Technical Documentation: College Scorecard Institution-Level Data

Contents

About

The College Scorecard project is designed to increase transparency, putting the power in the hands of students and families to compare how well individual postsecondary institutions are preparing their students to be successful. This project provides data to help students and families compare college costs and outcomes as they weigh the tradeoffs of different colleges, accounting for their own needs and educational goals.

College Scorecard provides 1) data files with data about institutions as a whole and 2) data files with data about specific fields of study within institutions. The purpose of this document is to describe College Scorecard data files for data elements calculated at the institution-level. For more information about the data files describing fields of study within institutions, please see the technical documentation for data files by field of study. These data are provided through federal reporting from institutions, data on federal financial aid, and tax information. These data provide insights into the performance of institutions that receive federal financial aid dollars, and the outcomes of the students of those institutions. A complete set of these data for all active Integrated Postsecondary Education Data System (IPEDS) institutions that participate in Title IV programs (either by disbursing aid or through deferments) and that are not solely administrative offices are available on the Scorecard data webpage and API. In addition, some institutions that do not participate in Title IV programs are also included in the data if they meet similar criteria to Title IV participating institutions[1]. A subset of these data elements is displayed on the consumer-facing College Scorecard website for undergraduate institutions that are currently operating.

Many data elements are drawn directly from, or derived from, data reported to the IPEDS.[2] Note that some colleges report combined data that pertain to more than one IPEDS institution. The reporting institution (called the "parent") also indicates an estimated proportion of the data that applies to each of the institutions for which it is reporting (called "child" institutions). In the Scorecard data, these proportions are used to allocate data reported by parent institutions to the child institutions to simplify both consumer and research use of the data.

Data Dictionary

The Scorecard data dictionary provides information on each metric in the API and downloadable data files, including the variable name from the data files, a longer descriptive name, the API location and developer-friendly metric name, values and value labels for the metrics, data source, and high-level notes for each metric. In addition, the data dictionary contains a cohort map that identifies the group of students each metric corresponds to within the data files. For instance, the predominant undergraduate award in the 2018-19 data file describes the 2017-18 academic year and is based on data collected during the IPEDS 2018-19 data collection. This cohort map information is provided for all yearly raw data

[1] The Scorecard universe of institutions excludes non-Title IV institutions that do not report data to IPEDS on degree/certificate completions and those that do not report enrollment data to IPEDS.
[2] For more information about IPEDS, please see http://nces.ed.gov/ipeds.

files and for the Featured Download – Most recent data file. The data dictionary file is available at https://collegescorecard.ed.gov/assets/CollegeScorecardDataDictionary.xlsx.

Accuracy and Privacy

Many elements are available only for Title IV recipients, or students who receive federal grants and loans. These data are reported at the individual level to the National Student Loan Data System (NSLDS), which is used to distribute federal aid, and published at the aggregate institutional level. While some institutions report these data at the campus level (8-digit OPE ID), data produced for this site are rolled up to the institution level (6-digit OPE ID). In these cases, IPEDS institutions sharing a common 6-digit OPEID are all assigned the same sum or (student-weighted) average outcome or median outcome for students across all branches of the institution for NSLDS or tax-data derived measures.

All NSLDS and Treasury elements are protected for privacy purposes; those data that do not meet reporting standards are shown as PrivacySuppressed. Note that for many elements, we have also taken additional steps to ensure data are stable from year to year and representative of a certain number of students. For many elements, data are pooled across two years of data to reduce year-over-year variability in figures (i.e. repayment rate, debt figures, earnings). Moreover, for elements that are highlighted on the consumer-facing College Scorecard, a separate version of the element is available that suppresses data for institutions with fewer than 30 students in the denominator to ensure data are as representative as possible.

Root

These select items refer to the most basic information in the data set.

Dev-category	root
ID	Integer Data files are provided at the UNITID level, which is the unique identification number assigned to postsecondary institutions as surveyed through IPEDS. OPEID is the identification number used by the U.S. Department of Education's Office of Postsecondary Education (OPE) and Federal Student Aid Office (FSA) to identify institutions that have Program Participation Agreements (PPA) so that its students are eligible to participate in Federal Student Financial Assistance programs under Title IV regulations. This is a 6-digit number and is also provided as an 8-digit number with a 2-digit suffix to identify branches, additional locations, and other entities that are part of the eligible institution.

