

DATA APPENDIX

In this article, we draw upon several data sources, including student-level surveys from the National Center of Education Statistics (NCES), institution-level data from the Integrated Postsecondary Educational Database System (IPEDS) and the Annual Survey of Colleges (ASC), and other data on college entrance examinations and degree attainment. In this data appendix, we describe these datasets in more detail and discuss issues of sample restrictions, dataset comparability, and important variable definitions.

Institutional Data: American Survey of Colleges

The American Survey of Colleges (ASC) is a near-census of four-year postsecondary educational institutions in the United States, conducted by the College Board. Containing detailed data on institutional classification, enrollment, applications, student body profiles, expenditures, and sources of funding, the ASC is ideal for examining trends in characteristics among different types of institutions. Our data run from 1986 through 2003. Missing data prevented us from using more recent years. Summary statistics for our sample are available in Appendix Table 1.

College Ranking Measure

In constructing our selectivity measures, we generally follow the methodology of Bound, Lovenheim, and Turner (2007), who in turn employ the rankings from the 2005 edition of *U.S. News and World Report's Best Colleges*. Selective private schools include the top 50 private research universities as well as the top 49 liberal arts colleges. Selective public schools include the top 47 public universities. These schools are listed in Appendix Table 2.

Student Data: National Center of Education Statistics Longitudinal Student Datasets

The National Center of Education Statistics (NCES) has conducted four nationally-representative, large-scale, longitudinal surveys of secondary students since 1972. Each of these surveys originally sampled between 12,000 and 16,000 students in a given grade cohort, with follow-up survey waves over the next several years. Designed to shed light on the school-to-work transition, the surveys ask questions about demographic background, school experiences, education and work expectations, and labor market outcomes. Additionally, each survey cohort was administered a cognitive test battery. In many cases, the data variables are directly comparable across the four different surveys.

These data allow us to construct measures of high school behavior, such as course-taking, test preparation, extracurriculars, and time spent on homework. Using the restricted-access versions of these datasets, we can also identify the top two choices of schools to which students applied and whether they were accepted. We condition our sample using inverse probability sampling weights to be nationally representative of the relevant high school senior classes.¹ Summary statistics for each cohort are provided in Appendix Table 4.

- *National Longitudinal Study of the High School Class of 1972 (NLS)*

The first NCES student panel study, the NLS surveyed approximately 17,000 high school seniors in the spring of 1972, with follow-up waves in 1973, 1974, 1979, and 1986. We use

¹ Because the NCES surveys initially sample the cohorts at different ages, we use the provided flags and probability weights that are meant to represent the universe of U.S. high school seniors. These flags and weights are described below.

the initial and 1973 waves.² The weight variable is W4, and the universe is already high school seniors, so no additional sample flag is necessary.

- *High School and Beyond (HSB)*
The second longitudinal study originally surveyed both 10th and 12th graders in 1980 (approximately 14,000 of each), with follow-ups in 1982, 1984, and 1986. Due to the richer high school data available for the sophomore cohort, we restrict our sample from the HSB to the sophomore cohort and use the 1982 and 1984 waves. The weight variable used is FU2WT, and the conditioning flags are FU1PART and FU2PART.
- *National Educational Longitudinal Survey (NELS)*
The third NCES panel study surveyed approximately 25,000 8th graders in 1988, with follow-up waves in 1990, 1992, 1994, and 2000. Using the 1992 and 1994 waves, the weight variable F3QWTG12, and the conditioning flag G12COHRT, our NELS sample is representative of the high school senior class of 1992.
- *Educational Longitudinal Survey (ELS)*
The fourth and newest longitudinal student survey, the ELS, initially surveyed approximately 15,000 10th graders in 2002 and has since conducted follow-ups in 2004 and 2006. Using the 2004 and 2006 waves, the weight variable F1QWT, and the conditioning flag F1UNIV2B, our ELS sample is representative of the high school class of 2004.

