017)	(0.027)	(0.031)	(0.018)	(0.019)
R-sq.	0.37	0.38	0.38	0.37	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 Appendix L.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 Appendix L.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 A44: The effect of accurate beliefs on the value added of students' tracks: results based on experimental preference estimates

	Change in value added: $V_{i,AS} - V_{i,IS}$		
	All students	Low-achieving	High-achieving
Panel A: Accurate			
Top 1	0.162	0.165	0.160
Top 2	0.168	0.171	0.167
Top 3	0.171	0.171	0.171
Top 4	0.161	0.160	0.161
Plausible: Top 2	0.154	0.161	0.149
Feasible: Top 2	0.096	0.107	0.090
Panel B: Two-thirds accurate			
Top 1	0.199	0.204	0.196
Top 2	0.206	0.210	0.203
Top 3	0.206	0.207	0.205
Top 4	0.185	0.185	0.185
Plausible: Top 2	0.183	0.193	0.178
Feasible: Top 2	0.102	0.113	0.096
Panel C: Half accurate			
Top 1	0.232	0.239	0.228
Top 2	0.231	0.237	0.228
Top 3	0.229	0.230	0.228
Top 4	0.199	0.199	0.199
Plausible: Top 2	0.201	0.211	0.195
Feasible: Top 2	0.098	0.108	0.092
Panel D: Accurate other than top 2			
Top 1	0.190	0.194	0.187
Top 2	0.197	0.201	0.195
Top 3	0.198	0.199	0.198
Top 4	0.185	0.185	0.185
Plausible: Top 2	0.184	0.194	0.179
Feasible: Top 2	0.118	0.133	0.110
Panel E: Accurate other than top 4			
Top 1	0.218	0.223	0.214
Top 2	0.231	0.237	0.228
Top 3	0.234	0.237	0.233
Top 4	0.220	0.221	0.219
Plausible: Top 2	0.217	0.230	0.210
Feasible: Top 2	0.147	0.166	0.137

The table summarizes the difference between $V_{i,AS}$ and $V_{i,IS}$ 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 Appendix L.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 A28 for more details. The sample is students in the follow-up survey.

References

Abdulkadiroglu, Atila, Parag A. Pathak, Jonathan Schellenberg, and Christopher R. Walters. 2020. "Do Parents Value School Effectiveness?" *American Economic Review* 110, no. 5 (May): 1502–39.

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.

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.

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.

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.

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.

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.

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, 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).

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.

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.

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.

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.

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.

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

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

Rothstein, Jesse. 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.