Why Have College Completion Rates Declined? An Analysis of Changing Student Preparation and Collegiate Resources

John Bound, Michael F. Lovenheim and Sarah Turner

Web Appendix

Appendix A: Technical Appendix

I. NLS72 and NELS:88 Data

a. Degree Completion

Degree completion is calculated using National Longitudinal Study of the High School Class of 1972 (NLS72) and the National Educational Longitudinal Study of 1988 (NELS:88) survey responses from the first through fifth follow-ups in NLS72 and the fourth follow-up in NELS:88. The NLS72 study participants were seniors in high school in the spring of 1972. Following the base year interview, participant follow-up surveys were administered in 1973, 1974, 1976, 1979, and 1986 (for a subsample), with questions covering collegiate participation and degree attainment. In addition, detailed high school records and postsecondary transcripts were collected by the Department of Education.

The NELS:88 survey started with students who were in the eighth grade in 1988 (high school class of 1992) and conducted follow-up surveys with participants in 1990, 1992, 1994, and 2000. Similar to the NLS72 survey, NELS:88 contains high school records and collegiate transcripts as well as a host of background information that may be relevant to degree completion.

Degree completion is defined as obtaining a BA within 8 years of cohort high school graduation conditional on beginning college within 2 years of cohort high school graduation. Cohort high school graduation is defined as June 1972 for the NLS72 sample and June 1992 for the NELS:88 sample.

Because the NELS:88 survey is comprised of eighth graders from 1988 and the NLS72 survey follows 12th graders from the class of 1972, the NELS:88 survey contains more students who graduate high school after their cohort's high school graduation. In our base sample, 1.3% of respondents in NLS72 and 4.4% of respondents in NELS:88 finish high school after June of their respective cohort graduation year. However, looking only at eight-year BA recipients, 0.3% and 0.6%, respectively in NLS72 and NELS:88 did not finish high school on time, and few of these students entered college within 2 years of cohort high school graduation. It is therefore unlikely the larger preponderance of late high school graduates in the NELS:88 survey biases our completion rate calculations.

Table A-1 of this Appendix contains variable names and definitions used to define the sample and to calculate degree completion in both the NLS72 and NELS:88 surveys.

b. School Type and Collegiate Start Dates

We define enrollment as those who start at an academic institution within two years of cohort high school graduation. Academic institutions are all four-year schools and public two-year schools. We exclude private two-year schools because they typically are not oriented towards allowing students to obtain a BA post-graduation.

College transcript data and self-reported enrollment records from the first through fourth follow-up surveys for the NLS72 survey and from NCES-aggregated responses in the NELS:88 survey are used to define the type of institution of initial collegiate enrollment. We use the transcript for the first institution post-high school attended by respondents in the transcript files to assign first institution attended for most respondents. In the cases in which there are multiple first transcripts from different institutions on the same date, we assign each student to the school at which she took the most credits during the first semester. There are some students who report attending college within two years of their cohort's high school graduation but do not have any transcripts. In NLS72, 6.8% of the sample reporting attendance do not have transcripts, and in NELS:88, 8.2% of the sample falls into this category. For these respondents, we use the first institution reported by them in the survey files.

In the NLS72 survey, we begin by determining the year in which a student first enrolls in an academic post-secondary institution, where "academic" is defined as granting at least an associates degree or BA. In each follow-up, students were asked about colleges they attended (up to three) in each year since the previous survey. The first college attended is identified from the entry the first time a student reports attending an academic institution and we record the institutional identifier (FICE code) either directly from transcript files or from the student survey responses about which institution they attended. We then merge institutional-level information that contains public/private status, 2-year/4-year identifiers, and collegiate rankings and classify the respondent's initial institution accordingly.

In the NELS:88 survey, we use a similar methodology to identify each respondent's initial institution. NCES has constructed variables that identify first institution attended in the transcript files (the "ref" variables). We use the transcript-based NCES-constructed institutional identifier ("unitid") code when it is available. For those who report college attendance and the sector of first attendance but are not assigned a transcript-based first institution identifier by NCES, we use the NCES-constructed variables that report individual enrollment histories from the survey data that identify first institution of enrollment ("unitid") and first institution type ("f4efsect").

For students with post-secondary experience preceding high school graduation, we use the first start date and institution after high school graduation taken from the post-secondary transcript files. For all other students in the NELS:88 survey, first start date is identified by f4efmy, which is the NCES-constructed date of first post-secondary attendance.