In general, item non-response in the NCES surveys for the measures used in this article was minor, often less than 5 percent of the sample. The main exception is for the cognitive test battery, a variable used for stratification of the other indicators, in the NELS92 survey. Rather than item non-response, per se, missing test battery scores result from students who were not administered the test battery (generally because they were not in school when the battery was administered and arrangements could not be made to re-test them). In the NELS, 20.6 percent of the eligible sample is missing a test battery score.³ However, a simple OLS regression showed almost no correlation between socioeconomic status and the likelihood of missing a test battery score. While this does not obviate the possibility of other latent correlates, it suggests bias in the results by test quintile is probably small. Consequently, Tables 2 through 5 treat this problem as ignorable.

HERI CIRP Freshman Survey

The Cooperative Institutional Research Program (CIRP) at the University of California, Los Angeles's Higher Education Research Institute (HERI) has conducted an annual, large-scale survey of college freshmen since the mid 1960s. The survey is administered to approximately 400,000 students at 300 to 400 participating schools each year and asks about many student characteristics, including demographic background, secondary school achievement and activities, educational and career plans, and college application behavior. We use aggregated, annual data

² As in all of the NCES surveys here, new individuals were often added in some of the later waves. We focus on individuals who were in the waves that coincide with the senior year of high school and the wave immediately following.

³ In the NLS, 4.9 percent of the sample is missing a score; in the HSB, 1.3 percent is missing a score; and in the ELS, 0 percent is missing a score.

that are weighted to be nationally representative of all first-year students at four-year colleges and universities, as published in selected editions of *The American Freshman* (1991, 1996, 2000, and 2004) and *The American Freshman: Forty Year Trends*. As individual institution-level identifiers are not available, we use the selectivity measures provided in the survey (which are based on mean SAT scores or equivalents of the entering class), aggregated across institution type. For more information about the Freshman Survey, see: <http://www.gseis.ucla.edu/heri/cirpoverview.php>

SAT and AP Exam Data

The College Board provides the number of SAT tests and AP exams taken each year by the students of each state.⁴ With population data at the state, single-year-of-age level from the National Cancer Institute's Surveillance Epidemiology and End Results (SEER) program (<http://seer.cancer.gov/popdata/>), we construct the test-taking rates. The data on score sending by SAT and ACT score band were kindly provided by Jesse Rothstein and Amanda Pallais, respectively.

SAT states

As mentioned in the text, the SAT is prevalent in some states, while the ACT is prevalent in others. We define an SAT state to be one in which the SAT participation rate weakly exceeds the national average in 2007 (43.7 percent).⁵ The resulting set of SAT states comprises: Alaska, California, Connecticut, Delaware, Florida, Georgia, Hawaii, Indiana, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, North Carolina, Oregon, Pennsylvania, Rhode Island, South Carolina, Texas, Vermont, Virginia, Washington, and the District of Columbia.

Geographical Definitions

The geographic regions we use in this article are based on the classification used in the American Survey of Colleges (ASC). They are defined as follows:

- *New England*: Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont.
- *Middle Atlantic States*: Delaware, Maryland, New Jersey, New York, Pennsylvania, and Washington, D.C.
- *South*: Alabama, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, and Virginia.
- *Midwest*: Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, West Virginia, and Wisconsin
- *Southwest*: Arkansas, New Mexico, Oklahoma, and Texas.
- *West*: Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, Oregon, Utah, Washington, and Wyoming.

⁴ The SAT data are for the high school graduating class of the specified year, while the AP data are for all exams administered in the specified year, regardless of cohort. See <http://professionals.collegeboard.com/data-reports-research/sat/cb-seniors-2008> and <http://professionals.collegeboard.com/data-reports-research/ap/nation> for more information.

⁵ The SAT participation rate in turn is constructed by dividing the number of test takers by state and cohort, as provided by the College Board, by the estimated number of high school graduates, as provided by WICHE (1998, 2008). The set of SAT states is not sensitive to using a different year cutoff.