There is not necessarily a one-to-one relationship between UNITID and the 8-digit OPEID. A UNITID may correspond to the aggregation of several OPEIDs (i.e., several locations of an institution combining their data for IPEDS reporting), or in rarer cases may correspond to part of an OPEID (e.g., a single location may elect to report their traditional campus-based and online operations under separate UNITIDs despite both being part of the same OPEID).

OPEID and UNITID Crosswalk

As a part of this project, the National Center for Education Statistics (NCES) and the Federal Student Aid (FSA) office collaborated to help create a crosswalk between IPEDS IDs (UNITID) and FSA IDs (OPEID). This is a map between the two systems' differing definitions of an institution. It includes the history of institutions going back to the 2000-01 academic year, including new locations, changes in affiliation, and closures. The OPEID/UNITID map across years represents this crosswalk.

The crosswalks were created by combining FSA Postsecondary Education Participants System (PEPS) data with data IPEDS reporting units provided to IPEDS about which 8-digit OPEIDs were included in their IPEDS reporting. The creation of the crosswalks is a multi-step process, beginning with defining the universe of Title IV entities via PEPS. This list was then electronically matched to IPEDS using available directory information and the mapping provided by IPEDS reporting units. Any remaining entities that were not electronically matched were hand-compared to IPEDS using name and address information. Following the manual match, entities that were still not paired with an IPEDS reporting unit were reported as "No Match".

Some IPEDS reporting units did not have a match amongst the set of Title IV entities from PEPS. These are included in the "UNITIDs Not Matched" tab of each crosswalk. These reporting units are either non-Title IV in IPEDS (but may report to IPEDS voluntarily) or were Title IV in IPEDS but no longer Title IV in PEPS.

Currently Operating

Boolean

Institutions that are noted as "currently operating" by FSA are noted with a 1; institutions that are not currently operating (i.e. institutions that were Title IV participating but no longer participate in Title IV programs by disbursing aid or through deferments, or institutions that

have closed[3] or merged) are noted with a 0. This data element is included only in the latest Scorecard data file.

About the Institution

A number of the elements produced in Scorecard and its API provide basic descriptive information about the institution in question. These include: identifiers, location, degree type and profile, programs offered, and the academic profile of students enrolled. Most of these elements are available through IPEDS.

Dev-category	**school**
Name	**String**
	The institution's name (INSTNM), as reported in IPEDS.
Alias	**String**
	Variations on the institution's name (ALIAS), e.g., acronyms, that also identify the institution, as reported in IPEDS.
Location	String
	The institution's location, as reported in IPEDS (ADDR, CITY, STABBR, ZIP). This is also reported using latitude and longitude (float; LATITUDE and LONGITUDE). Note that several locations (8-digit OPEIDs) of an institution may report to IPEDS as a single entity. The locating elements referenced here correspond to the 8-digit OPEID that is listed as the primary entity for the group reporting jointly to IPEDS.
URLs	String
	Each institution reports the URL of its homepage (INSTURL) and the URL of its net price calculator (NPCURL). These URLs are included only in the latest Scorecard data file, and represent the value reported in IPEDS.

[3] Title IV institution closures are identified through PEPS. Non-Title IV institutions may not report closures to PEPS, so such institutions are considered "closed" if they do not report to IPEDS in the year following the establishment of the Scorecard universe (e.g., 2018-19 non-Title IV Scorecard institutions are considered closed if they are not part of the 2019-20 IPEDS collection).

Main Campus/Branch Boolean and Integer

The main campus column (MAIN[4]) identifies whether the institution's IPEDS-derived data elements represent the main campus of the institution or not, where 1 is a main campus and 0 is not[5]. The branch campus column (NUMBRANCH) identifies the number of branch campuses at that institution.

Title IV Eligibility Type Integer

Institutions are classified into categories corresponding to their Title IV participation: Participates in Title IV (OPEFLAG=1), Branch campus of a Title IV participating institutions (OPEFLAG=2), Deferment-only Title IV participant (OPEFLAG=3), and Non-title IV (OPEFLAG=5, 6, 7, and 8).

Accrediting Agency String

The institutional accreditor[6] for each institution is identified (ACCREDAGENCY) to facilitate analysis by accreditor. The PEPS code (ACCREDCODE) corresponding to the accreditor is also included in the data file. These data are maintained by FSA and are included only in the latest Scorecard data file.

