Online Appendix for
“The returns to STEM programs for less-prepared students”
by Kevin Ng and Evan Riehl

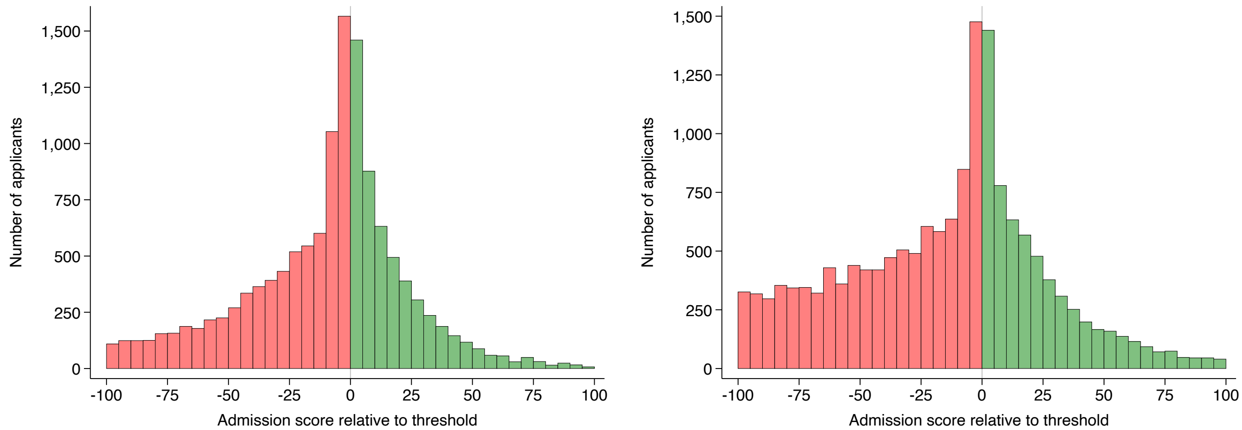
Outline:

A. Appendix figures and tables

B. Theoretical appendix

C. Empirical appendix

A. Appendix figures and tables



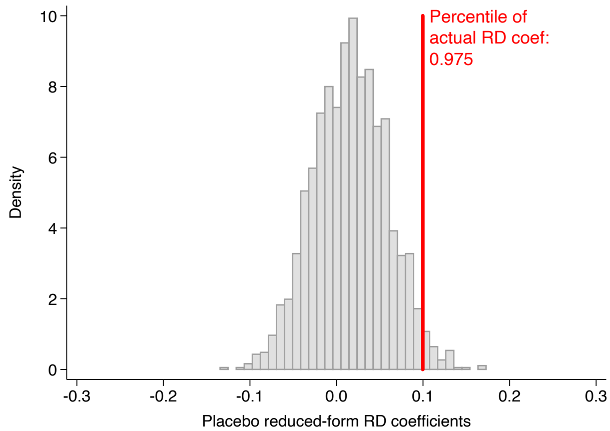
Panel A. STEM applicants

Panel B. Other applicants

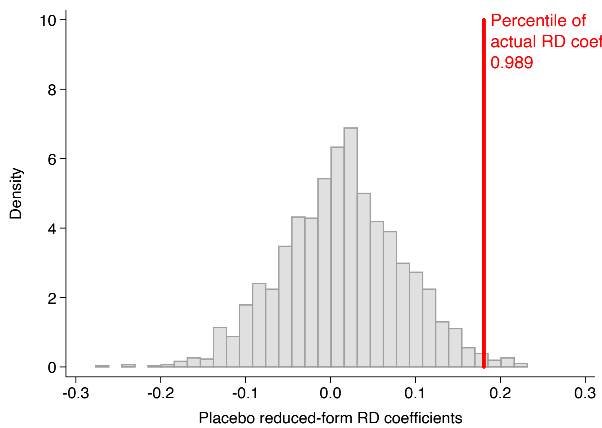
Figure A1. Density of admission scores relative to the threshold

Notes: This figure shows the density of admission scores relative to the admission thresholds. The x-axis is a student's admission score normalized to zero at the threshold. The y-axis shows the number of applicants within five unit bins of the admission score. The graphs are limited to those with normalized scores between -100 and 100.

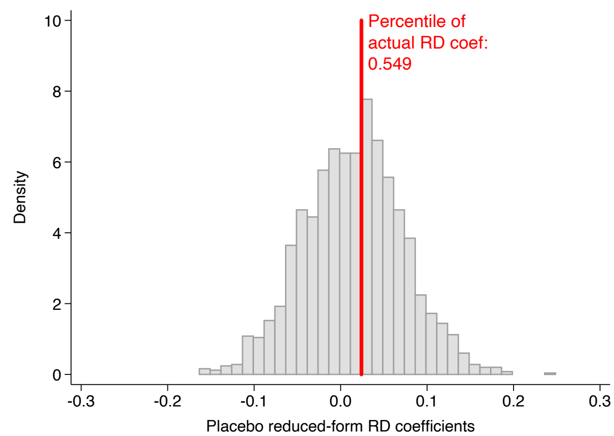
Panel A shows the distribution of admission scores for applicants to Univalle STEM programs. Using the McCrary (2008) density test, the estimated discontinuity—i.e., the log difference in height at the threshold—is -0.049 with a standard error of 0.030. Panel B shows the distribution of admission scores for applicants to non-STEM Univalle programs. The estimated density discontinuity is 0.017 with a standard error of 0.026.



Panel A. All STEM applicants



Panel B. Less-prepared STEM applicants



Panel C. More-prepared STEM applicants

FIGURE A2. Placebo RD estimates for log monthly earnings — STEM applicants

Notes: This figure displays placebo RD estimates for STEM applicants' log monthly earnings.

We follow Beuermann and Jackson (2022)'s method of generating these placebo RD estimates (see their Appendix Figure A2). First, we randomly choose an admission rank as the placebo cutoff in each application pool. We then estimate our reduced-form RD regression (equation 1) with log monthly earnings as the dependent variable, and we define the placebo running variable, x_{ip}, and above-threshold indicator, D_{ip}, relative to the placebo cutoffs.

The gray bars in each graph plot the distribution of 2,000 placebo reduced-form RD coefficients estimated using this method. The sample for Panel A includes all STEM applicants. The samples for Panels B and C include less- and more-prepared STEM applicants, respectively. In each graph, the vertical red lines depict the actual reduced-form RD coefficients for log monthly earnings, and we report the percentile of these actual coefficients in the placebo distribution. The actual reduced-form RD earnings coefficients are 0.100 for all STEM applicants, 0.181 for less-prepared STEM applicants, and 0.024 for more-prepared STEM applicants.

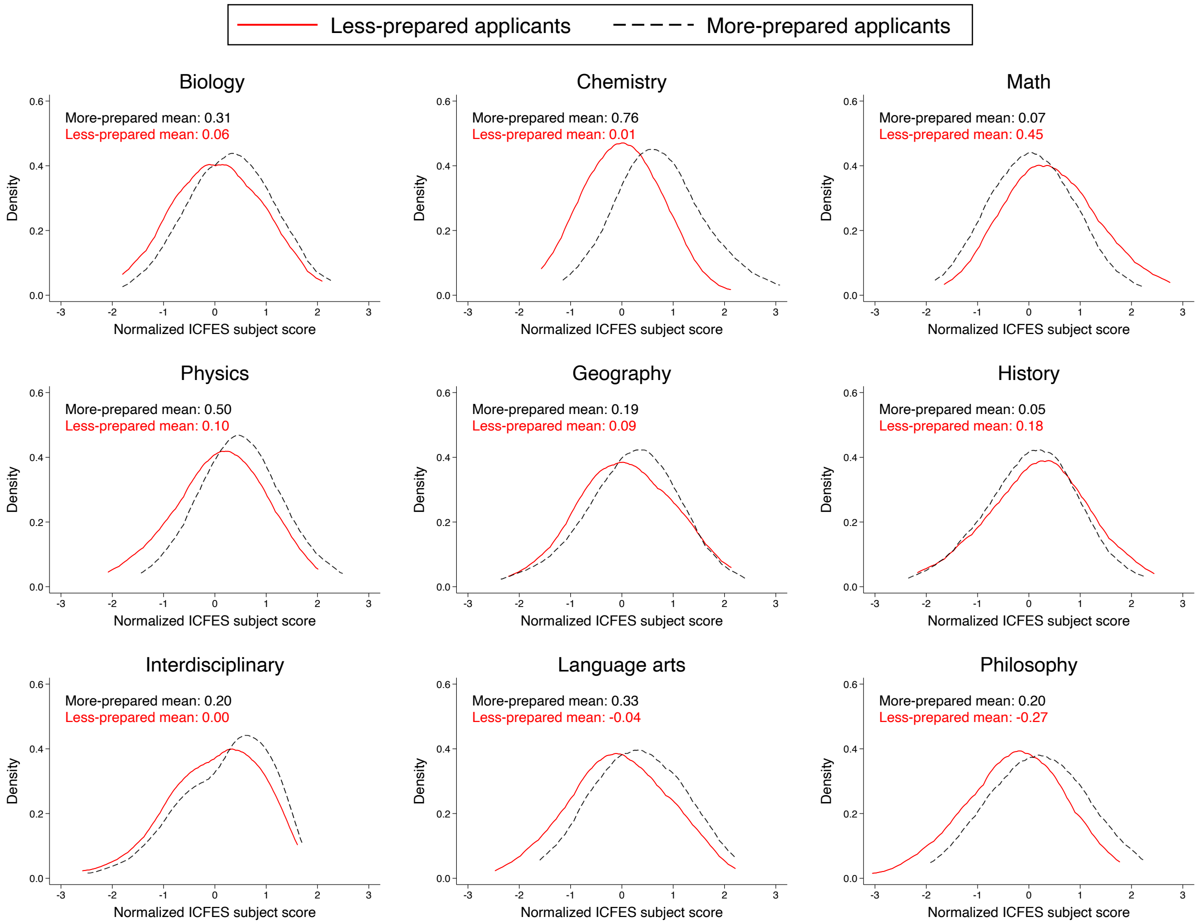
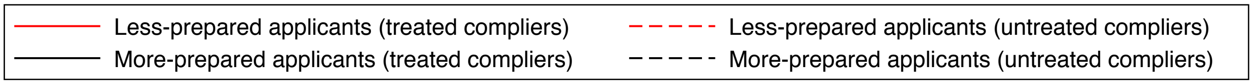
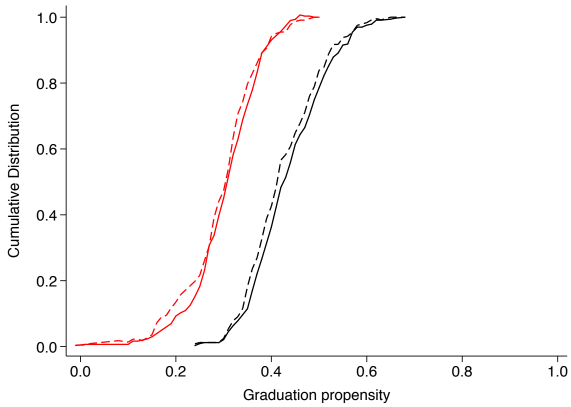


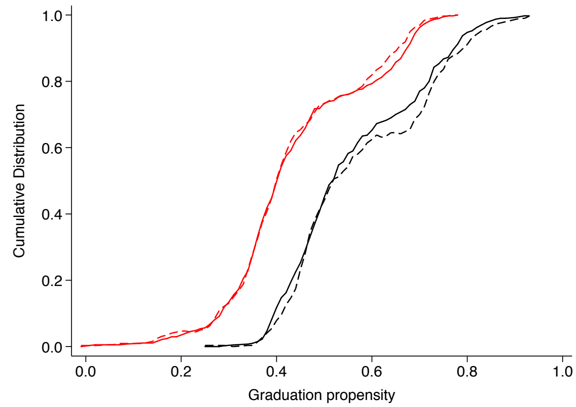
FIGURE A3. ICFES subject score distributions for marginal STEM applicants by academic preparation

Notes: This figure plots distributions of ICFES subject scores for marginal STEM applicants by academic preparation. The sample includes applicants to Univalle STEM programs who were within five positions of the admission thresholds and who took the post-2000 version of the ICFES exam. Each graph shows score distributions for a different ICFES subject exam, as indicated in the graph title. Subject scores are normalized to be mean zero and standard deviation one for the full sample of applicants to all Univalle programs. Solid red lines show score distributions for less-prepared applicants, and black dashed lines show score distributions for more-prepared applicants. Each graph reports the mean normalized score in the less- and more-prepared samples.