A list of the top-50 public schools from the 2005 U.S. News and World Report rankings as well as the top-65 private schools and the top-50 liberal arts colleges plus the United States Armed Services Academies, which constitute the highly selective private schools, is shown is shown in Table A-2.

c. Background Characteristics

i. Math Tests

In both surveys, tests of academic achievement were administered to students in the senior year. The NLS72 exam was administered as a 69-minute test book with sections on vocabulary, picture numbers (associative memory), reading, letter groups, mathematics, and mosaic comparisons. Each section was 15 minutes (except for the mosaic comparison, which was 9 minutes). The math test included 25 items and contained only quantitative comparisons in order to measure basic quantitative competence. We use the reported scaled math score (scmatsc) test score measure in NLS72.

The NELS:88 cognitive test batteries were administered in each of the first three waves, with sections on reading, math, science and social studies. The tests were 85 minutes and consisted of 116 questions, 40 of which were on math and 21 of which were on reading comprehension. Unlike the NLS72 exams, the NELS:88 tests covered more material and tested more skills. The math exam consisted of word problems, graphs, equations, quantitative comparisons, and geometric figures. Further, because the NELS:88 tests were given in subsequent waves, students were given harder or easier tests in the first and second follow-ups depending on their scores in the previous wave to guard against floor and ceiling effects. We use the math IRT theta score (f22xmth) from the second follow-up as the base measure of test scores. These scores are psychometric evaluation scores of each student's ability that account for the difficulty of the exam.

Because the tests in NLS72 and NELS:88 covered different subject matter, were of different lengths, and were graded on different scales, the scores are not directly comparable across surveys. Instead, we construct the percentile of the score distribution for each survey among all high school graduates. The comparison of students in the same test percentile across surveys is based on the assumption overall achievement did not change over this time period. This assumption is supported by the observation that there is little change in the overall level of test scores on the nationally-representative NAEP over our period of observation. Similarly, examination of time trends in standard college entrance exams such as the SAT provides little support for the proposition that achievement declined appreciable over the interval. For the SAT, the ratio of test takers to high school graduates increased from 33% to 42%, while mean math scores declined from 509 to 501 over the 1972 to 1992 interval (*Digest of Education Statistics*, 2005, Table 129).

In the NLS72 survey, we use high school GPA as an imputation variable in order to measure pre-collegiate academic ability for students with missing test scores. The GPA measure we use is "imptaver" from the NLS72 survey. In the multiple imputation of missing variables in the NELS:88 survey, we use IRT theta test scores from the first follow-up for math (f12xmth) and from the base year for math (by2xmth). The IRT theta scores are scaled to a common metric across years by NCES. The imputed math test scores from the senior year in each survey are used to construct the test percentiles used in the main analysis.

ii. Parental Education

We obtain student reported measures of father's and mother's education separately. In the NLS72 survey, we have three different measures of this variable. For mother's education, we use the variables cmoed, bq90b, and fq78b. For father's

education, we use the variables cfaed, bq90a, and fq78a. If there are disagreements across measures, fq78b and fq78a take precedence.

In the NELS:88 survey, we also use student reports of father's education (bys34a) and mother's education (bys34b). For the multiple imputation model, we include parent self-reports of their own education from the base year and second follow-up parental surveys. In the base year parent survey, we combine information on whether the respondent and his/her spouse is the father or mother (byp1a1 and byp1a2) with reported self (byp30) and spouse (byp31) educational attainment. A similar methodology is used for the second-follow up parent survey, using f2p1a and f2p1b to identify the gender of the respondent and the spouse, respectively, and f2p101a and f2p101b to identify educational attainment of the respondent and the spouse, respectively. The base year and second follow-up parental education information is aggregated into two variables, father's education and mother's education, used in the multiple imputation model.

iii. Parental Income Levels

The parental income variables are bq93 for NLS72 and f2p74 for NELS:88. The former is reported by the student while the latter is reported by the parents. Unfortunately, NLS72 does not contain a parent-reported measure and the NELS:88 survey does not contain a student-reported measure, so these variables are the most closely aligned parental income measures across the two surveys.

Rather than asking directly for parental income levels, the NELS:88 and NLS72 surveys ask for income ranges from respondents. Because we are interested in measuring parents' ability to finance college, the variable of interest is the real income level, not one's place in the income distribution. We thus align the income blocks across the two surveys using the CPI. In NLS72, the measured income groups we construct are less than $3000, $3000-$6000, $6000-$7500, $7500-$10500, $10500-$15000, and greater than $15000. In NELS:88, the corresponding real income blocks we create are less than $10000, $10000-$20000, $20000-$25000, $25000-$35000, $35000-$50000, and greater than $50000. Across surveys, the six income groups are comparable in real terms.

iv. Race

Race is measured in the NLS72 survey using "crace" and "race86." The latter is used if the former is blank due to non-response. In the NELS:88 survey, race is measured using the "race" variable available in the data files.