Degree Type Integer

There are several elements that identify the degree/certificate profile of the institution. Highest award (HIGHDEG) identifies the highest award level conferred at the institution[7]. Level of institution (ICLEVEL) conveys the highest level of award offered at the institution: 4-year, 2-year, or less-than-2-year. This designation differs from the highest degree element in that it is based on an institution's reported offerings, rather than on degree or certificate completions. For example, an institution

[4] Beginning in 2017-18, the main-campus indicator algorithm was modified for cases where a single OPEID is associated with multiple IPEDS UNITIDs. The new calculation allows MAIN to be used to select a single record for each 6-digit OPEID grouping.

[5] Main campus status is derived from the institution's OPEID. Any institution with an OPEID that begins with "0" and ends with "00" is classified as the main campus.

[6] In a small number of instances where the FSA data center does not list an active primary institutional accreditor, a NULL value is assigned.

[7] HIGHDEG is the highest category of award conferred by the institution, in descending order of graduate degree/certificate, bachelor's degree, associate's degree, and certificate, calculated from the IPEDS Completions component. The values are limited to award categories where the institution conferred at least one first-major award. IPEDS Completions data contain the number of awards conferred during the 12-month period ending June 30 prior to the IPEDS collection year. E.g., the 2018-19 IPEDS collection counts awards conferred from July 1, 2017 to June 30, 2018.

may offer a bachelor's degree program but only infrequently have students that earn the degree. The level of institution element would consistently classify this institution as 4-year, but the highest degree element would potentially vary from year-to-year. Predominant undergraduate award (PREDDEG) identifies the type of award that the institution primarily confers[8]; for instance, an institution that awards 40 percent bachelor's degrees, 30 percent associate degrees, and 30 percent certificate programs would be classified as predominantly bachelor's degree awarding. Institution award (SCH_DEG) uses the predominant undergraduate award calculations at the six-digit OPEID level but recodes missing elements (0=N/A) and predominantly graduate schools (4=Graduate) according to the program length reported to NSLDS.

Highest award, level of institution, and predominant undergraduate award are all derived from IPEDS.[9]

Public/Private Nonprofit/ Private For-Profit

String

Two versions of this element are included in the data file. CONTROL is reported to IPEDS by the institution, and CONTROL_PEPS is derived directly from PEPS. Both indicators identify whether the institution's governance structure is public, private nonprofit, or private for-profit.

Carnegie Classifications

String

The Carnegie Foundation classifies institutions in several ways. These data include the basic classification (CCBASIC), the undergraduate profile (CCUGPROF), and the size and setting classification (CCSIZSET). These data are included only in the most recent Scorecard data file and represent the 2021 Carnegie Classifications.

[8] If there are any undergraduate awards, PREDDEG is assigned based on the undergraduate award category with the largest number of awards (ties going to the higher level). If there are no awards reported for an institution (e.g., new institutions), PREDDEG is assigned by using the IPEDS institutional category (INSTCAT), which maps "Degree-granting, graduate with no undergraduate degrees" to the graduate PREDDEG, "Degree-granting, primarily baccalaureate or above" to the Bachelor's PREDDEG category, both "Degree-granting, not primarily baccalaureate or above" and "Degree-granting, associates and certificates" to the associate's PREDDEG category, and "Nondegree-granting, sub-baccalaureate" to the certificate's PREDDEG category.

[9] Institution award is used in the calculation of NSLDS completion and transfer rates. Since these completion and transfer rates are not calculated in the most recent data file, institution award is also not calculated in the most recent data file.

Special Mission or Religious Affiliation

String

Institutions that are identified as minority-serving institutions, including (for this purpose) Historically Black Colleges and Universities and Tribal Colleges and Universities, are flagged with an indicator (HBCU=Historically Black Colleges and Universities; PBI=Predominantly Black Institutions; ANNHI=Alaska Native-/Native Hawaiian-serving Institutions; TRIBAL=Tribal Colleges and Universities; AANAPII=Asian American-/Native American-Pacific Islander-serving Institutions; HSI=Hispanic-serving Institutions; NANTI=Native American Non-Tribal Institutions). With the exception of a small number of HBCU's, identified institutions are classified by their eligibility to apply for the Department's minority-serving institution program grants.