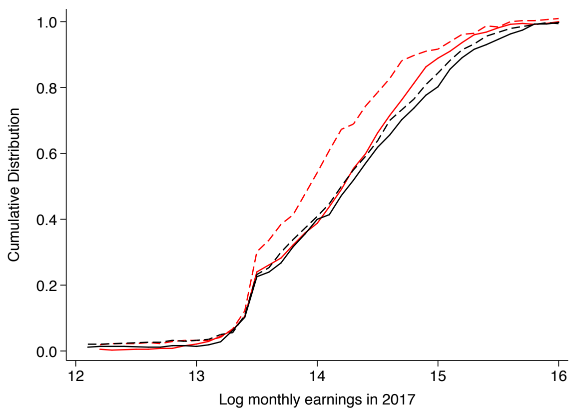




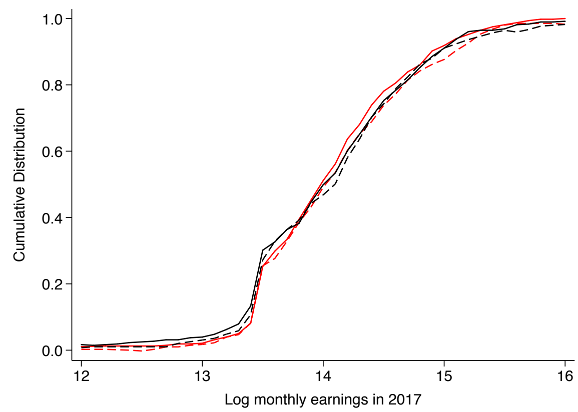
Panel A. Graduation propensity
(STEM applicants)



Panel B. Graduation propensity
(Non-STEM applicants)



Panel C. Log monthly earnings
(STEM applicants)

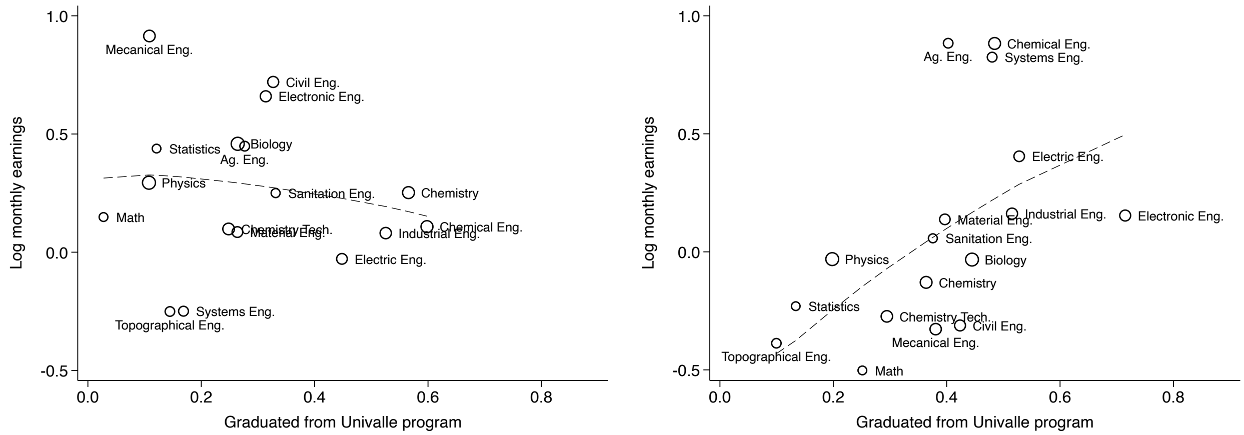


Panel D. Log monthly earnings
(Non-STEM applicants)

FIGURE A4. Abadie (2002) Cumulative Distributions of Treated and Untreated Compliers

Notes: This figure presents cumulative distribution functions (CDFs) for treated and untreated compliers following Abadie (2002). "Treated" and "untreated" are defined by admission to Univalle (above and below the admission threshold). "Compliers" are applicants who would have enrolled in the Univalle program they applied to if and only if they were admitted. We compute CDFs separately for treated and untreated compliers and for less- and more-prepared applicants defined by graduation propensity (see Section 4.1), as indicated by the legend.

The sample includes applicants to STEM programs (Panels A and C) and non-STEM programs (Panels B and D) who are within 10 positions of the admission thresholds. Panels A–B show CDFs of graduation propensity. Panels C–D show CDFs of log monthly earnings in 2017.



Panel A. Less-prepared applicants

Panel B. More-prepared applicants

FIGURE A5. Earnings and graduation RD coefficients for each Univalle STEM program

Notes: This figure plots 2SLS RD estimates for graduation rates and earnings returns for each of Univalle's 18 STEM programs. The x-axis in each panel is the program's graduation rate for marginal enrollees, which is β coefficient from separate estimation of equations (1)–(2) with an indicator for graduating as the dependent variable. The y-axis in each panel is the earnings return for marginal enrollees, which is the 2SLS RD estimate of β for each program with log monthly earnings in 2017 as the dependent variable. Panel A shows estimates for less-prepared applicants and Panel B shows estimates for more-prepared applicants. We define our less- and more-prepared samples as described in the notes to Table 5. Dashed lines show the non-parametric relationships between the earnings and graduation coefficients.

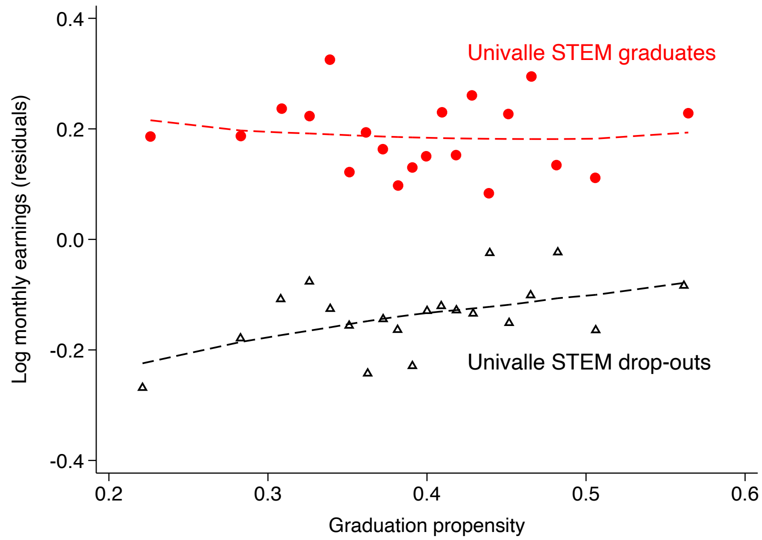
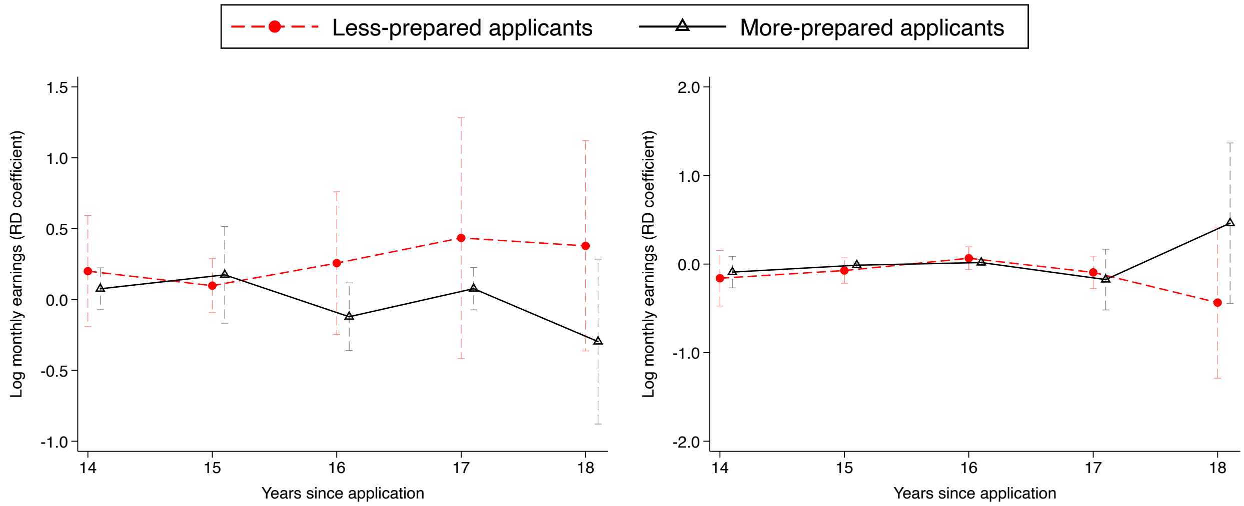


FIGURE A6. Log earnings residuals for Univalle graduates and dropouts

Notes: This figure is similar to Panel A of Figure 3, but the dependent variable is earnings residuals rather than raw earnings. We plot earnings residuals (y-axis) by graduation propensity (x-axis) for students who enrolled in Univalle's STEM programs. These residuals are generated from a regression of log monthly earnings in 2017 on a vector of individual covariates (gender, age, and dummies for high schools, mother's education categories, father's education categories, family income bins, and ICFES exam years). Markers depict means in ventiles of graduation propensity, with red circles representing students who completed the Univalle STEM program and hollow triangles representing students who dropped out. Dashed lines are predicted values from local linear regressions.

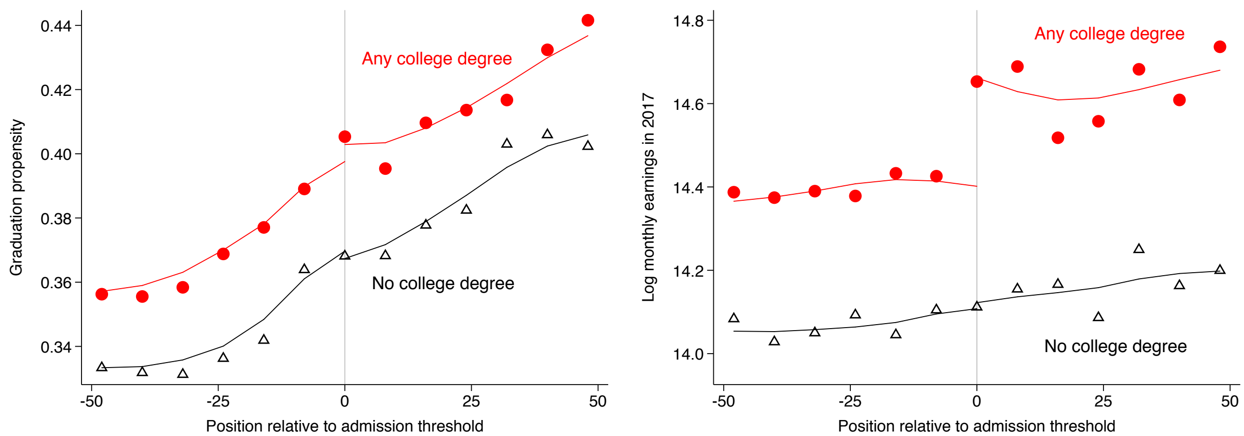


Panel A. STEM applicants

Panel B. Non-STEM applicants

FIGURE A7. Earnings returns to Univalle enrollment by years since application

Notes: This figure plots RD estimates of the earnings returns to Univalle enrollment estimated separated by years since application. We use our 2SLS RD specification (1)–(2) with log monthly earnings in 2017 as the dependent variable and estimate this regression separately for each application year in Fall 1999 through Spring 2004. The y-axis in each panel represents the RD coefficients. The x-axis represents years since application, defined as 2017 minus the year of the fall term of each academic year (e.g., 18 years since application includes applicants in Fall 1999 and Spring 2000). Panel A shows estimates for applicants to STEM programs and Panel B shows estimates for applicants to non-STEM programs. Red circles show RD coefficients for less-prepared applicants, and hollow black triangles show estimates for more-prepared applicants. We define our less- and more-prepared samples as described in the notes to Table 5. Dashed vertical lines are 95 percent confidence intervals using standard errors clustered at the individual level.