Appendix Table 1: American Survey of Colleges – Summary Statistics for Four-Year Institutions

	Full Sample	Private Institutions		Public Institutions	
		<i>Top 50 Schools</i>	<i>Other Schools</i>	<i>Top 50 Schools</i>	<i>Other Schools</i>
Percent Public	32.55	0.00	0.00	100.00	100.00
Percent Private	67.45	100.00	100.00	0.00	0.00
New England	13.86	26.31	12.24	7.08	12.93
Middle States	25.23	33.56	25.52	10.62	23.66
South	18.95	14.25	17.42	26.55	22.71
Midwest	27.56	13.15	32.96	31.86	22.24
Southwest	4.18	0.00	4.47	5.31	5.21
West	10.20	12.73	7.39	18.58	13.25
1986 Mean Characteristics					
In-State Tuition	\$3970 (3455)	\$7753 (4278)	\$4964 (2772)	\$862 (855)	\$896 (713)
Undergraduate Enrollment	4603 (6093)	4086 (3978)	1667 (1526)	20980 (9567)	7852 (5679)
Percent Minority	11.46	12.04	9.85	11.65	14.31
2003 Mean Characteristics					
In-State Tuition	\$14,347 (8747)	\$25,719 (6419)	\$18,048 (4643)	\$4918 (1570)	\$3685 (1635)
Undergraduate Enrollment	4982 (6118)	4374 (4418)	1898 (1442)	20,188 (6471)	9025 (6568)
Percent Minority	20.85	25.61	18.17	24.11	23.29
Schools	766	91	425	38	212
Observations	13,770	1642	7646	678	3804

Source: American Survey of Colleges, College Board, 1986-87 to 2003-04.

Notes: The Top 50 Private Institutions includes the top 50 private universities and top 49 liberal arts colleges. To be included in the sample, institutions must have had at least 16 of the 18 possible years of data. The regional statistics are percentages.

Appendix Table 2: Most Selective Four-year Colleges and Universities

Top 50 Private Universities	Top 49 Private Liberal Arts Colleges	Top 47 Public Universities
American	Amherst	Auburn Univ.
Baylor	Bard	Clemson Univ.
Boston College	Barnard	Coll. of William and Mary
Boston Univ.	Bates	Georgia Tech
Brandeis	Bowdoin	Indiana Univ., Bloomington
Brigham Young	Bryn Mawr	Iowa State Univ.
Brown	Bucknell	Miami Univ.
Cal Tech	Carleton	Michigan State Univ.
Carnegie Mellon	Centre	N. Carolina State Univ.
Case Western Reserve	Claremont McKenna	Ohio State Univ.
Clark	Colby	Ohio Univ.
Columbia	Colgate	Pennsylvania State Univ.
Cornell	Colorado Coll.	Purdue Univ.
Dartmouth	Connecticut Coll.	Rutgers, New Brunswick
Duke	Davidson	SUNY, Binghamton
Emory	DePauw	Texas A&M
		Univ. of California, San Diego
Fordham	Dickinson	Univ. of Arizona
George Washington	Franklin and Marshall	Univ. of California, Berkeley
Georgetown	Furman	Univ. of California, Davis
Harvard	Gettysburg	Univ. of California, Irvine
Howard	Grinnell	Univ. of California, Los Angeles
		Univ. of California, Riverside
Johns Hopkins	Hamilton Coll.	Univ. of California, Santa Barbara
		Univ. of California, Santa Cruz
Lehigh	Harvey Mudd	Univ. of Colorado, Boulder
		Univ. of Connecticut
MIT	Haverford	Univ. of Delaware
		Univ. of Florida
New York Univ.	Holy Cross	Univ. of Georgia
Northwestern	Kenyon	Univ. of Illinois, Urbana-Champaign
Notre Dame	Lafayette Coll.	Univ. of Iowa
Pepperdine	Macalester	Univ. of Kansas
Princeton	Middlebury	Univ. of Maryland, College Park
Rensselaer Polytechnic	Mount Holyoke	Univ. of Massachusetts, Amherst
Rice	Oberlin	
Southern Methodist	Occidental	
St. Louis University	Pitzer	
Stanford	Pomona	
Stevens Inst. of Technology	Rhodes Coll.	