II. Procedures to Handle Missing Data

a. Multiple Imputation

There is a considerable amount of missing data in the NLS72 and NELS:88 surveys. Table A-3 of this Appendix presents the number of unweighted missing observations by variable and survey. These observations are not missing completely at random; respondents who have no math test scores are less likely to finish college conditional on starting.

Casewise deletion of missing observations will therefore cause a bias in the calculation of the base trends we are seeking to explain in this analysis. To deal with this

problem, we use the multiple imputation by chained equation (MICE) algorithm developed by Van Buuren, Boshuizen, and Knook (1999) that is implemented through the STATA module "ICE" (see Royston (2004) for a detailed discussion of ICE).

MICE is implemented by first defining the set of predictor variables ($x_1 \ldots x_k$) and the set of variables with missing values to be imputed: math test scores, father's education, mother's education, and parental income levels ($y_1 \ldots y_5$). The MICE algorithm implemented by ICE first randomly fills in all missing values from the posterior distribution of each variable. Then, for each variable with missing data, y_i, STATA runs a regression (or ordered logit) of y_i on $y_{\sim i}$ and $x_1 \ldots x_k$ and calculates expected values from these regressions for all missing data points. The expected values then replace the randomly assigned values for the missing data points. A sequence of regressions for each y_i is a cycle, and this process is repeated for 10 cycles, replacing the missing values with the new expected values from each regression in each cycle. The imputed values after 10 cycles constitute one imputed data set, and this process is repeated five different times to generate five imputed data sets.

There are two important specifications in implementing MICE: determination of the predictor variables and determination of the imputation models. Because of the different structure of the two surveys, different variables are used in the imputation procedure across surveys. In both surveys, we include dummy variables for cumulative time to degree from four to eight years, dummy variables for initial school type, interactions between these variables, an indicator for college attendance within two years of cohort high school graduation, as well as race and gender indicators.

For imputations with the NLS72 sample, we include a measure of high school GPA in order to proxy for unobserved ability among those without test score information. Due to the structure of the NELS:88 survey, there is more background information with which to impute missing data. We use 8^{th} and 10^{th} grade math test scores, parental reports of their education from the base year and second follow-up parent surveys, and parental reports of their income level from the base year parent survey. The definitions of the variables used in the imputation models are discussed in the preceding section.

Because the math test scores are continuous variables, we use OLS regressions to impute these variables. Mother's and father's education and income, however, are categorical variables. Because of the ordered nature of these variables, we use ordered logits to impute the missing values of these variables. While these model choices are reasonably arbitrary, they are only used to draw ranges of plausible estimates of missing data.

The multiple imputation procedure creates five different data sets, each with different imputed values for the missing observations. All reported statistics and results in our analysis are averages across data sets. In other words, we conduct each analysis separately for each data set and average the final result. The average of final results is what is reported in the tables and figures in the paper. For the completion logit estimates in Table 5, we use the "micombine" command in STATA that is designed to estimate coefficients and standard errors from multiply imputed data sets. The coefficient estimates represent the simple arithmetic mean across the five coefficients from each imputed data set for each variable. Standard error estimation in this context is more complex. See Carlin et al. (2003) for a complete description of how standard errors are calculated.

b. Dropped Observations and Missing Transcript Data

The base sample in this analysis consists of all respondents who graduate high school and attend college within two years of their cohort's high school graduation. We further restrict the sample to exclude those whose only enrollment over this time period is at a private two-year institution as these schools are predominantly professional without a BA track. Table A-4 presents information on the number of observations that are dropped by survey and the reason for dropping the observation. For example, 168 respondents are dropped because they are not high school graduates in NLS72 whereas 722 are dropped in NELS:88 for this reason. The apparently higher dropout rate in NELS:88 is because the universe of students are all those enrolled in the 8^{th} grade in 1988, whereas the universe in NLS72 are all those enrolled in 12^{th} grade in 1972.

In the NLS72 survey, 63 observations are dropped because they report attending college but provide no information on either the type of institution or the date they first began attending this institution, and in NELS:88, 51 respondents do not provide this information. In addition, 51 observations were dropped because they were not in all four waves of the NELS:88 survey. In other words, they have a sample weight of zero.

Of potential concern in constructing our sample is the exclusion of those beginning college more than two years post-high school cohort graduation. We exclude these observations because we are interested in the truncated, eight-year completion rate. These statistics have a different interpretation for a student who began college directly after high school than for a student who began college, for instance, five years after high school. In NLS72, 889 respondents attend college more than 2 years after their cohort's high school graduation, and in NELS:88, 971 do so. Given the similarity of these numbers, shifts in when students began attending college cannot account for the trends in completion rates reported in the main text.