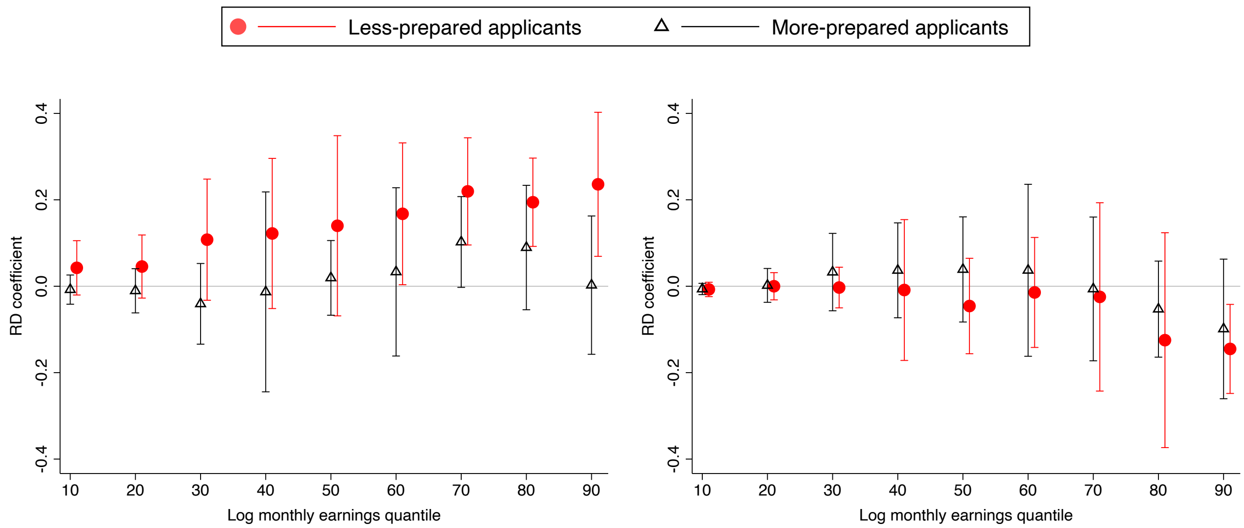


Panel A. Graduation propensity

Panel B. Log monthly earnings in 2017

FIGURE A8. STEM returns by college degree attainment

Notes: This figure plots academic preparation and earnings for STEM applicants based on whether or not they ultimately earned a college degree. The x-axis in each panel is an applicant's position in their application pool normalized to zero at the threshold. The y-axis is an individual's graduation propensity (Panel A) or log monthly earnings in 2017 (Panel B). The sample includes all STEM applicants within 50 positions of the admission threshold. Markers show the means of each variable in 8-rank bins of the admission score. Red circles include applicants who earned *any* college degree, regardless of whether it was at Univalle. Hollow triangles include those who did not. Lines are local linear regressions estimated separately above and below the thresholds for each sample.



Panel A. STEM applicants

Panel B. Non-STEM applicants

Figure A9. RD quantile regressions for log monthly earnings

Notes: This figure presents reduced form RD quantile estimates of the effects of admission to Univalle programs. The x-axis in each panel is the quantile of log monthly earnings. The y-axis shows the estimated RD coefficient at each quantile. We estimate these coefficients using the reduced form RD specification (1) with log earnings as the dependent variable. Markers show the point estimates and vertical bars show 95% confidence intervals. Panel A shows estimates for STEM applicants and Panel B show estimates for applicants to non-STEM programs. Hollow triangles show estimates for more-prepared applicants and red circles show estimates for less-prepared applicants.

Table A1. Robustness to RD specification — STEM applicants

	(A)	(B)	(C)	(D)	(E)
Bandwidth:	$h = 30$	$h = 15$	$h = 45$	CCT	$h = 30$
Kernel:	Uniform	Uniform	Uniform	Uniform	Triang.
Panel A. All applicants					
Enrolled in Univalle program	0.746	0.716	0.766	0.729	0.730
	(0.015)	(0.021)	(0.012)	(0.017)	(0.016)
Graduated from Univalle program	0.344	0.338	0.338	0.335	0.341
	(0.020)	(0.029)	(0.017)	(0.020)	(0.023)
Employed in formal sector in 2017	0.040	0.046	0.039	0.055	0.051
	(0.027)	(0.040)	(0.023)	(0.025)	(0.030)
Log monthly earnings in 2017	0.133	0.074	0.101	0.141	0.119
	(0.061)	(0.088)	(0.049)	(0.060)	(0.069)
N	6,699	3,789	8,994	5,215	6,519
Panel B. Less-prepared applicants					
Enrolled in Univalle program	0.726	0.697	0.750	0.692	0.709
	(0.022)	(0.031)	(0.018)	(0.026)	(0.024)
Graduated from Univalle program	0.288	0.271	0.274	0.281	0.281
	(0.029)	(0.042)	(0.024)	(0.030)	(0.032)
Employed in formal sector in 2017	0.019	0.024	0.027	0.027	0.034
	(0.042)	(0.062)	(0.035)	(0.049)	(0.047)
Log monthly earnings in 2017	0.244	0.176	0.142	0.235	0.257
	(0.094)	(0.143)	(0.075)	(0.100)	(0.105)
N	3,306	1,850	4,348	2,456	3,215
Panel C. More-prepared applicants					
Enrolled in Univalle program	0.761	0.743	0.777	0.766	0.750
	(0.021)	(0.029)	(0.017)	(0.018)	(0.023)
Graduated from Univalle program	0.375	0.383	0.386	0.386	0.377
	(0.028)	(0.040)	(0.023)	(0.023)	(0.031)
Employed in formal sector in 2017	0.044	0.067	0.040	0.050	0.059
	(0.037)	(0.053)	(0.031)	(0.035)	(0.040)
Log monthly earnings in 2017	0.032	0.017	0.057	0.034	0.013
	(0.083)	(0.116)	(0.068)	(0.072)	(0.089)
N	3,390	1,937	4,642	4,177	3,301

Notes: This table displays RD coefficients for STEM applicants with different bandwidths and kernels. Panel A shows estimates for our full sample of STEM applicants and Panels B–C show estimates separately for less- and more-prepared applicants. The specifications are the same as in Tables 3 and 5, but we vary the bandwidth or kernel as indicated in the column header. Column (A) replicates our benchmark results from those tables, which use an RD bandwidth of 30 positions and a uniform kernel. Columns (B) and (C) use bandwidths of 15 and 45 positions. Column (D) uses the RD bandwidth from the benchmark method of Calonico, Cattaneo and Titiunik (2014) estimated separately for each sample and outcome variable. Column (E) uses a triangular kernel with a bandwidth of 30 positions.

Parentheses contain standard errors clustered at the individual level.

Table A2. Robustness to RD specification — Non-STEM applicants

	(A)	(B)	(C)	(D)	(E)
Bandwidth:	$h = 30$	$h = 15$	$h = 45$	CCT	$h = 30$
Kernel:	Uniform	Uniform	Uniform	Uniform	Triang.
Panel A. All applicants					
Enrolled in Univalle program	0.784	0.765	0.788	0.785	0.771
	(0.013)	(0.018)	(0.011)	(0.012)	(0.014)
Graduated from Univalle program	0.498	0.514	0.499	0.501	0.509
	(0.018)	(0.024)	(0.015)	(0.014)	(0.020)
Employed in formal sector in 2017	0.035	0.034	0.018	0.022	0.028
	(0.025)	(0.035)	(0.021)	(0.020)	(0.027)
Log monthly earnings in 2017	−0.047	−0.009	−0.070	−0.050	−0.058
	(0.051)	(0.071)	(0.043)	(0.038)	(0.056)
N	7,664	4,439	10,026	8,203	7,476
Panel B. Less-prepared applicants					
Enrolled in Univalle program	0.808	0.799	0.811	0.804	0.801
	(0.018)	(0.025)	(0.015)	(0.017)	(0.020)
Graduated from Univalle program	0.475	0.495	0.491	0.485	0.488
	(0.025)	(0.034)	(0.022)	(0.019)	(0.027)
Employed in formal sector in 2017	0.038	0.072	0.001	0.018	0.040
	(0.036)	(0.049)	(0.030)	(0.032)	(0.038)
Log monthly earnings in 2017	−0.093	−0.054	−0.094	−0.086	−0.112
	(0.074)	(0.108)	(0.061)	(0.058)	(0.082)
N	3,773	2,208	4,870	4,170	3,681
Panel C. More-prepared applicants					
Enrolled in Univalle program	0.765	0.740	0.770	0.773	0.746
	(0.019)	(0.027)	(0.015)	(0.016)	(0.021)
Graduated from Univalle program	0.519	0.533	0.507	0.516	0.527
	(0.027)	(0.037)	(0.023)	(0.021)	(0.029)
Employed in formal sector in 2017	0.032	−0.014	0.032	0.024	0.017
	(0.035)	(0.051)	(0.030)	(0.026)	(0.039)
Log monthly earnings in 2017	−0.026	−0.051	−0.067	−0.063	−0.066
	(0.074)	(0.106)	(0.062)	(0.055)	(0.079)
N	3,884	2,225	5,149	4,797	3,788

Notes: This table displays RD coefficients for non-STEM applicants with different bandwidths and kernels. Panel A shows estimates for our full sample of non-STEM applicants and Panels B–C show estimates separately for less- and more-prepared applicants. The specifications are the same as in Table 3 and Appendix Table A8, but we vary the bandwidth or kernel as indicated in the column header. Column (A) replicates our benchmark results from those tables, which use an RD bandwidth of 30 positions and a uniform kernel. Columns (B) and (C) use bandwidths of 15 and 45 positions. Column (D) uses the RD bandwidth from the benchmark method of Calonico, Cattaneo and Titiunik (2014) estimated separately for each sample and outcome variable. Column (E) uses a triangular kernel with a bandwidth of 30 positions.

Parentheses contain standard errors clustered at the individual level.

Table A3. RD balance tests

	(A)	(B)	(C)	(D)	(E)	(F)
	STEM applicants			Non-STEM applicants		
Dependent variable	All	Less-prepared	More-prepared	All	Less-prepared	More-prepared
Panel A. Balance tests using individual characteristics						
ICFES percentile	0.002	0.004	0.003	0.003	0.001	0.004
	(0.005)	(0.008)	(0.006)	(0.007)	(0.010)	(0.008)
Age	0.022	0.006	0.032	−0.155	−0.180	−0.125
	(0.114)	(0.172)	(0.146)	(0.120)	(0.196)	(0.156)
College educated father	−0.011	−0.019	−0.003	0.003	−0.053	0.033
	(0.024)	(0.035)	(0.034)	(0.022)	(0.032)	(0.030)
College educated mother	0.021	0.004	0.046	0.003	−0.020	0.016
	(0.022)	(0.033)	(0.032)	(0.020)	(0.030)	(0.029)
Family income > 2x min wage	0.016	0.001	0.021	0.035	0.006	0.047
	(0.023)	(0.033)	(0.032)	(0.021)	(0.031)	(0.029)
Female	−0.024	−0.019	−0.031	−0.004	0.008	−0.004
	(0.020)	(0.029)	(0.029)	(0.020)	(0.030)	(0.028)
N	6,699	3,309	3,391	7,664	3,780	3,888
p value: Jointly zero	0.693	0.980	0.622	0.566	0.606	0.707
Panel B. Balance tests using predicted outcomes						
Enrolled in Univalle program	0.007	0.009	0.009	0.002	0.003	−0.001
	(0.007)	(0.011)	(0.009)	(0.006)	(0.009)	(0.008)
Graduated from Univalle program	−0.006	−0.006	−0.005	0.002	−0.005	0.008
	(0.005)	(0.008)	(0.007)	(0.005)	(0.007)	(0.006)
Employed in formal sector in 2017	−0.000	−0.001	0.000	−0.001	−0.001	−0.001
	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Log monthly earnings in 2017	−0.002	−0.005	0.002	0.003	−0.007	0.009
	(0.007)	(0.011)	(0.009)	(0.006)	(0.009)	(0.007)
N (w/ all characteristics defined)	5,031	2,419	2,605	5,345	2,573	2,765

Notes: This table displays RD balance tests. We estimate our reduced form RD specification (1) using the dependent variable listed in the row header and display the θ coefficient. In Panel A, the dependent variables are individual characteristics, and the last row reports p values from F tests that the coefficients on all characteristics are jointly equal to zero. In Panel B, the dependent variables are predicted outcomes based on individual characteristics. To define these predicted outcomes, we regress the outcome listed in the row header on all of the covariates from Panel A and application pool dummies. We estimate these regressions separately for STEM and non-STEM applicants and take the predicted values. We then use these predicted outcomes as dependent variables in the balance tests in Panel B.

Columns (A)–(C) include applicants to STEM programs and columns (D)–(F) include non-STEM applicants. Columns (A) and (D) include all applicants to these programs. Columns (B) and (E) include less-prepared applicants. Columns (C) and (F) include more-prepared applicants. We define our less- and more-prepared samples as described in the notes to Table 5.

Parentheses contain standard errors clustered at the individual level.