Syracuse	Richmond	Univ. of Michigan
Tufts	Scripps	Univ. of Minnesota, Twin Cities
Tulane	Sewanee	Univ. of Missouri, Columbia
Univ. of Chicago	Skidmore	Univ. of N. Carolina, Chapel Hill
Univ. of Miami	Smith	Univ. of Nebraska, Lincoln
Univ. of Pennsylvania	Swarthmore	Univ. of New Hampshire
Univ. of Pittsburgh	Trinity Coll.	Univ. of Tennessee
Univ. of Rochester	Union	Univ. of Texas, Austin
Univ. of Southern California	Vassar	Univ. of Vermont
Vanderbilt	Washington and Lee	Univ. of Virginia
Wake Forest	Wellesley	Univ. of Washington
Washington Univ. (St. Louis)	Wesleyan	Univ. of Wisconsin, Madison
Worcester Polytechnic	Whitman	
Yale	Williams	
Yeshiva		

Source: Bound, Lovenheim, and Turner (2007) who in turn employ the rankings from the 2005 edition of *U.S. News and World Report's Best Colleges*.

Appendix Table 3: Counterfactual College Acceptance, Conditional on Applying

<i>HS Cohort</i>	<i>Percentage who Applied to a Four-year Institution</i>				<i>Percentage who Applied to a Private Selective Four-year Institution</i>				<i>Percentage who Applied to a Public Selective Four-year Institution</i>			
	<i>1972</i>	<i>1982</i>	<i>1992</i>	<i>2004</i>	<i>1972</i>	<i>1982</i>	<i>1992</i>	<i>2004</i>	<i>1972</i>	<i>1982</i>	<i>1992</i>	<i>2004</i>
U.S. average	94.7	98.0	91.3	88.8	84.1	81.4	71.6	62.7	88.4	86.3	86.3	82.6
Test Quintile												
First	88.5	94.9	75.0	64.3	—	—	—	—	—	—	—	—
Second	88.0	95.4	81.8	73.8	—	—	—	—	70.9	63.1	65.4	61.7
Third	92.1	98.2	89.4	84.6	78.0	65.7	52.5	53.7	83.8	83.4	78.3	78.6
Fourth	94.2	98.7	92.9	90.2	74.6	87.3	64.5	52.9	87.4	88.1	79.7	74.2
Fifth	97.8	98.2	94.1	94.3	87.6	81.5	75.7	68.9	91.8	89.1	93.3	89.7
Region												
New England	85.5	97.3	95.5	89.0	76.1	78.8	72.1	62.6	79.7	68.9	89.6	86.0
Mid. Atlantic	92.7	97.2	90.0	90.1	87.4	78.3	68.1	62.0	84.8	78.1	81.5	83.2
South	97.2	97.5	91.3	86.5	85.7	87.9	75.9	62.4	82.5	91.8	83.8	74.0
Midwest	98.1	99.4	94.6	90.9	86.2	81.8	73.3	71.3	94.4	91.9	88.3	86.9
Southwest	95.3	99.6	86.7	83.7	79.0	89.0	81.0	67.7	88.2	93.7	79.6	88.2
West	94.0	96.7	86.7	86.6	78.1	84.1	73.2	55.3	89.0	86.0	89.6	76.5

Source: *National Center for Education Statistics*, longitudinal surveys (NLS72, HSB82, NELS92, ELS04). The cohort year refers to the year on-time students would have graduated high school.

Notes: Data represent mean fitted probabilities from logistic regressions using the NLS72 data but allowing coefficients to be survey-wave specific. Each number thus represents the mean conditional probability that a student from 1972 in a given cell is admitted to a given college type during the respective survey wave. Application behavior is based on the top two school choices of respondents. Geography is according to the high school of the student. The test quintile comes from a survey-specific cognitive test battery given to the respondents of each survey during the spring of their senior year; by construction, it is normalized by cohort. (The test batteries are similar but not identical across surveys.) See the Data Appendix for the definitions of selective schools and the regional breakdowns. The covariates used for the regression include class rank decile dummy variables (GPA for 2004); categorical homework time dummy variables; region dummy variables; test quintile dummy variables; high school semesters each of English, math, social studies, science, and foreign languages; and participation dummy variables for each of sports, music/debate/drama, newspaper/yearbook, and student government.