III. Counterfactual Simulation Procedure

The counterfactual math test scores, student-faculty ratios and initial school types are computed using a matching algorithm that holds relative rank constant. The goal of this procedure is to generate a counterfactual of the variable *X* on the NELS:88 sample that has the same distribution as the variable in the NLS72 sample. To construct this counterfactual, we first expand both data sets based on the analysis sample weights and then balance the two surveys by randomly deleting observations from the survey with more weighted observations. The weight-expanded samples then have the same number of observations in each survey. We then use the psmatch2 program in STATA to match each NELS:88 respondent to a NLS72 respondent on the variable *X*, using single nearest neighbor matching without replacement. For each NELS:88 respondent, we assign a counterfactual value of *X* that is equal to the value from the unique match. This matching method creates a counterfactual distribution of *X* in the NELS:88 sample that has the same mean as in the NLS72 sample and has a distribution that holds relative rank-order constant across the surveys. Using single nearest neighbor matching on *X* without replacement with psmatch2 ensures that we are finding a unique match for each NELS:88

respondent and that we are breaking ties randomly. Finally, we collapse the data back down to the individual level to conduct our analyses.

Because the math test percentiles we use are multiply imputed, we perform this procedure separately for each imputed data set. However, the imputations are done separately by survey, so it is not correct to match similarly numbered imputations with each other across cohorts. Instead, for each imputation in NELS:88, we perform our matching algorithm separately for each imputation in NLS72 and take the average counterfactual math test percentile across NLS72 imputations for each NELS:88 imputation. This procedure requires us to perform the matching algorithm 25 times when we simulate counterfactual math test percentiles.

When we simulate counterfactual student-faculty ratios and initial school types, we only have to conduct the matching algorithm once because none of the data are imputed. But, for student-faculty ratios, there are a number of missing observations. To handle this difficulty, we treat missing status as a fixed attribute and perform the matching only for observations with non-missing student-faculty ratio values. For those with missing student-faculty ratios, they are assumed to have missing data both in the observed and counterfactual completion rate calculations.

With test score percentiles and student-teacher ratios, establishing rank-order is straightforward as these variables are cardinal. For initial school type, however, how to rank-order the samples is less clear. For the matching procedure, we order school types in ascending order from least to most selective. Thus, school type 1 is community colleges, followed by non-top 50 public universities, less selective private universities, top 50 public schools, and finally highly selective private universities. The logic behind this ordering is that we are assigning students in our counterfactual simulations to sectors that are most similar to their observed sector; we are not assigning, for example, students in community colleges to the highly selective private sector when finding counterfactual school types. This ordering creates conservative estimates of the effect of school type shifts on completion rates because we are minimizing the quality increases for each student associated with the shift across sectors that has occurred over time.

References

Carlin, John B., Ning Li, Philip Greenwood, and Carolyn Coffey. 2003. “Tools for Analyzing Multiple Imputed Datasets.” *Stata Journal*, 3(3): 226-244.

Royston, Patrick. 2004. “Multiple Imputation of Missing Values.” *Stata Journal*, 4(3): 227-241.

Van Buuren, Stef, Hendriek C. Boshuizen, and Dick L. Knook. 1999. “Multiple Imputation of Missing Blood Pressure Covariates in Survival Analysis.” *Statistics in Medicine*, 18(6): 681-694.

Table A-1. Variable Names and Definitions for Calculation of Completion Rates in NLS72 and NELS:88

Panel A: NLS72		
Variable Name	**Variable Definition**	**Follow Up**
Fq2	High school completion dummy	2
Edatt86	Educational attainment as of 1986	1-5
Fq3b	High school graduation year	2
Fq3a	High school graduation month	2
Tq48ea	BA completion dummy as of 10/1/1976	3
Tq48eb	Month BA received as of third follow-up	3
Tq48ec	Year BA received as of third follow-up	4
Ft76ea	BA completion as of fourth follow-up	4
Ft76eb	Month BA received as of fourth follow-up	4
Ft76ec	Year BA received as of fourth follow-up	5
Fi19b1ey - Fi19b4ey	Year ended most recent school attended, first through fourth time	5
Fi19b1em–Fi19b4em	Month ended most recent school attended, first through fourth time	5
Fi19h	Course of study in most recent school attended	5
Fi19i	Completed requirements in most recent school attended	5
Fi20b1ey–Fi20b4ey	Year ended 2nd most recent school attended, first through fourth time	5
Fi20b1em–Fi20b4em	Month ended 2nd most recent school attended, first through fourth time	5
Fi19h	Course of study in 2nd most recent school attended	5
Fi19i	Completed requirements in 2nd most recent school attended	5