Table A4. Formal and informal sector monthly earnings in 2017

	(A)	(B)	(C)	(D)	(E)	(F)	(G)
				Mean earnings (1000s of COP)		SD of earnings (1000s of COP)	
Education/industry group	*N*	Prop. employed	Prop. in formal sector	Formal sector	Informal sector	Formal sector	Informal sector
Panel A. By highest degree completed							
High school degree	29,204	0.78	0.50	1,069	670	650	527
Technical degree	12,710	0.82	0.70	1,237	742	780	694
Bachelor's degree	10,440	0.86	0.80	2,459	1,291	2,338	1,562
All high school and above	52,354	0.80	0.60	1,435	732	1,398	716
Panel B. By industry of employment							
A. Agriculture and livestock	917	1.00	0.30	980	552	451	513
B. Fishing	37	1.00	0.09	790	527	545	265
C. Mining	241	1.00	0.81	3,371	663	4,160	389
D. Manufacturing	4,691	1.00	0.68	1,264	723	1,331	621
E. Electricity, gas and water utilities	408	1.00	0.96	1,506	617	1,299	147
F. Construction	1,962	1.00	0.41	1,590	844	1,784	485
G. Wholesale and retail	9,688	1.00	0.50	1,218	706	1,318	821
H. Hotels and restaurants	2,659	1.00	0.39	1,352	705	1,251	541
I. Transportation and communications	3,878	1.00	0.48	1,328	820	1,325	536
J. Financial organizations	1,059	1.00	0.88	1,909	1,236	2,030	1,911
K. Real estate and business	3,766	1.00	0.70	1,618	984	1,495	1,085
L. Public administration and defense	2,719	1.00	0.99	1,846	1,255	1,129	882
M. Education	2,326	1.00	0.88	1,601	517	1,422	438
N. Social and health services	3,588	1.00	0.87	1,344	705	931	873
O. Other community services	2,550	1.00	0.39	1,138	628	779	558
P. Domestic services	801	1.00	0.13	918	664	362	686

Notes: This table shows formal and informal sector earnings in 2017 from the GEIH Colombian household survey (*Gran Encuesta Integrada de Hogares*). The sample includes all individuals surveyed between January 2017 and December 2017 who were born between 1980–1986 and whose highest degree is high school, technical college, or university. Panel A presents summary statistics by individuals' highest degree. Panel B displays statistics by workers' industry of employment, defined by the section categories in the third revision of the CIIU economic activity codes (*Clasificación Industrial Internacional Uniforme*).

Column (A) shows the number of surveyed individuals. Column (B) shows the proportion employed at the time of the survey. Column (C) shows the proportion of employed individuals who worked in the formal sector; we define formally-employed workers as those who either: 1) have a written contract; or 2) run a business that is registered with a government agency. This is our best approximation of the definition of formal employment that we use throughout the paper, which is having earnings at a firm that is tracked by the Ministry of Social Protection. Column (D) shows mean monthly earnings for formal sector workers in thousands of Colombian Pesos and column (E) shows mean monthly earnings for informal sector workers. Columns (F)–(G) show the standard deviation of monthly earnings in the formal and informal sectors. All statistics in columns (B)–(G) are computed using survey weights.

Table A5. Mean returns to enrollment in Univalle STEM and other programs using different earnings measures

	(A)	(B)	(C)	(D)
	STEM programs		Other programs	
Dependent variable	Mean below threshold	RD coef	Mean below threshold	RD coef
Panel A. Regressions that exclude individuals with no formal earnings				
Log monthly earnings	14.168	0.133	14.158	−0.047
		(0.061)		(0.051)
Monthly earnings (in 2017 USD)	633.910	80.662	624.063	−33.735
		(41.852)		(32.008)
Monthly earnings/Mean below threshold	1.000	0.127	1.000	−0.054
		(0.066)		(0.051)
N (with earnings defined)		4,845		5,441
Panel B. Regressions that include zeroes for individuals with no formal earnings				
Monthly earnings (in 2017 USD)	447.466	80.141	434.733	−16.769
		(35.902)		(27.774)
Monthly earnings/Mean below threshold	1.000	0.179	1.000	−0.039
		(0.080)		(0.064)
N		6,699		7,664

Notes: This table displays RD estimates of mean returns to enrolling in Univalle's STEM (columns A–B) and other (columns C–D) programs using different earnings measures.

In Panel A, all regressions exclude individuals who do not appear in our formal sector earnings data. The dependent variable in first row is log monthly earnings in 2017, which replicates our benchmark results from Table 3. In the second row, the dependent variable is monthly earnings in levels converted to 2017 U.S. dollars. In the third row, the dependent variable is monthly earnings in levels divided by the control complier means reported in columns (A) and (C).

In Panel B, the dependent variables are the same as those in the second and third rows of Panel A, except we include zeroes for individuals who do not appear in our formal sector earnings data.

Columns (A) and (C) show control complier means estimated following Katz, Kling and Liebman (2001). Columns (B) and (D) display 2SLS RD coefficients, β, from equation (2) using samples of applicants within 30 positions of the admission thresholds.

Parentheses contain standard errors clustered at the individual level.

Table A6. Pre-2000 ICFES subject score weights in admission scores and outcomes

(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)
	STEM programs				Other programs			
ICFES subject (Pre-2000 exam)	Admit score	Grad-uated	Exit score	Log earnings	Admit score	Grad-uated	Exit score	Log earnings
Biology	0.13	−0.38	0.03	−0.03	0.06	−0.09	0.17	0.08
Chemistry	0.10	1.97	0.09	1.01	0.05	1.20	0.01	0.15
Math aptitude	0.22	−0.10	0.22	−0.17	0.13	−0.03	0.46	0.28
Math knowledge	0.19	0.11	0.02	0.55	0.06	−0.54	−0.55	0.11
Physics	0.12	−0.08	0.05	0.46	0.07	0.12	0.27	0.07
Language arts	0.13	−0.37	0.44	0.00	0.34	0.19	0.68	−0.09
Social science	0.11	−0.14	0.15	−0.82	0.29	0.16	−0.04	0.40
Quantitative subjects	0.76	1.51	0.41	1.82	0.38	0.65	0.36	0.69
Qualitative subjects	0.24	−0.51	0.59	−0.82	0.62	0.35	0.64	0.31
Mean absolute deviation from admit score		0.58	0.15	0.60		0.33	0.23	0.16
N (enrollees w/ outcome)	1,007	1,007	302	742	1,042	1,042	372	743

Notes: This table shows how subject scores on the ICFES exam relate to four outcomes: 1) the Univalle admission score; 2) an indicator for graduating from the Univalle program; 3) scores on a field-specific college *exit* exam called *Saber Pro* (formerly *ECAES*); and 4) log monthly earnings in 2017. We regress each outcome variable on the nine ICFES subject scores using all Univalle enrollees in our sample who took the pre-2000 version of the ICFES. (See Table 4 for analogous results using post-2000 ICFES exam takers.) We run these regressions separately for each of Univalle's 48 programs and normalize the estimated coefficients to sum to one. Columns (A)–(D) show the subject weights for each outcome averaged across Univalle's 18 STEM programs. Columns (E)–(H) show the subject weights for each outcome averaged across Univalle's 30 non-STEM programs. We report the sum of the weights for quantitative subjects (biology, chemistry, math, and physics) and qualitative subjects (geography, history, interdisciplinary, language arts, and philosophy). We also report the mean absolute deviation between the average admission score weights (columns A and D) and the average weights for each other outcome (columns B–D and F–H). The last row shows the number of Univalle enrollees for which each outcome is defined.

Table A7. Heterogeneity in returns to Univalle STEM enrollment by academic preparation using different earnings measures

	(A)	(B)	(C)	(D)	(E)
	Less-prepared applicants		More-prepared applicants		
Dependent variable	Mean below threshold	RD coef	Mean below threshold	RD coef	p value diff
Panel A. Regressions that exclude individuals with no formal earnings					
Log monthly earnings	13.992	0.244	14.307	0.032	0.091
		(0.094)		(0.083)	
Monthly earnings (in 2017 USD)	510.877	133.107	725.820	33.935	0.242
		(58.519)		(59.865)	
Monthly earnings/Mean below threshold	1.000	0.261	1.000	0.047	0.134
		(0.115)		(0.082)	
N (with earnings defined)		2,338		2,500	
Panel B. Regressions that include zeroes for individuals with no formal earnings					
Monthly earnings (in 2017 USD)	370.107	107.233	517.938	44.796	0.393
		(49.981)		(52.266)	
Monthly earnings/Mean below threshold	1.000	0.290	1.000	0.086	0.233
		(0.135)		(0.101)	
N		3,306		3,390	

Notes: This table displays RD estimates of the returns to enrolling in Univalle's STEM programs for less-prepared (columns A–B) and more-prepared (columns C–D) applicants using different earnings measures. We define our less- and more-prepared samples in the same way as described in the notes to Table 5.

In Panel A, all regressions exclude individuals who do not appear in our formal sector earnings data. The dependent variable in first row is log monthly earnings in 2017, which replicates our benchmark results from Table 5. In the second row, the dependent variable is monthly earnings in levels converted to 2017 U.S. dollars. In the third row, the dependent variable is monthly earnings in levels divided by the control complier means reported in columns (A) and (C).

In Panel B, the dependent variables are the same as those in the second and third rows of Panel A, except we include zeroes for individuals who do not appear in our formal sector earnings data.

Columns (A) and (C) show control complier means estimated following Katz, Kling and Liebman (2001). Columns (B) and (D) display 2SLS RD coefficients, β, from equation (2) using samples of STEM applicants within 30 positions of the admission thresholds. Column (E) displays the p value from an F test for equality of the RD coefficients in columns (B) and (D).

Parentheses contain standard errors clustered at the individual level.

TABLE A8. Heterogeneity in returns to Univalle enrollment by academic preparation
Applicants to non-STEM programs

	(A)	(B)	(C)	(D)	(E)
	Less-prepared applicants		More-prepared applicants		
Dependent variable	Mean below threshold	2SLS coef	Mean below threshold	2SLS coef	p value diff
Panel A. First stage					
Enrolled in Univalle program	0.077	0.808	0.134	0.765	0.105
		(0.018)		(0.019)	
N		3,773		3,884	
Panel B. 2SLS regressions					
Graduated from Univalle program	−0.000	0.475	0.000	0.519	0.235
		(0.025)		(0.027)	
Employed in formal sector in 2017	0.674	0.038	0.699	0.032	0.899
		(0.036)		(0.035)	
Log monthly earnings in 2017	14.172	−0.093	14.142	−0.026	0.524
		(0.074)		(0.074)	
N (with earnings defined)		2,654		2,771	
Panel C. Log monthly earnings returns with imputed informal earnings					
$\beta^{\text{Informal}} = 0$	13.979	−0.032	13.956	0.014	
$\beta^{\text{Informal}} = 0.133$	13.979	0.006	13.956	0.050	
$\beta^{\text{Informal}} = 0.266$	13.979	0.045	13.956	0.086	
N		3,773		3,884	

Notes: This table displays RD coefficients from separate regressions for less-prepared (columns A–B) and more-prepared (columns C–D) applicants to Univalle's non-STEM programs. We define the less- and more-prepared samples using a leave-cohort-out version of the graduation ICFES score weights from column (F) of Table 4. Specifically, for each program m and cohort t, we regress an indicator for Univalle graduation on the ICFES subject scores in a sample that includes all enrollees in program m in cohorts *other than* t. We take the predicted values from this regression as a measure of the graduation propensity of applicants to program m and t. Lastly, we regress graduation propensity on individuals' admission ranks with application pool dummies and take the residuals from this regression. Less-prepared applicants are those with below median residuals of graduation propensity in their application pool. More-prepared applicants are those with above median residuals.

Columns (A) and (C) present means of each dependent variable for applicants who were just below the admission thresholds. In Panel A, these columns show means over all applicants who were 1–5 positions below the thresholds. In Panels B–C, these columns show control complier means estimated following Katz, Kling and Liebman (2001).

Columns (B) and (D) present RD coefficients using samples of applicants within 30 positions of the admission thresholds. Panel A displays reduced-form RD coefficients, θ, from equation (1). Panel B displays 2SLS RD coefficients, β, from equation (2) using the dependent variable listed in the row header. Panel C displays 2SLS RD coefficients for log monthly earnings in which we impute values for individuals with missing earnings; see the notes to Table 3 for details on this imputation method. We estimate the 2SLS RD specification (1)–(2) using log monthly earnings including these imputed values as the outcome and display the θ coefficients in Panel C.

Column (E) displays the p value from an F test for equality of the RD coefficients in columns (B) and (D).

Parentheses contain standard errors clustered at the individual level.

Table A9. Heterogeneity in returns to Univalle STEM enrollment using alternative measures of academic preparation

	(A)	(B)	(C)	(D)	(E)
	Less-prepared applicants		More-prepared applicants		
Dependent variable	Mean below threshold	2SLS coef	Mean below threshold	2SLS coef	p value diff
Panel A. Graduation propensity in national administrative data					
Enrolled in Univalle program	0.158	0.728	0.141	0.762	0.263
		(0.022)		(0.021)	
Graduated from Univalle program	−0.000	0.267	0.000	0.397	0.001
		(0.029)		(0.028)	
Employed in formal sector in 2017	0.706	0.041	0.715	0.012	0.609
		(0.042)		(0.038)	
Log monthly earnings in 2017	13.959	0.253	14.356	0.004	0.044
		(0.091)		(0.085)	
N		3,307		3,390	
Panel B. Predicted log monthly earnings					
Enrolled in Univalle program	0.138	0.751	0.161	0.736	0.602
		(0.021)		(0.022)	
Graduated from Univalle program	0.000	0.290	0.000	0.393	0.012
		(0.028)		(0.029)	
Employed in formal sector in 2017	0.714	0.004	0.701	0.051	0.410
		(0.041)		(0.039)	
Log monthly earnings in 2017	14.009	0.188	14.316	0.092	0.435
		(0.087)		(0.089)	
N		3,306		3,390	

Notes: This table displays RD coefficients for less- and more-prepared STEM applicants using different measures of academic preparation. The specifications and dependent variables are the same as in Table 5, but we define our less- and more-prepared samples using two different methods. In Panel A, we define graduation propensity using our national Ministry of Education data rather than our sample of Univalle applicants. We regress an indicator for graduating from any college program on the ICFES subject scores in a sample that includes all 1998–2003 ICFES exam takers except for those who appear in our Univalle sample. We estimate this regression separately for each program area in the Ministry's data, and we take the predicted values from this regression as a measure of the graduation propensity for our Univalle sample based on the area of the program the applicant applied to. We then define our less- and more-prepared samples in the same way as in Table 5. In Panel B, we define academic preparation based on predicted earnings rather than graduation propensity within our Univalle sample. The method is the same as in Table 5, except we use log monthly earnings in 2017 rather than an indicator for Univalle graduation to define the two samples.