Appendix Table 4: NCES Longitudinal Student Datasets

	NLS72		HSB82		NELS92		ELS04	
	<i>Un-Weighted</i>	<i>Weighted</i>	<i>Un-Weighted</i>	<i>Weighted</i>	<i>Un-Weighted</i>	<i>Weighted</i>	<i>Un-Weighted</i>	<i>Weighted</i>
Female	51.0%	50.1%	51.9%	51.1%	51.3%	50.0%	50.5%	50.2%
Black	12.6%	8.5%	13.1%	11.8%	10.0%	11.5%	12.4%	13.3%
Hispanic	4.5%	3.3%	21.4%	12.2%	12.6%	9.9%	13.7%	15.1%
Asian	1.2%	0.9%	3.4%	1.4%	8.3%	4.5%	10.5%	4.5%
Native American	1.1%	1.0%	1.9%	1.1%	1.1%	1.1%	0.8%	0.9%
White	77.8%	83.4%	60.1%	73.2%	67.9%	73.0%	58.2%	62.2%
Mom: HS degree	69.6%	72.8%	81.2%	80.2%	85.2%	86.8%	89.0%	88.7%
Mom: College degree	10.7%	10.6%	14.1%	15.6%	26.1%	26.0%	29.7%	26.9%
Dad: HS degree	64.9%	68.7%	76.9%	77.1%	84.0%	85.8%	88.4%	87.9%
Dad: College degree	17.2%	16.8%	24.1%	22.5%	32.5%	32.6%	34.4%	31.2%
New England	5.3%	6.9%	5.7%	6.9%	4.3%	4.7%	4.4%	5.2%
Middle States	19.3%	22.3%	22.4%	20.2%	16.4%	17.1%	16.3%	15.6%
South	21.9%	16.4%	16.7%	18.5%	21.4%	22.1%	25.3%	23.3%
Midwest	28.3%	31.3%	28.1%	29.5%	27.2%	26.7%	26.0%	25.2%
Southwest	8.9%	6.9%	9.8%	9.4%	11.1%	10.7%	8.4%	9.1%
West	16.3%	16.3%	17.4%	15.5%	19.5%	18.7%	19.7%	21.6%
Observations	15,635	3,043,599	11,189	3,161,622	12,312	2,502,578	13,370	2,983,515

Sources: NCES surveys, described in appendix text.

Notes: Numbers are percentages. Categories may not sum to 100 due to rounding and, for race, an omitted “other” category. Observations with zero or invalid (missing) weights were dropped. Race categories in ELS04 exclude the multiracial (about 4 percent of the student population). Education categories in HSB82 and NELS92 exclude those who report that they don’t know their parents’ education.

Appendix Table 5: Composite Index of Competitive Pressure

Rank	State	Index	Rank	State	Index
1	New Jersey	1.763	27	Indiana	0.876
2	Rhode Island	1.736	28	Wyoming	0.847
3	District of Columbia	1.708	29	Alabama	0.835
4	Connecticut	1.608	30	South Carolina	0.826
5	Massachusetts	1.551	31	Tennessee	0.824
6	New York	1.512	32	Wisconsin	0.822
7	Delaware	1.284	33	Minnesota	0.818
8	Virginia	1.240	34	Nebraska	0.813
9	California	1.204	35	Maine	0.800
10	Colorado	1.160	36	Ohio	0.796
11	Georgia	1.157	37	Nevada	0.794
12	Maryland	1.055	38	Washington	0.766
13	Illinois	1.045	39	Oklahoma	0.761
14	Florida	1.034	40	Kansas	0.753
15	New Hampshire	1.032	41	Louisiana	0.732
16	North Carolina	0.993	42	Montana	0.727
17	Alaska	0.987	43	Mississippi	0.702
18	Utah	0.984	44	Idaho	0.676
19	Hawaii	0.962	45	Iowa	0.673
20	Oregon	0.955	46	North Dakota	0.671
21	Pennsylvania	0.944	47	New Mexico	0.663
22	Kentucky	0.939	48	Missouri	0.658
23	Vermont	0.929	49	West Virginia	0.624
24	Texas	0.902	50	Arkansas	0.482
25	Michigan	0.893	51	South Dakota	0.398
26	Arizona	0.884			

Notes: The competitiveness index is the sum of the fractions of students doing each of the following five indicators in 1992: taking the PSAT, taking an AP exam, spending 10+ hours on homework per week, using private test preparation services, and sending 5+ college applications. Each indicator is taken from authors' calculations of NCES data, except for AP exam participation, which is from the College Board.