Panel B: NELS:88		
Variable Name	**Variable Definition**	**Follow Up**
F4hsgradt	High school graduation date	4
F4ed1	Degree receipt date–first degree received	4
F4edgr1	Degree type received–first degree	4
F4ed2	Degree receipt date–second degree received	4
F4edgr2	Degree type received–second degree	4
F4ed3	Degree receipt date–third degree received	4
F4edgr3	Degree type received–third degree	4
F4ed4	Degree receipt date–fourth degree received	4
F4edgr4	Degree type received–fourth degree	4
F4ed5	Degree receipt date–fifth degree received	4
F4edgr5	Degree type received–fifth degree	4
F4ed6	Degree receipt date–sixth degree received	4
F4edgr6	Degree type received–sixth degree	4

Table A-2. Top-50 Public Schools, Top-65 Private Schools and Top-50 Liberal Arts Colleges from the 2005 U.S. News and World Report Rankings

	Highly Selective Private Schools		
Top-50 Public Schools	**Top-65 Private Schools**		**Top-50 Liberal Arts**
University of California – Berkeley	Harvard University	University of Tulsa	Amherst College
University of Virginia	Princeton University	Texas Christian University	Williams College
University of Michigan – Ann Arbor	Yale University	University of Dayton	Swarthmore College
University of California – Los Angeles	University of Pennsylvania	Drexel University	Wellesley College
University of North Carolina – Chapel Hill	Duke University	Illinois Institute of Technology	Carleton College
College of William and Mary	MIT	University of San Diego	Middlebury College
University of Wisconsin – Madison	Stanford University	Catholic University	Pomona College
University of California – San Diego	California Institute of Tech.	Loyola University	Bowdoin College
University of Illinois	Columbia University	Univ. of San Francisco	Davidson College
Georgia Institute of Technology	Dartmouth College	University of the Pacific	Haverford College
University of California – Davis	Northwestern University	New School	Claremont-McKenna
University of California – Irvine	Washington Univ. of St. Louis	Northeastern University	Wesleyan University
University of California – Santa Barbara	Brown University	Seton Hall University	Grinell College
University of Texas – Austin	Cornell University	University of St. Thomas	Vassar College
University of Washington	Johns Hopkins University		Harvey Mudd College
Pennsylvania State University	University of Chicago		Washington and Lee
University of Florida	Rice University		Smith College
University of Maryland – College Park	Notre Dame University		Hamilton College
Rutgers University – New Brunswick	Vanderbilt University		Colgate University
University of Georgia	Emory University		Oberlin College
University of Iowa	Carnegie Mellon University		Colby College
Miami University (Ohio)	Georgetown University		Bates College
Ohio State University	Wake Forest University		Bryn Mawr College
Purdue University	Tufts University		Colorado College
Texas A&M – College Station	Univ. of Southern California		Macalester College
University of Connecticut	Brandeis University		Scripps College

University of Delaware	New York University	Mt. Holyoke College
University of Minnesota – Twin Cities	Case Western Reserve	Barnard College
University of Pittsburgh	Boston College	Bucknell University
Indiana University	Lehigh University	Kenyon College
Michigan State University	Univ. of Rochester	College of the Holy Cross
Clemson University	Tulane University	Trinity College
SUNY at Binghamton	Rensselaer Polytechnic	Lafayette College
University of California – Santa Cruz	Yeshiva University	Occidental College
University of Colorado – Boulder	George Washington Univ.	Bard College
Virginia Tech.	Pepperdine University	Furman University
University of California – Riverside	Syracuse University	Whitman College
Iowa State University	Worcester Polytechnic	Union College
North Carolina State University	Boston University	Franklin and Marshall
University of Alabama	University of Miami	Sewanee College
University of Missouri – Columbia	Fordham University	University of Richmond
Auburn University	Southern Methodist Univ.	Connecticut College
University of Kansas	Brigham Young University	Centre College
University of Tennessee – Knoxville	Clark University	Dickinson College
University of Vermont	Stevens Inst. of Technology	Skidmore College
Ohio University	St. Louis University	Gettysburg College
University of Arizona	Baylor University	Pitzer College
University of Massachusetts – Amherst	American University	DePauw University
University of Nebraska – Lincoln	Howard University	Rhodes College
University of New Hampshire	Marquette University	Reed College
	University of Denver	

Source: 2005 U.S. News and World Report Rankings of colleges and universities in the United States. Schools are listed in the order they appear in the U.S. News and World Report ranking. The rankings include many ties, in which case schools are listed alphabetically within rank. This table lists schools within rank in the same manner. The highly selective private school category also includes the four U.S. Armed Services Academies: U.S. Naval Academy, U.S. Air Force Academy, U.S. Military Academy at West Point and U.S. Coast Guard Academy.