Columns (A)–(B) show results for less-prepared applicants defined in these two ways, and columns (C)–(D) show results for more-prepared applicants. Columns (A) and (C) present means of each dependent variable for applicants who were just below the admission thresholds. Columns (B) and (D) present RD coefficients using samples of applicants within 30 positions of the admission thresholds. Column (E) displays the p value from an F test for equality of the RD coefficients in columns (B) and (D).

Parentheses contain standard errors clustered at the individual level.

Table A10. Heterogeneity in returns to STEM enrollment by ICFES year and academic preparation

	(A)	(B)	(C)	(D)	(E)	(F)
	All levels of academic preparation		Less-prepared applicants		More-prepared applicants	
Dependent variable	Pre-2000	Post-2000	Pre-2000	Post-2000	Pre-2000	Post-2000
Panel A. Graduation and earnings						
Graduated from Univalle program	0.466	0.323	0.401	0.259	0.529	0.371
	(0.051)	(0.024)	(0.072)	(0.035)	(0.075)	(0.034)
Log monthly earnings in 2017	0.095	0.111	0.335	0.194	−0.124	0.039
	(0.168)	(0.072)	(0.264)	(0.111)	(0.210)	(0.099)
Panel B. Enrollment in college programs						
Enrolled in any STEM BA program	0.559	0.550	0.600	0.603	0.543	0.499
	(0.059)	(0.027)	(0.088)	(0.041)	(0.081)	(0.037)
Enrolled in any BA program	0.320	0.260	0.393	0.302	0.275	0.210
	(0.053)	(0.025)	(0.083)	(0.039)	(0.069)	(0.034)
Enrolled in any college program	0.236	0.142	0.245	0.173	0.246	0.104
	(0.049)	(0.022)	(0.075)	(0.035)	(0.066)	(0.030)
Panel C. Log mean earnings in college program						
Mean earnings in college	0.054	0.026	0.063	0.041	0.040	0.012
	(0.024)	(0.011)	(0.036)	(0.016)	(0.032)	(0.015)
Mean earnings in major	0.082	0.095	0.085	0.131	0.081	0.064
	(0.029)	(0.013)	(0.043)	(0.020)	(0.041)	(0.018)
Mean earnings in college/major	0.197	0.154	0.232	0.187	0.162	0.125
	(0.034)	(0.015)	(0.048)	(0.021)	(0.051)	(0.021)
N	1,062	4,912	521	2,434	541	2,477

Notes: This table displays heterogeneity in the returns to Univalle STEM enrollment by academic preparation and applicants' version of the ICFES exam. The sample, specifications, and dependent variables are the same as in Tables 5–6. In column (A), the sample includes applicants who took the pre-2000 version of the ICFES exam (1998–1999 cohorts). Column (B) includes applicants with post-2000 ICFES scores (2000–2003 cohorts). Columns (C)–(D) include less-prepared applicants with pre- and post-2000 ICFES scores. Columns (E)–(F) include more-prepared applicants with pre- and post-2000 ICFES scores. We define our less- and more-prepared samples as described in the notes to Table 5.

All columns displays 2SLS RD coefficients β from equations (1)–(2). Panel A shows effects of Univalle STEM enrollment on Univalle graduation and log monthly earnings in 2017, as in Panel B of Table 5. Panel B shows effects of Univalle STEM enrollment on college program characteristics using our national higher education census data, as in Panel A of Table 6. Panel C shows effects of Univalle STEM enrollment on log mean earnings in an applicant's college and/or major, as in Panel C of Table 6. Parentheses contain standard errors clustered at the individual level.

Table A11. Heterogeneity in returns to STEM enrollment by gender and academic preparation

	(A)	(B)	(C)	(D)	(E)	(F)
	All levels of academic preparation		Less-prepared applicants		More-prepared applicants	
Dependent variable	Men	Women	Men	Women	Men	Women
Panel A. Graduation and earnings						
Graduated from Univalle program	0.306	0.422	0.232	0.373	0.348	0.469
	(0.025)	(0.034)	(0.037)	(0.051)	(0.036)	(0.050)
Log monthly earnings in 2017	0.173	0.061	0.287	0.141	0.032	0.071
	(0.081)	(0.096)	(0.127)	(0.147)	(0.105)	(0.144)
Panel B. Enrollment in college programs						
Enrolled in any STEM BA program	0.479	0.635	0.540	0.625	0.411	0.646
	(0.029)	(0.037)	(0.044)	(0.055)	(0.039)	(0.050)
Enrolled in any BA program	0.271	0.252	0.304	0.315	0.229	0.187
	(0.027)	(0.035)	(0.042)	(0.053)	(0.035)	(0.048)
Enrolled in any college program	0.168	0.151	0.171	0.203	0.151	0.102
	(0.024)	(0.032)	(0.037)	(0.047)	(0.032)	(0.044)
Panel C. Log mean earnings in college program						
Mean earnings in college	0.034	0.017	0.052	0.040	0.019	−0.001
	(0.012)	(0.014)	(0.017)	(0.022)	(0.016)	(0.020)
Mean earnings in major	0.090	0.074	0.101	0.113	0.078	0.031
	(0.014)	(0.017)	(0.022)	(0.026)	(0.019)	(0.025)
Mean earnings in college/major	0.170	0.135	0.198	0.172	0.149	0.096
	(0.016)	(0.021)	(0.023)	(0.028)	(0.022)	(0.031)
N	4,558	2,132	2,225	1,066	2,329	1,045

Notes: This table displays heterogeneity in the returns to Univalle STEM enrollment by academic preparation and gender. The sample, specifications, and dependent variables are the same as in Tables 5–6. In column (A), the sample includes male applicants, and column (B) includes female applicants. Columns (C)–(D) include less-prepared male and female applicants. Columns (E)–(F) include more-prepared male and female applicants. We define our less- and more-prepared samples as described in the notes to Table 5.

All columns displays 2SLS RD coefficients β from equations (1)–(2). Panel A shows effects of Univalle STEM enrollment on Univalle graduation and log monthly earnings in 2017, as in Panel B of Table 5. Panel B shows effects of Univalle STEM enrollment on college program characteristics using our national higher education census data, as in Panel A of Table 6. Panel C shows effects of Univalle STEM enrollment on log mean earnings in an applicant's college and/or major, as in Panel C of Table 6. Parentheses contain standard errors clustered at the individual level.

Table A12. Heterogeneity in returns to STEM enrollment by program type and academic preparation

	(A)	(B)	(C)	(D)	(E)	(F)
	All levels of academic preparation		Less-prepared applicants		More-prepared applicants	
Dependent variable	Eng.	N. Sci.	Eng.	N. Sci.	Eng.	N. Sci.
Panel A. Graduation and earnings						
Graduated from Univalle program	0.377	0.289	0.308	0.253	0.418	0.316
	(0.026)	(0.032)	(0.037)	(0.047)	(0.036)	(0.044)
Log monthly earnings in 2017	0.190	0.035	0.249	0.236	0.133	−0.142
	(0.080)	(0.094)	(0.122)	(0.151)	(0.106)	(0.131)
Panel B. Enrollment in college programs						
Enrolled in any STEM BA program	0.480	0.623	0.538	0.631	0.418	0.605
	(0.029)	(0.035)	(0.043)	(0.055)	(0.041)	(0.044)
Enrolled in any BA program	0.269	0.269	0.307	0.310	0.219	0.222
	(0.027)	(0.033)	(0.040)	(0.053)	(0.036)	(0.042)
Enrolled in any college program	0.165	0.171	0.200	0.162	0.126	0.163
	(0.024)	(0.030)	(0.037)	(0.046)	(0.032)	(0.040)
Panel C. Log mean earnings in college program						
Mean earnings in college	0.025	0.036	0.044	0.047	0.007	0.021
	(0.012)	(0.013)	(0.017)	(0.020)	(0.018)	(0.016)
Mean earnings in major	0.135	0.006	0.180	−0.010	0.097	0.013
	(0.014)	(0.016)	(0.021)	(0.026)	(0.020)	(0.021)
Mean earnings in college/major	0.181	0.126	0.213	0.142	0.149	0.111
	(0.016)	(0.018)	(0.022)	(0.028)	(0.024)	(0.024)
N	4,385	2,314	2,163	1,143	2,220	1,170

Notes: This table displays heterogeneity in the returns to Univalle STEM enrollment by academic preparation and type of STEM program. The sample, specifications, and dependent variables are the same as in Tables 5–6. In column (A), the sample includes applicants to Univalle's Engineering programs. Column (B) includes applicants to Univalle's Natural Science programs. Columns (C)–(D) include less-prepared applicants to Engineering and Natural Science programs. Columns (E)–(F) include more-prepared applicants to Engineering and Natural Science programs. We define our less- and more-prepared samples as described in the notes to Table 5. See the notes to Table 2 for the Univalle Engineering and Natural Science programs included in our sample.

All columns displays 2SLS RD coefficients β from equations (1)–(2). Panel A shows effects of Univalle STEM enrollment on Univalle graduation and log monthly earnings in 2017, as in Panel B of Table 5. Panel B shows effects of Univalle STEM enrollment on college program characteristics using our national higher education census data, as in Panel A of Table 6. Panel C shows effects of Univalle STEM enrollment on log mean earnings in an applicant's college and/or major, as in Panel C of Table 6. Parentheses contain standard errors clustered at the individual level.

Table A13. Graduation propensity, GPA, and log earnings by year in program for Univalle STEM graduates

Panel A. Relationship between GPA and graduation propensity by year of course

	(A)	(B)	(C)	(D)	(E)
	Dependent variable				
Covariate	Year 1 GPA	Year 2 GPA	Year 3 GPA	Year 4 GPA	Year 5 GPA
Constant	3.010	3.320	3.418	3.244	3.686
	(0.198)	(0.145)	(0.160)	(0.198)	(0.128)
Graduation propensity	1.588	0.731	0.378	0.491	0.439
	(0.448)	(0.337)	(0.353)	(0.453)	(0.282)
N	152	152	152	152	152

Panel B. Relationship between log earnings and GPA by year of course

	(A)	(B)	(C)	(D)	(E)
	Dependent variable				
Covariate	Log earnings	Log earnings	Log earnings	Log earnings	Log earnings
Year 1 GPA	0.456				
	(0.145)				
Year 2 GPA		0.555			
		(0.183)			
Year 3 GPA			0.451		
			(0.185)		
Year 4 GPA				0.376	
				(0.141)	
Year 5 GPA					0.474
					(0.191)
N	121	121	121	121	121

Notes: This table shows the relationship between graduation propensity, Univalle GPA, and log earnings for students who completed a Univalle STEM degree. The sample and definition of Univalle GPA is the same as in Panel B of Figure 3. The sample includes graduates from the 2000 and 2001 cohorts of five Univalle engineering programs for which we have transcript data: Chemical, Electrical, Electronic, Materials, and Mechanical Engineering. To compute GPA, we include only courses that were required for the major and we group courses based on the modal year in the program in which students take them. See the text in Section 4.4 for details on the transcript data and grades at Univalle.

Panel A shows results from regressions of GPA in each year on graduation propensity. Panel B shows results from regressions of log monthly earnings in 2017 on GPA in each year. All regressions include program × enrollment cohort fixed effects. Parentheses contain standard errors clustered at the individual level.