Table A-3. Number of Imputed Observations by Survey and Variable (Unweighted)

	Number of Imputed Observations	
Variable	**NLS72**	**NELS:88**
Math Test Score	1,961	1,560
Mother's Education	46	1,199
Father's Education	46	1,408
Parent Income	1,634	1,158
Total	3,687	5,325

Observation counts include only those respondents who enroll in college within two years of cohort high school graduation at a four-year institution or a non-private two-year college.

Table A-4. Number of Dropped Observations by Category (Un-weighted)

NLS72		
Sample Change	**Dropped Observations**	**Remaining Observations**
Original Base - 5th Follow Up Sample		12841
High School Dropouts	168	12673
Missing Initial School Information	63	12610
Never Attended College	4503	8107
Time between HS and College >2 Years	889	7218
NELS:88		
Sample Change	**Dropped Observations**	**Remaining Observations**
Original Base-4th Follow Up Sample		12144
High School Dropouts	722	11422
Observations not in all 4 Waves	196	11226
Missing Initial School Information	51	11175
Never Attended College	2836	9257
Time between HS and College >2 Years	971	8286

Appendix B: Supplemental Decomposition Tables

Table B-1. Decompositions Using Average Derivatives From Logit Models of College Completion in the NELS:88 Survey by Type of Institution

	Full Sample	Public Non-Top 50	Public Top 50	Private Less Selective	Private Highly Selective	Community College
NLS72	50.5	61.8	73.5	58.2	80.1	20.2
NELS:88	45.9	56.9	82.5	70.5	90.3	17.6
Total Change	-4.6	-4.9	9.0	12.3	10.3	-2.5
Change Due to Observables (X)	-4.0	-1.7	3.4	3.0	6.5	-3.5
Change Due to Student Characteristics	-0.4	2.1	2.8	3.6	6.3	-3.0
Math Test Percentiles	-1.6	0.1	1.0	0.2	0.4	-2.6
Other Student Characteristics	1.2	2.0	1.8	3.4	5.9	-0.4
Change Due to Supply-Side Factors	-3.6	-3.8	0.6	-0.6	0.2	-0.5
Student-Faculty Ratios	-1.2	-3.8	0.6	-0.6	0.2	-0.5
Initial School Types	-2.4					
Residual	-0.6	-3.2	5.6	9.3	3.8	1.0

[a] Source: Authors' calculations as described in the text from the NLS72 and NELS:88 surveys . NLS72 calculations were made using the fifth follow-up weights included in the survey. Fourth follow-up weights were used for the NELS:88 survey calculations. Only those participating in these follow-ups are included in the regression. School type samples refer to first institution attended. The NLS72 and NELS:88 samples are restricted to those who attend college within 2 years of cohort high school graduation. Cohort high school graduation is defined as June 1972 for the NLS72 sample and June 1992 for the NELS:88 sample.

[b] All decompositions are computed using average marginal effects from logit models of completion likelihood estimated with the NELS:88 data as described in the main text. Simulated changes are calculated by multiplying the average marginal effect for a given variable by the observed change in means across surveys.

[c] The Change Due to Observables is the difference between the observed NELS:88 completion rate and the simulated NELS:88 completion rate using the distribution of all observables from NLS72. The changes due to math test percentiles, student-faculty ratios and initial school types are the differences between the observed NELS:88 completion rate and the simulated NELS:88 completion rates in which we simulate the completion rate if the each variable were distributed as in the NLS72 survey and all other observables were unchanged. The Change Due to Supply Side Factors is the sum of the changes due to student-faculty ratios and initial school types. We calculate the Change Due to Student Characteristics as the difference between the Change Due to Observables (X) and the Change Due to Supply-Side Factors. The Change Due to Other Student Characteristics is calculated by subtracting the Change Due to Math Test Percentiles from the Change Due to Student Characteristics.

[d] Data on faculty and enrollment are from the HEGIS/IPEDS surveys from the Department of Education.

Table B-2. Decompositions Using Average Derivatives From Logit Models of College Completion in the NELS:88 Survey by Type of Institution and Gender

Panel A: Men

	Full Sample	Public Non-Top 50	Public Top 50	Private Less Selective	Private Highly Selective	Community College
NLS72	51.7	61.2	73.8	59.9	82.7	21.6
NELS:88	43.2	51.6	77.5	67.7	89.2	17.7
Total Change	-8.5	-9.6	3.7	7.8	6.5	-3.9
Change Due to Observables (X)	-5.4	-2.1	1.9	1.1	7.1	-3.6
Change Due to Student Characteristics	-1.2	1.5	1.3	1.5	6.6	-3.0
Math Test Percentiles	-2.5	-0.7	0.6	-1.9	0.4	-3.1
Other Student Characteristics	1.3	2.2	0.7	3.4	6.2	0.1
Change Due to Supply-Side Factors	-4.2	-3.6	0.6	-0.4	0.5	-0.6
Student-Faculty Ratios	-1.3	-3.6	0.6	-0.4	0.5	-0.6
Initial School Types	-2.9					
Residual	-3.1	-7.5	1.8	6.7	-0.6	-0.3