Table A14. Number of students admitted to Univalle programs by cohort

Quota variation	Program	Aug 1999	Jan 2000	Aug 2000	Jan 2001	Aug 2001	Jan 2002	Aug 2002	Jan 2003	Aug 2003	Jan 2004
		Number admitted by semester of application									
Panel A. STEM programs											
Program expansions	Biology	**101**		**99**		**82**	43	**92**	45	53	62
	Systems Eng.	62		82		**126**		61		63	
Tracking admissions	Chemical Eng.	61		**130**		66		43	41	39	36
	Electrical Eng.	56		**127**		57		45	51	49	45
	Electronic Eng.	64		**141**		71		54	44	55	46
	Mechanical Eng.	62		67		**123**		56	50	44	42
Minimal	Other programs (mean)	60	45	62	44	62	50	63	47	63	46
Panel B. Non-STEM programs											
Tracking admissions	Accounting (day)		25	**97**		**194**		**178**		**96**	
	Accounting (night)			**99**		**101**		**95**		**93**	
	Architecture	49	35	**102**		**125**		**100**		**132**	
	Business (day)	51		**106**		**196**		**184**		**100**	
	Business (night)		48	**105**		**103**		**89**		**90**	
	Foreign Trade							54		**92**	
Minimal	Other programs (mean)	38	12	51	39	50	42	48	40	47	60

Notes: This table shows the number of students in our sample who were admitted to Univalle programs in each application cohort. Columns denote the semester of application, which we observe from August (Fall) 1999 to January (Spring) 2004. Panel A includes STEM programs, as depicted in Figure 4. Panel B includes non-STEM programs.

The first six rows in each panel show programs in which the admission quotas changed significantly during this time period. In STEM, this includes two programs with class size expansions (Biology and Systems Engineering) and four programs that used "tracking" admissions (Chemical, Electrical, Electronic, and Mechanical Engineering). In all six non-STEM programs with significant quota variation, the increase in quotas was due to tracking admissions. The last row in each panel shows the mean number of admits for the other programs in our sample without significant quota variation during this time period. See Section 5.1 for details on program expansions and tracking admissions.

Bold numbers are cohorts that we define as having large quotas for our binary measure of L_{mt} (see Section 5.2).

Table A15. Single-step regressions for effects of quota expansions on returns to Univalle STEM enrollment

	(A)	(B)	(C)	(D)
		Effect of quota expansion		
Dependent variable	Mean below threshold	Large quota (binary)	60 extra admits (integer)	Stacked DD (binary)
Panel A. Characteristics of applicants at threshold (DD coefficients)				
Graduation propensity	0.375	−0.092	−0.088	−0.092
		(0.015)	(0.012)	(0.015)
Female	0.311	0.120	0.107	0.122
		(0.064)	(0.051)	(0.066)
College educated mother	0.351	−0.181	−0.104	−0.199
		(0.073)	(0.068)	(0.068)
Family income > 2x min wage	0.622	−0.098	−0.127	−0.099
		(0.075)	(0.053)	(0.073)
N		657	657	1,479
Panel B. Returns to Univalle enrollment (RDDD coefficients)				
Enrolled in Univalle program	0.149	0.153	0.149	0.146
		(0.058)	(0.061)	(0.059)
Graduated from Univalle program	−0.000	−0.125	−0.116	−0.132
		(0.072)	(0.057)	(0.079)
Employed in formal sector in 2017	0.704	0.033	0.041	0.009
		(0.110)	(0.088)	(0.111)
Log monthly earnings in 2017	14.168	0.401	0.211	0.347
		(0.216)	(0.206)	(0.205)
N		6,699	6,699	14,901

Notes: This table shows how the characteristics (Panel A) and outcomes (Panel B) of applicants near the admissions threshold for Univalle's STEM programs changed when the quotas increased. This table is similar to Table 7, except we estimate the DD or RDDD coefficients in a single-step using individual-level observations.

Panel A presents results from difference-in-differences (DD) regressions using a sample of STEM applicants whose admission scores were 1–5 positions below the thresholds. Column (A) shows the mean of each dependent variable, and columns (B)–(D) show π coefficients from equation (3) estimated at the individual-level in this sample.

Panel B presents results from our RD difference-in-differences (RDDD) specification using all STEM applicants whose admission scores were within 30 positions of the thresholds. Column (A) shows control complier means for each dependent variable estimated following Katz, Kling and Liebman (2001). Columns (B)–(D) show π coefficients from a single-step 2SLS RDDD specification, which we derive by plugging equation (3) into our first-step 2SLS specification (1)–(2). Our single-step 2SLS RDDD specification is:

$$E_{ip} = \theta_p D_{ip} + \alpha_p x_{ip} + \psi_p D_{ip} x_{ip} + \gamma_p + \epsilon_{ip} \quad \text{if } |x_{ip}| \leq h$$

$$Y_{ip} = \left(\tilde{\gamma}_m + \tilde{\gamma}_t + \pi L_{mt}\right) E_{ip} + \tilde{\alpha}_p x_{ip} + \tilde{\psi}_p D_{ip} x_{ip} + \tilde{\gamma}_p + \tilde{\epsilon}_{ip} \quad \text{if } |x_{ip}| \leq h.$$

Column (B) reports estimates of π in which the variable of interest, L_{mt}, is an indicator for programs and cohorts with large quotas (the solid symbols in Figure 4). Column (C) reports π coefficients in which we define L_{mt} as the total number of admits in each program/cohort divided by 60 (the y-axis in Figure 4). Column (D) is similar to column (B), but we "stack" our dataset so that the π coefficients are identified *only* by comparing programs with quota expansions to those without expansions. See the notes to Table 7 for details on this stacking procedure.

Regressions are at the individual level. Parentheses contain standard errors clustered at the program/cohort level.

Table A16. Effects of quota expansions on returns to Univalle STEM enrollment using different earnings measures

	(A)	(B)	(C)	(D)
		Effect of quota expansion		
Dependent variable	Control complier mean	Large quota (binary)	60 extra admits (integer)	Stacked DD (binary)
Panel A. Regressions that exclude individuals with no formal earnings				
Log monthly earnings	14.168	0.393	0.183	0.357
		(0.217)	(0.201)	(0.231)
Monthly earnings (in 2017 USD)	633.910	381.382	256.362	375.565
		(136.316)	(133.388)	(144.351)
Monthly earnings/Mean below threshold	1.000	0.602	0.404	0.592
		(0.215)	(0.210)	(0.228)
N (# program/cohorts)		104	104	232
Panel B. Regressions that include zeroes for individuals with no formal earnings				
Monthly earnings (in 2017 USD)	443.903	289.911	201.357	279.311
		(135.022)	(125.451)	(132.799)
Monthly earnings/Mean below threshold	1.000	0.653	0.454	0.629
		(0.304)	(0.283)	(0.299)
N (# program/cohorts)		104	104	232

Notes: This table shows RDDD estimates of how quota increases changed the returns to enrolling in Univalle's STEM programs under different earnings measures.

The row headers describe the earnings measures that we use as dependent variables in our first-step 2SLS RD regressions (equations 1–2). In Panel A, all first-step regressions exclude individuals who do not appear in our formal sector earnings data. The dependent variable in first row is log monthly earnings in 2017, which replicates our benchmark results from Table 7. In the second row, the dependent variable is monthly earnings in levels converted to 2017 U.S. dollars. In the third row, the dependent variable is monthly earnings in levels divided by the control complier mean reported in column (A).

In Panel B, the dependent variables for our first-step regressions are the same as those in the second and third rows of Panel A, except we include zeroes for individuals who do not appear in our formal sector earnings data.

Column (A) shows control complier means for each dependent variable estimated following Katz, Kling and Liebman (2001). Columns (B)–(D) show π coefficients from equation (3) in which the dependent variables are program/cohort-specific RD coefficients, β_{mt}, from our 2SLS specification (1)–(2). Column (B) reports estimates of π in which the variable of interest, L_{mt}, is an indicator for programs and cohorts with large quotas (the solid symbols in Figure 4). Column (C) reports π coefficients in which we define L_{mt} as the total number of admits in each program/cohort divided by 60 (the y-axis in Figure 4). Column (D) is similar to column (B), but we "stack" our dataset so that the π coefficients are identified *only* by comparing programs with quota expansions to those without expansions. See the notes to Table 7 for details on this stacking procedure.

Regressions are at the program/cohort level with observations weighted by the inverse squared standard errors of the RD coefficients. Parentheses contain standard errors clustered at the program/cohort level.

Table A17. Effects of quota expansions on returns to non-STEM enrollment

	(A)	(B)	(C)	(D)
		Effect of quota expansion		
Dependent variable	Control complier mean	Large quota (binary)	60 extra admits (integer)	Stacked DD (binary)
Panel A. Characteristics of marginally-admitted compliers (DD coefficients)				
Graduation propensity	0.505	−0.106	−0.038	−0.106
		(0.051)	(0.026)	(0.051)
Female	0.592	−0.233	−0.063	−0.230
		(0.198)	(0.073)	(0.200)
College educated mother	0.346	−0.188	−0.023	−0.193
		(0.110)	(0.057)	(0.114)
Family income > 2x min wage	0.554	0.008	0.080	0.002
		(0.130)	(0.067)	(0.129)
N (# program/cohorts)		130	130	239
Panel B. Returns to Univalle enrollment (RDDD coefficients)				
Enrolled in Univalle program	0.106	−0.049	−0.030	−0.046
		(0.146)	(0.056)	(0.144)
Graduated from Univalle program	−0.000	−0.128	−0.072	−0.130
		(0.125)	(0.080)	(0.126)
Employed in formal sector in 2017	0.690	−0.074	0.012	−0.085
		(0.200)	(0.101)	(0.208)
Log monthly earnings in 2017	14.158	−0.375	−0.054	−0.398
		(0.325)	(0.283)	(0.319)
N (# program/cohorts)		130	130	239

Notes: This table shows how the characteristics (Panel A) and outcomes (Panel B) of applicants near the admissions threshold for Univalle's non-STEM programs changed when the quotas increased. This table is similar to Table 7, except the sample includes non-STEM applicants.

In both panels, column (A) shows control complier means for each dependent variable estimated following Katz, Kling and Liebman (2001). In Panel A, columns (B)–(D) show π coefficients from equation (3) in which the dependent variables are program/cohort-specific mean complier characteristics. In Panel B, columns (B)–(D) show π coefficients from equation (3) in which the dependent variables are program/cohort-specific RD coefficients, β_{mt}, from our 2SLS specification (1)–(2).

Column (B) reports estimates of π in which the variable of interest, L_{mt}, is an indicator for programs and cohorts with large quotas (the bold numbers in Appendix Table A14). Column (C) reports π coefficients in which we define L_{mt} as the total number of admits in each program/cohort divided by 60 (the number values in Appendix Table A14). Column (D) is similar to column (B), but we "stack" our dataset so that the π coefficients are identified *only* by comparing programs with quota expansions to those without expansions. We combine the six "treated" non-STEM programs into two groups based on the cohort(s) in which their quotas expanded: 1) Accounting, Architecture, and Business (Fall 2000–2003); and 2) Foreign Trade (Fall 2003). We then create two datasets that include all 24 "control" non-STEM programs plus the treated programs in each group. Lastly, we stack these datasets and estimate the DD or RDDD specification with all covariates (except L_{mt}) interacted with dummies for each dataset.

Regressions are at the program/cohort level with observations weighted by the inverse squared standard errors of the means (Panel A) and RD coefficients (Panel B). Parentheses contain standard errors clustered at the program/cohort level.

Table A18. Effects of STEM quota expansions on college program and degree characteristics

	(A)	(B)	(C)	(D)	(E)
		Effect on mean below threshold (DD coefficients)		Effect on returns to STEM enrollment (RDDD coefficients)	
Dependent variable	Mean below threshold	Large quota (binary)	60 extra admits (integer)	Large quota (binary)	60 extra admits (integer)
Panel A. Enrollment in college programs					
Enrolled in any STEM BA program	0.564	−0.177	−0.150	0.241	0.248
		(0.079)	(0.073)	(0.070)	(0.068)
Enrolled in any BA program	0.766	−0.161	−0.143	0.175	0.174
		(0.068)	(0.058)	(0.091)	(0.067)
Enrolled in any technical program	0.202	0.051	0.029	0.041	0.090
		(0.067)	(0.055)	(0.098)	(0.076)
Enrolled in any college program	0.838	−0.066	−0.069	0.092	0.123
		(0.069)	(0.059)	(0.082)	(0.054)
N (# program/cohorts)		104	104	104	104
Panel B. Log mean earnings in college program					
Mean earnings in college	14.081	−0.067	−0.048	0.045	0.040
		(0.029)	(0.028)	(0.038)	(0.032)
Mean earnings in major	14.122	−0.023	−0.024	0.010	0.038
		(0.045)	(0.039)	(0.050)	(0.040)
Mean earnings in college/major	14.131	−0.099	−0.084	0.053	0.053
		(0.049)	(0.039)	(0.046)	(0.044)
N (# program/cohorts)		104	104	104	104

Notes: This table shows how the college enrollment outcomes of applicants near the admissions threshold for Univalle's STEM programs changed when the quotas increased. Panel A shows effects on college program characteristics using our national higher education census data, as in Panel A of Table 6. Panel B shows effects on log mean earnings in an applicant's college and/or major, as in Panel C of Table 6.

Column (A) shows the mean of each dependent variable for STEM applicants 1–5 positions below the thresholds. Columns (B)–(E) show π coefficients from equation (3). In columns (B)–(C), the dependent variables are the mean outcomes of STEM applicants whose admission scores were 1–5 positions below the thresholds in each program/cohort (as in Panel A of Table 7). In columns (D)–(E), the dependent variables are program/cohort-specific RD coefficients, β_{mt}, from our 2SLS specification (1)–(2) estimated in a sample of all STEM applicants whose admission scores were within 30 positions of the thresholds (as in Panel B of Table 7). Columns (B) and (D) report estimates in which the variable of interest, L_{mt}, is an indicator for programs and cohorts with large quotas (the solid symbols in Figure 4). Columns (C) and (E) report estimates in which we define L_{mt} as the total number of admits in each program/cohort divided by 60 (the y-axis in Figure 4).

Regressions are at the program/cohort level with observations weighted by the number of observations (columns B–C) and the inverse squared standard errors of the RD coefficients (columns D–E). Parentheses contain standard errors clustered at the program/cohort level.

Table A19. Effects of quota expansions on returns for top Univalle STEM enrollees

	(A)	(B)	(C)	(D)
		Effect of quota expansion		
Dependent variable	Mean in small cohorts	Large quota (binary)	60 extra admits (integer)	Stacked DD (binary)
Panel A. Characteristics of top enrollees (DD coefficients)				
Graduation propensity	0.403	0.018	0.012	0.022
		(0.007)	(0.005)	(0.006)
Female	0.199	−0.004	−0.020	−0.008
		(0.022)	(0.019)	(0.028)
College educated mother	0.409	0.014	−0.000	0.014
		(0.045)	(0.039)	(0.042)
Family income > 2x min wage	0.620	−0.015	−0.030	−0.016
		(0.041)	(0.035)	(0.044)
N (# program/cohorts)		106	106	238
Panel B. Returns to Univalle enrollment (DD coefficients)				
Graduated from Univalle program	0.364	0.071	0.024	0.078
		(0.045)	(0.030)	(0.050)
Employed in formal sector in 2017	0.772	0.043	0.046	0.036
		(0.046)	(0.040)	(0.046)
Log monthly earnings in 2017	14.363	−0.004	0.009	0.016
		(0.076)	(0.060)	(0.083)
N (# program/cohorts)		106	106	238

Notes: This table shows how the characteristics (Panel A) and outcomes (Panel B) of highly-ranked Univalle STEM enrollees changed when the quotas increased. The specifications and outcome variables are similar to those in Table 7, but we use a sample of "top enrollees" in Univalle's STEM programs. To define this sample, we first compute the minimum rank of a student who was admitted and enrolled in each Univalle program and cohort. We then compute the *maximum* of these minimum ranks across all cohorts for each program. Our top enrollee sample includes all students who enrolled in the Univalle STEM program to which they applied and whose rank was higher than this maximum rank. Thus this sample contains students whose admission ranks were high enough such that they could have enrolled in *any* cohort of their program, regardless of the quota size.

The dependent variables are the mean characteristics (Panel A) and outcomes (Panel B) of top enrollees in each program/cohort. Column (A) shows the mean of each dependent variable, and columns (B)–(D) show π coefficients from equation (3). Column (B) reports estimates of π in which the variable of interest, L_{mt}, is an indicator for programs and cohorts with large quotas (the solid symbols in Figure 4). Column (C) reports π coefficients in which we define L_{mt} as the total number of admits in each program/cohort divided by 60 (the y-axis in Figure 4). Column (D) is similar to column (B), but we "stack" our dataset so that the π coefficients are identified *only* by comparing programs with quota expansions to those without expansions. See the notes to Table 7 for details on this stacked specification.

Regressions are at the program/cohort level with observations weighted by the inverse squared standard errors of the means. Parentheses contain standard errors clustered at the program/cohort level.

B. Theoretical appendix

This section presents a framework that illustrates the mechanisms through which the returns to enrolling in a selective STEM program can vary with a student's academic preparation.

We consider a population of high school graduates indexed by i with pre-college academic preparation α_i. Students can choose from a large number of college programs $p \in P$, where programs are defined by both an institution and a field of study. The set P also includes the option of not enrolling in college at all, which we denote by $p = 0$. For simplicity, our framework assumes that academic preparation, α_i, is unidimensional. In our empirical analysis, we allow individuals to have different levels of preparation for different college programs p.

We define the following potential outcomes that describe an individual's returns to enrolling in each program:

- Let v^e_{ip} represent individual i's potential skill value added from *enrolling* in program p. This term reflects, for example, the skills an individual learns in first-year courses.
- Let g_{ip} denote individual i's potential graduation outcome in program p. In other words, $g_{ip} = 1$ for individuals who would successfully complete the program if they enrolled and $g_{ip} = 0$ for individuals who would drop out.
- Let v^g_{ip} represent the additional skill that individual i would gain if they *graduate* from program p.

We assume $v^e_{ip} \geq 0$ and $v^g_{ip} \geq 0$ for all p and that $v^e_{i0} = v^g_{i0} = 0$ for the option of not attending college. Importantly, each of these three potential outcomes can depend on an individual's academic preparation, α_i.

After college, individuals enter a competitive labor market and earn a wage equal to their skill. Under the above assumptions, individual i's potential log wage from enrolling in program p is given by:

$$w_{ip} = \alpha_i + v^e_{ip} + g_{ip} v^g_{ip}. \tag{B1}$$

An individual's wage is equal to $\alpha_i + v^e_{ip} + v^g_{ip}$ if they complete program p and it is equal to $\alpha_i + v^e_{ip}$ if they drop out of the program.

Our empirical estimates pertain to a population of "compliers" for a selective STEM program that we denote by s. By "compliers," we mean a group of students who would enroll in program s if and only if they are offered admission. If these students are not admitted, they enroll in their next-choice program that we denote by $c(i) \in P$. Next-choice programs can vary across individuals in the complier group, and they may differ from program s in institution and/or field of study.

We begin by examining the average wage returns to *enrolling* in Univalle's STEM program in Section 3. We denote this return by $E[w_{is} - w_{i,c(i)}]$, where the expectation is defined over all compliers who are close to Univalle's admission threshold. Using the wage equation (B1) and that fact that g_{ip} is binary, this return is given by:

$$E[w_{is} - w_{i,c(i)}] = E[v_{is}^e - v_{i,c(i)}^e] + \Big\{E[v_{is}^g | g_{is} = 1] - E[v_{i,c(i)}^g | g_{i,c(i)} = 1]\Big\} E[g_{is}]$$
$$+ E[v_{i,c(i)}^g | g_{i,c(i)} = 1] E[g_{is} - g_{i,c(i)}] \qquad \text{(B2)}$$

Our results in Sections 4–5 show how the returns to enrolling in a Univalle STEM program vary with academic preparation. In notation this estimand is $dE[w_{is} - w_{i,c(i)} | \alpha_i = \alpha]/d\alpha$—the change in the mean wage return to program s from an increase in academic preparation, α. Using equation (B2) and letting $E_\alpha[x] \equiv E[x|\alpha_i = \alpha]$ denote the expected value of a variable x conditional on academic preparation level $\alpha_i = \alpha$, this term is given by:

(B3)

$$\frac{dE_\alpha[w_{is} - w_{i,c(i)}]}{d\alpha} = \underbrace{\frac{dE_\alpha[v_{is}^e - v_{i,c(i)}^e]}{d\alpha}}_{\text{Term 1}} + \underbrace{\frac{dE_\alpha[v_{is}^g | g_{is} = 1]}{d\alpha} E_\alpha[g_{is}]}_{\text{Term 2}} - \underbrace{\frac{dE_\alpha[v_{i,c(i)}^g | g_{i,c(i)} = 1]}{d\alpha} E_\alpha[g_{i,c(i)}]}_{\text{Term 3}}$$
$$+ \underbrace{\Big\{E_\alpha[v_{is}^g | g_{is} = 1] - E_\alpha[v_{i,c(i)}^g | g_{i,c(i)} = 1]\Big\} \frac{dE_\alpha[g_{is}]}{d\alpha}}_{\text{Term 4}} + \underbrace{E_\alpha[v_{i,c(i)}^g | g_{i,c(i)} = 1] \frac{dE_\alpha[g_{is} - g_{i,c(i)}]}{d\alpha}}_{\text{Term 5}}$$

Our RD analysis of the returns to Univalle's STEM programs yields three main results. First, there is a positive mean earnings return to enrolling in these STEM programs for marginal admits (Table 3). Second, STEM graduation rates at Univalle increase with academic preparation (Tables 5 and 7), while the effect of enrollment on the probability of earning any college degree does not differ significantly by academic preparation (Table 6). Third, mean earnings returns to enrolling in these STEM programs *decrease* with academic preparation (Tables 5 and 7).

These results lead us to explore the mechanisms through which less-prepared students can have larger earnings returns to selective STEM programs. All else equal, earnings returns increase with the probability of graduating, but there are three reasons why returns can be larger for less-prepared students despite lower graduation rates. We summarize these three mechanisms in the following proposition.

Proposition. *Suppose that:*

(i) The skill return to graduating from program s is non-negative for all levels of academic preparation,

$$E_\alpha[v_{is}^g | g_{is} = 1] - E_\alpha[v_{i,c(i)}^g | g_{i,c(i)} = 1] \geq 0;$$

(ii) Graduation rates in program s are increasing in academic preparation,

$$\frac{dE_\alpha[g_{is}]}{d\alpha} > 0;$$

(iii) Relative graduation rates between program s and next-choice programs are unrelated to academic preparation,

$$\frac{dE_\alpha[g_{is} - g_{i,c(i)}]}{d\alpha} = 0.$$

Then if the wage return to enrolling in program s is decreasing in academic preparation, $dE_\alpha[w_{is} - w_{i,c(i)}]/d\alpha < 0$*, at least one of the following conditions must hold:*

(a) There is a skill return to enrolling in program s that decreases with academic preparation,

$$\frac{dE_\alpha[v^e_{is} - v^e_{i,c(i)}]}{d\alpha} < 0;$$

(b) Less-prepared students choose counterfactual programs with less degree value added,

$$\frac{dE_\alpha[v^g_{i,c(i)}|g_{i,c(i)} = 1]}{d\alpha} > 0;$$

(c) Less-prepared students have greater value added to a degree from program s,

$$\frac{dE_\alpha[v^g_{is}|g_{is} = 1]}{d\alpha} < 0.$$

This proposition follows from inspection of equation (B3). Conditions (i) and (ii) ensure that Terms 4 and 5 are non-negative. Mechanisms (a)–(c) determine the sign of Terms 1–3 since $E_\alpha[g_{is}] \geq 0$ and $E_\alpha[g_{i,c(i)}] \geq 0$.

We explore the empirical evidence on these three mechanisms in Sections 4.3–4.5 and in Section 5.5.

C. Empirical appendix

C.1. **Data and merging.** This section provides details on our data sources and merging.

Our base dataset includes lists of all applicants to Universidad del Valle's undergraduate programs from Fall 1999 to Spring 2004 (Univalle, 2017). These data were provided by Univalle, and they include the program/cohort that applicants applied to, their admission scores, and their admission decisions.

We combine the Univalle application records with three individual-level administrative datasets provided by the Colombian government. The first dataset includes records from Colombia's national standardized college entrance exam, which was formerly called the ICFES exam and is now called *Saber 11* (ICFES, 2013*a*). The data were provided by the agency that administers the exam and it contains all students who took the exam between 1998–2003. The ICFES exam is also used by the Colombian government for high school accountability, so it is taken by nearly every high school graduate in the country. The main variables of interest are individuals' scores on each exam subject and demographic characteristics.

The second administrative dataset includes enrollment and graduation records from the Ministry of Education (SPADIES, 2013). These records include the institution, program of study, and graduation outcome for students who enrolled in college between 1998–2012. The Ministry's records cover almost all colleges in Colombia, although it omits a few schools due to their small size or inconsistent reporting. To describe the set of colleges that are included in the Ministry of Education records, we use another administrative dataset from a college exit exam called *Saber Pro* (ICFES, 2013*b*). This national exit exam is administered by the same agency that runs the ICFES college admission exam and it became a requirement for graduation from any higher education institution in 2009. Column (A) in Table C1 depicts the 310 colleges that have any exit exam takers in these administrative records in 2009–2011. These colleges are categorized into the Ministry of Education's five types of higher education institutions, which are listed in descending order of their on-time program duration.[38] Column (B) shows the number of exit exam takers per year. The majority of exam takers are from university-level institutions, with fewer students from technical colleges. Column (C) shows the fraction of these 310 colleges that appear in the Ministry of Education records that we use in our analysis. These proportions are weighted by the number of exam takers depicted in column (B). Column (C) shows that the Ministry of Education records

[38] Most programs at universities require 4–5 years of study, while programs at Technical/Professional Institutes typically take 2–3 years.