Panel B: Women						
	Full Sample	Public Non-Top 50	Public Top 50	Private Less Selective	Private Highly Selective	Community College
NLS72	49.2	62.4	73.1	56.5	75.6	18.6
NELS:88	48.5	61.8	87.7	72.6	91.4	17.5
Total Change	-0.8	-0.6	14.6	15.9	15.8	-1.1
Change Due to Observables (X)	-2.8	-1.4	3.5	4.3	5.3	-3.5
Change Due to Student Characteristics	0.2	2.4	2.9	5.1	5.3	-3.0
Math Test Percentiles	-0.6	1.0	1.4	2.1	0.5	-2.0
Other Student Characteristics	0.8	1.4	1.5	3.0	4.8	-1.0
Change Due to Supply-Side Factors	-3.0	-3.8	0.6	-0.8	0.0	-0.5
Student-Faculty Ratios	-1.2	-3.8	0.6	-0.8	0.0	-0.5
Initial School Types	-1.8					
Residual	2.0	0.8	11.1	11.6	10.5	2.4

[a] Source: Authors' calculations as described in the text from the NLS72 and NELS:88 surveys . NLS72 calculations were made using the fifth follow-up weights included in the survey. Fourth follow-up weights were used for the NELS:88 survey calculations. Only those participating in these follow-ups are included in the regression. School type samples refer to first institution attended. The NLS72 and NELS:88 samples are restricted to those who attend college within 2 years of cohort high school graduation. Cohort high school graduation is defined as June 1972 for the NLS72 sample and June 1992 for the NELS:88 sample.

[b] All decompositions are computed using average marginal effects from logit models of completion likelihood estimated with the NELS:88 data as described in the main text. Simulated changes are calculated by multiplying the average marginal effect for a given variable by the observed change in means across surveys.

[c] The Change Due to Observables is the difference between the observed NELS:88 completion rate and the simulated NELS:88 completion rate using the distribution of all observables from NLS72. The changes due to math test percentiles, student-faculty ratios and initial school types are the differences between the observed NELS:88 completion rate and the simulated NELS:88 completion rates in which we simulate the completion rate if the each variable were distributed as in the NLS72 survey and all other observables were unchanged. The Change Due to Supply Side Factors is the sum of the changes due to student-faculty ratios and initial school types. We calculate the Change Due to Student Characteristics as the difference between the Change Due to Observables (X) and the Change Due to Supply-Side Factors. The Change Due to Other Student Characteristics is calculated by subtracting the Change Due to Math Test Percentiles from the Change Due to Student Characteristics.

[d] Data on faculty and enrollment are from the HEGIS/IPEDS surveys from the Department of Education.

Table B-3. Oaxaca-Blinder Decompositions based on NELS:88 Estimates of College Completion by Type of Institution

	Full Sample	Public Non-Top 50	Public Top 50	Private Less Selective	Private Highly Selective	Community College
NLS72	50.5	61.8	73.5	58.2	80.1	20.2
NELS:88	45.9	56.9	82.5	70.5	90.3	17.6
Total Change	-4.6	-4.9	9.0	12.3	10.3	-2.5
Change Due to Observables (X)	-4.8	-0.8	3.6	3.6	9.0	-3.8
Change Due to Student Characteristics	-0.3	3.0	3.0	4.3	8.9	-3.2
Math Test Percentiles	-2.0	0.1	1.3	0.2	0.5	-3.4
Other Student Characteristics	1.7	2.9	1.7	4.1	8.4	0.2
Change Due to Supply-Side Factors	-4.5	-3.8	0.6	-0.7	0.1	-0.6
Student-Faculty Ratios	-1.4	-3.8	0.6	-0.7	0.1	-0.6
Initial School Types	-3.1					
Residual	0.2	-4.1	5.4	8.7	1.3	1.3

[a] Source: Authors' calculations as described in the text from the NLS72 and NELS:88 surveys . NLS72 calculations were made using the fifth follow-up weights included in the survey. Fourth follow-up weights were used for the NELS:88 survey calculations. Only those participating in these follow-ups are included in the regression. School type samples refer to first institution attended. The NLS72 and NELS:88 samples are restricted to those who attend college within 2 years of cohort high school graduation. Cohort high school graduation is defined as June 1972 for the NLS72 sample and June 1992 for the NELS:88 sample.

[b] All decompositions are computed using a linear probability model of completion likelihood estimated with the NELS:88 data. Simulated changes are calculated by multiplying the coefficient on a given variable by the observed change in means across surveys.