TABLE C1. Higher education institutions in the Ministry of Education records

	(A)	(B)	(C)
	Number of colleges	Number of exit exam takers/year	Prop. of colleges in records
University	122	134,496	1.00
University Institute	103	53,338	0.88
Technology School	3	2,041	1.00
Technology Institute	47	15,092	0.82
Technical/Professional Institute	35	11,408	0.99
Total	310	216,375	0.96

Notes: Column (A) depicts the number of colleges that have *Saber Pro* exit exam takers in 2009–2011 using administrative records from the testing agency. Colleges are categorized into the Ministry of Education's five higher education institution types. Column (B) shows the number of 2009–2011 exam takers per year. Column (C) shows the proportion of colleges that appear in the Ministry of Education records, where colleges are weighted by the number of exit exam takers.

include all universities but are missing a few technical colleges.[39] Overall, 96 percent of exit exam takers attend colleges that appear in the Ministry of Education records.

The third administrative dataset includes earnings records collected by the Ministry of Social Protection (PILA, 2019). The records are from the Ministry's electronic tax record system called *Planilla Integrada de Liquidación de Aportes* (PILA). Our data include monthly earnings in 2017 for any individual who worked at a firm that was registered with the Ministry. Our main income measure is average monthly earnings, which we compute by dividing total annual earnings by the number of employment months in 2017. We also use an indicator for appearing in the PILA dataset as a measure of formal employment.

We merge the Univalle application data into the ICFES data using applicants' full names. Since the ICFES exam is required for admission to Univalle, most applicants appear in the ICFES administrative dataset. Most individuals match uniquely on name, but in cases with duplicate names we use information on ICFES exam cohort and high school location to identify the correct match.[40] Through this process, we are able to match 84 percent of individuals in the Univalle application data to the ICFES records, as shown in columns (A)–(B) in Table C4 below. The vast majority of non-matches occur because individuals took the ICFES exam prior to 1998, when our records begin.[41]

We merge the ICFES and Ministry of Education datasets using individuals' national ID numbers, birth dates, and names. We define a match from this merge as observations that

[39] The largest omitted institutions are the national police academy (*Dirección Nacional de Escuelas*) and the Ministry of Labor's national training service (*Servicio Nacional de Aprendizaje*).

[40] If there are duplicates, we select the individual who took the ICFES exam prior to Univalle application and who attended a high school in the Valle del Cauca region. If these criteria do not identify a unique ICFES exam taker, we consider the applicant to be a non-match.

[41] Many Colombians wait a year or more after high school before applying to college.

have either: 1) the same ID number and a fuzzy name match; 2) the same birth date and a fuzzy name match; or 3) an exact name match for a name that is unique in both records.[42] 39 percent of the 1998–2003 ICFES exam takers appear in the Ministry of Education records, which is comparable to the higher education enrollment rate in Colombia during the same time period.[43] A better indicator of merge success is the percentage of college enrollees that appear in the admission exam records because all domestic college students must take the exam. We match 91 percent of enrollees who took the admission exam between 1998 and 2003.[44]

Lastly, the combined dataset from the above merges was matched to the PILA earnings records by the Colombian statistical agency *Departamento Administrativo Nacional de Estadística* (DANE). DANE also merged these datasets using national ID numbers, names, and birth dates. The fraction of individuals in the 1998–2003 ICFES exam cohorts who were matched to the 2017 earnings dataset is 56 percent. To benchmark this merge rate, we use Colombian household survey data (GEIH) on individuals in the 1981–1987 birth cohorts with at least a high school degree (GEIH, 2019). In this population, the fraction of individuals who worked and had a contract for their employment was also 56 percent in 2017. This suggests that the DANE merge identified nearly all individuals in our sample with formal sector jobs.

C.2. **Analysis sample.** This section provides details on the sample we use for our analysis.

Our sample includes all of Univalle's bachelor's degree programs where we can identify the effects of admission. Our initial dataset includes applicants to 74 different degree programs from Fall 1999 to Spring 2004. We exclude 26 of these programs from our sample for one of two reasons, as shown in Table C2. First, we exclude technical/professional programs to focus on bachelor's degree attainment (column C). Second, we exclude programs with fewer than two cohorts in which any applicant was rejected (column E), which is necessary for our RD difference-in-differences design. Excluded programs tend to attract fewer applicants and were offered only a few times during our data period. Our sample includes the remaining 48 degree programs listed in Appendix Table C3.

[42] Nearly all students in these records have national ID numbers, but Colombians change ID numbers around age 17. Most students in the admission exam records have below-17 ID numbers (*tarjeta*), while most students in the college enrollment and earnings records have above-17 ID numbers (*cédula*). Merging using ID numbers alone would therefore lose a large majority of students.

[43] The gross tertiary enrollment rate ranged from 23 percent to 28 percent between 1998 and 2003 (World Bank World Development Indicators, available at: http://data.worldbank.org/country/colombia). This rate is not directly comparable to our merge rate because not all high school aged Colombians take the ICFES exam. About 70 percent of the secondary school aged population was enrolled in high school in this period. Dividing the tertiary enrollment ratio by the secondary enrollment ratio gives a number roughly comparable to our 39 percent merge rate.

[44] Approximately 16 percent of students in the Ministry of Education records have missing birth dates, which accounts for most of the non-matches.

TABLE C2. Programs excluded from sample

(A)		(B)	(C)	(D)	(E)	(F)
Faculty area	#	Program	Degree level	Application cohorts	Cohorts with rejects	Total applied
Engineering	1	Environmental Management	Technical	4	2	538
	2	Food Science	Technical	1	1	195
	3	Forest Protection	Technical	1	0	12
	4	Information Systems	Technical	1	1	201
	5	Soil and Water Conservation	Technical	4	2	253
Health	6	Prehospital Care	Technical	4	2	3,014
Humanities	7	Geography	Bachelor's	2	1	97
	8	Philosophy	Professional	2	2	106
	9	Physical Education	Professional	2	1	316
	10	Political Studies	Bachelor's	3	1	337
	11	Recreation (night)	Bachelor's	2	1	112
	12	Teaching (Biology & Chemistry)	Bachelor's	2	0	44
	13	Teaching (Elem. Math, day)	Bachelor's	1	0	34
	14	Teaching (Elem. Math, mixed)	Bachelor's	1	0	30
	15	Teaching (Elem. N. Science, day)	Bachelor's	1	1	138
	16	Teaching (Elem. N. Science, mixed)	Bachelor's	1	0	13
	17	Teaching (Math & Physics, day)	Bachelor's	1	1	65
	18	Teaching (Math & Physics, mixed)	Bachelor's	1	0	18
	19	Teaching (Modern Languages, day)	Bachelor's	1	1	39
	20	Teaching (Modern Languages, night)	Bachelor's	1	0	37
	21	Teaching (Phys. Ed. & Health)	Bachelor's	2	1	111
	22	Teaching (Physical Education, day)	Bachelor's	1	1	55
	23	Teaching (Physical Education, mixed)	Bachelor's	2	0	43
	24	Teaching (Physical Math)	Bachelor's	1	0	23
	25	Teaching (Popular Education)	Bachelor's	1	1	45
Integrated arts	26	Music	Bachelor's	1	1	110
Total			Bachelor's	44	21	5,986

Notes: Columns (A)–(B) list the Univalle programs that we exclude from our sample and their faculty areas at the university. Column (C) reports the program's degree level (technical, professional, or bachelor's). Column (D) shows the total number of application cohorts from August 1999 to January 2004. Column (E) shows the number of cohorts during this period in which any applicant was rejected. Column (F) shows the total number of applicants during this period.

Table C4 shows the applicants to these 48 programs that we include in our sample. Column (A) shows that our initial dataset includes 20,001 applicants to the STEM programs in our sample (Panel A) and 29,041 applicants to other programs (Panel B). We exclude applicants for the three reasons shown in columns (B)–(D) of Table C4. First, we drop applicants who do not appear in our ICFES dataset (column B), as described in Section C.1. Second, we exclude applicants in special disadvantaged admission groups who were not subject to Univalle's primary admission thresholds (column C). During this time period, Univalle maintained special admission quotas for disabled, indigenous, and military applicants. Third, we drop applicants from cohorts where no applicants were rejected (column D), which is necessary

Table C3. Programs included in sample

Group	(A) Faculty area	#	(B) Program	(C) Application cohorts	(D) Total applied	(E) Main RD sample
STEM	Engineering	1	Agricultural Engineering	6	532	313
		2	Chemical Engineering	7	1,220	478
		3	Civil Engineering	7	590	405
		4	Electrical Engineering	7	717	380
		5	Electronic Engineering	7	1,027	407
		6	Industrial Engineering	7	1,183	423
		7	Materials Engineering	6	857	379
		8	Mechanical Engineering	7	849	443
		9	Sanitary Engineering	4	541	274
		10	Statistics	5	627	254
		11	Systems Engineering	5	1,758	323
		12	Topographical Engineering	6	517	306
	N. sciences	13	Biology	8	2,021	567
		14	Chemical Technology (day)	3	883	238
		15	Chemical Technology (night)	3	295	212
		16	Chemistry	7	1,073	473
		17	Math	3	481	259
		18	Physics	9	851	565
Other	Administration	19	Accounting (day)	5	845	250
		20	Accounting (night)	4	758	274
		21	Business (day)	5	1,065	275
		22	Business (night)	5	770	299
		23	Foreign Trade	2	359	107
	Health	24	Audiology	5	579	294
		25	Bacteriology	5	1,657	301
		26	Dentistry	5	818	286
		27	Medicine	5	2,551	327
		28	Nursing	5	1,149	261
		29	Occupational Therapy	5	889	286
		30	Physical Therapy	5	1,742	297
	Humanities	31	History	4	531	190
		32	Recreation	2	228	123
		33	Social Work	4	1,016	233
		34	Teaching (Elem. S. Science)	2	154	108
		35	Teaching (Foreign Lang., day)	2	188	114
		36	Teaching (Foreign Lang., night)	2	107	93
		37	Teaching (History)	4	596	213
		38	Teaching (Literature)	4	588	260
		39	Teaching (Philosophy)	4	411	261
		40	Teaching (Social Science)	3	171	139
	Integrated arts	41	Architecture	6	1,346	311
		42	Communication	5	356	268
		43	Dramatic Arts	9	363	336
		44	Teaching (Music)	5	571	332
		45	Visual Arts	5	423	280
	S. sciences	46	Economics	9	983	585
		47	Psychology	5	1,264	323
		48	Sociology	4	961	238
Total				242	39,461	14,363

Notes: Columns (A)–(B) list each Univalle program in our sample and its faculty area (see Section 2.2). Column (C) shows the total number of application cohorts from August 1999 to January 2004. Column (D) reports the total number of applicants in our sample and column (E) shows the number of applicants within 30 positions of the admission thresholds.

Table C4. Analysis sample

	(A)	(B)	(C)	(D)	(E)	(F)
		Excluded applicants				
	All applicants	Missing ICFES scores	Special admission group	No rejected applicants	Full sample	RD sample
Panel A. STEM applicants						
Ability percentile	0.783		0.811	0.838	0.780	0.842
Age	18.713		19.544	18.997	18.686	18.947
College educated father	0.426		0.355	0.454	0.427	0.440
College educated mother	0.361		0.330	0.358	0.361	0.373
Family income > 2x min wage	0.576		0.470	0.612	0.576	0.599
Female	0.357		0.274	0.312	0.360	0.319
N	20,001	3,077	310	592	16,022	6,699
Panel B. Other applicants						
Ability percentile	0.735		0.778	0.846	0.733	0.810
Age	18.923		19.735	20.596	18.879	19.353
College educated father	0.408		0.431	0.370	0.408	0.424
College educated mother	0.344		0.375	0.300	0.344	0.354
Family income > 2x min wage	0.560		0.498	0.609	0.560	0.589
Female	0.637		0.539	0.541	0.641	0.588
N	29,041	4,746	462	394	23,439	7,664

Notes: Column (A) shows the total number of applicants to the 48 Univalle programs in our sample (see Appendix Table C3). Column (B) shows the number of applicants who do not appear in the ICFES dataset. Column (C) lists the number of students who were admitted through special quotas for disadvantaged groups. Column (D) shows the number of applicants to program/cohort pairs in which no applicants were rejected. Column (E) shows our full analysis sample, which is equal to column (A) minus the applicants in columns (B)–(D). Column (F) shows the subset of applicants from column (E) who are within 30 positions of the admission threshold in their application pool.

Panel A includes applicants to Univalle's STEM programs and Panel B includes applicants to non-STEM programs. Demographic characteristics are not reported in column (B) because these variables come from the ICFES dataset.

for our RD strategy. After these restrictions, our sample includes 16,022 STEM applicants and 23,439 applicants to other programs.

Most of our regressions focus on the subset of applicants whose admission scores are within h ranks of the tracking threshold. Our benchmark model uses $h = 30$, which is roughly the mean of the Calonico, Cattaneo and Titiunik (2014) bandwidths across all dependent variables. Column (F) shows that this RD sample includes 6,699 STEM applicants and 7,664 applicants to other programs. Applicants in our RD sample tend to have higher pre-college ability than those in the full sample. In addition, these applicants come from slightly more advantaged socioeconomic backgrounds, and are less likely to identify as female.