[c] The Change Due to Observables is the difference between the observed NELS:88 completion rate and the simulated NELS:88 completion rate using the distribution of all observables from NLS72. The changes due to math test percentiles, student-faculty ratios and initial school types are the differences between the observed NELS:88 completion rate and the simulated NELS:88 completion rates in which we simulate the completion rate if the each variable were distributed as in the NLS72 survey and all other observables were unchanged. The Change Due to Supply Side Factors is the sum of the changes due to student-faculty ratios and initial school types. We calculate the Change Due to Student Characteristics as the difference between the Change Due to Observables (X) and the Change Due to Supply-Side Factors. The Change Due to Other Student Characteristics is calculated by subtracting the Change Due to Math Test Percentiles from the Change Due to Student Characteristics.

[4] Data on faculty and enrollment are from the HEGIS/IPEDS surveys from the Department of Education.

Table B-4. Oaxaca-Blinder Decompositions based on NELS:88 Estimates of College Completion by Type of Institution and Gender

Panel A: Men

	Full Sample	Public Non-Top 50	Public Top 50	Private Less Selective	Private Highly Selective	Community College
NLS72	51.7	61.2	73.8	59.9	82.7	21.6
NELS:88	43.2	51.6	77.5	67.7	89.2	17.7
Total Change	-8.5	-9.6	3.7	7.8	6.5	-3.9
Change Due to Observables (X)	-6.1	-1.4	2.7	1.2	8.1	-3.5
Change Due to Student Characteristics	-0.9	2.1	2.2	1.7	7.8	-2.9
Math Test Percentiles	-2.9	-0.7	0.6	-2.2	0.4	-4.0
Other Student Characteristics	2.0	2.8	1.6	3.9	7.4	1.1
Change Due to Supply-Side Factors	-5.2	-3.5	0.5	-0.5	0.3	-0.6
Student-Faculty Ratios	-1.5	-3.5	0.5	-0.5	0.3	-0.6
Initial School Types	-3.7					
Residual	-2.4	-8.2	1.0	6.6	-1.6	-0.4

Panel B: Women						
	Full Sample	Public Non-Top 50	Public Top 50	Private Less Selective	Private Highly Selective	Community College
NLS72	49.2	62.4	73.1	56.5	75.6	18.6
NELS:88	48.5	61.8	87.7	72.6	91.4	17.5
Total Change	-0.8	-0.6	14.6	15.9	15.8	-1.1
Change Due to Observables (X)	-3.7	-0.4	4.0	5.3	9.8	-4.0
Change Due to Student Characteristics	0.2	3.6	3.3	6.2	9.8	-3.4
Math Test Percentiles	-0.8	1.1	2.2	2.5	0.8	-2.6
Other Student Characteristics	1.0	2.5	1.1	3.7	9.0	-0.8
Change Due to Supply-Side Factors	-3.9	-4.0	0.7	-0.9	0.0	-0.6
Student-Faculty Ratios	-1.3	-4.0	0.7	-0.9	0.0	-0.6
Initial School Types	-2.6					
Residual	2.9	-0.2	10.6	10.6	6.0	2.9

[a] Source: Authors' calculations as described in the text from the NLS72 and NELS:88 surveys . NLS72 calculations were made using the fifth follow-up weights included in the survey. Fourth follow-up weights were used for the NELS:88 survey calculations. Only those participating in these follow-ups are included in the regression. School type samples refer to first institution attended. The NLS72 and NELS:88 samples are restricted to those who attend college within 2 years of cohort high school graduation. Cohort high school graduation is defined as June 1972 for the NLS72 sample and June 1992 for the NELS:88 sample.

[b] All decompositions are computed using a linear probability model of completion likelihood estimated with the NELS:88 data. Simulated changes are calculated by multiplying the coefficient on a given variable by the observed change in means across surveys.

[c] The Change Due to Observables is the difference between the observed NELS:88 completion rate and the simulated NELS:88 completion rate using the distribution of all observables from NLS72. The changes due to math test percentiles, student-faculty ratios and initial school types are the differences between the observed NELS:88 completion rate and the simulated NELS:88 completion rates in which we simulate the completion rate if the each variable were distributed as in the NLS72 survey and all other observables were unchanged. The Change Due to Supply Side Factors is the sum of the changes due to student-faculty ratios and initial school types. We calculate the Change Due to Student Characteristics as the difference between the Change Due to Observables (X) and the Change Due to Supply-Side Factors. The Change Due to Other Student Characteristics is calculated by subtracting the Change Due to Math Test Percentiles from the Change Due to Student Characteristics.

[d] Data on faculty and enrollment are from the HEGIS/IPEDS surveys from the Department of Education.