Minority-serving institution indicators (except HBCU) are derived from the Eligibility Matrix[10] constructed by the Department. HBCU status is not determined by the eligibility matrix. Instead, HBCU status is determined using IPEDS.[11]

Institutions are also identified by their religious affiliation (RELAFFIL), including not reported or no affiliation. These data are reported by institutions directly to IPEDS.

Data on special missions are provided only in the latest Scorecard data file.

Distance-Only

Boolean

Institutions are identified as distance education-only (DISTANCEONLY) if all their programs are available only via distance education. Users may find specific programs of study that are offered as distance education programs through the CIP variables (e.g. CIP01ASSOC is an online-only program if it is identified as a 2)[12]. The values of these elements within each Scorecard data file are derived from data reported to IPEDS. These data are not reported prior to 2011-12.

[10] The Department publishes the Eligibility Matrix at https://www2.ed.gov/about/offices/list/ope/idues/eligibility.html#tips.

[11] The IPEDS HBCU status designation was selected to allow a small number of historically black institutions that do not meet the criteria in the eligibility matrix for technical reasons to be properly included as HBCUs.

[12] Program designations at the 2-digit CIP level are aggregations of individual 6-digit CIP code offerings. If at least one 6-digit CIP code in the 2-digit group is offered (or is offered through an exclusively distance education modality), then the full 2-digit group is designated as offered (or offered through distance education).

Institution Revenues/Expenses Integer

The data files include several elements related to institutional finance. The net tuition revenue per full-time equivalent (FTE) student (TUITFTE) uses tuition revenue minus discounts and allowances, and divides that by the number of FTE undergraduate and graduate students. Instructional expenditures per FTE student (INEXPFTE) uses instructional expenditures divided by the number of FTE students. Financial data within each Scorecard data file match to the most recent fiscal year ending prior to December 31 of the IPEDS collection used as a source for that Scorecard file. FTE enrollment covers the 12-month period ending June 30 prior to the IPEDS collection year.

The average faculty salary (AVGFACSAL) produces the average faculty salary per month, by dividing the total salary outlays by the number of months worked for all full-time, nonmedical instructional staff. “Months worked” was grouped for reporting prior to the 2012-13 academic year, and as a result the value for the 9-10 months group is estimated as 9.5 months and the value for the 11-12 months group is estimated as 11.5 months during that period. Values prior to the 2003-04 academic year are limited to degree-granting institutions for consistency with values in subsequent years. Average faculty salaries cover the full academic year corresponding to the relevant IPEDS collection year.

Heightened Cash Monitoring 2 Boolean

The Department places institutions on a Heightened Cash Monitoring (HCM) payment method to provide additional oversight of cash management. HCM2 is the type of HCM that indicates more serious financial or federal compliance issues. These data are maintained by FSA

Academics

This information describes the types of academic offerings available at each institution.

Dev-category **academics**

Academic Areas Offered Integer

The Classification of Instructional Programs (CIP) provides a taxonomy for different academic disciplines. Two types of CIP data are included in these institution-level data files. The first set (PCIP[01-54]) provide the percentage of degrees awarded in each two-digit CIP code category of academic areas. The second set (CIP[01-

54][CERT1/CERT2/ASSOC/BACHL/CERT4]) identifies which academic areas are offered at which levels and whether or not specific offerings are available through distance education. [13] These CIP elements within each Scorecard data file are derived from the IPEDS Completions component. Reported awards cover the 12-month period ending June 30 prior to the IPEDS collection year. The College Scorecard data files that provide information about specific fields of study within institutions offer a more granular picture (at the four-digit CIP code level) of what fields of study are available at each institution. For more information on these data, please see the technical documentation for field of study data files.

Number of Programs Offered and Six Largest Programs Integer and String

In addition to the two-digit CIP code field of study data, information is available on the number of programs offered (at the 6-digit CIP code level; PRGMOFR) and six largest programs of study at institutions that primarily follow a program-based or continuous enrollment calendar system. The data identifying the programs (CIPCODE[1-6] and CIPTITLE[1-6]) are provided at the six-digit CIP code level. Also provided is the number of months typically needed to complete the full program of study for each program (MTHCMP[1-6]). These data are reported in the IPEDS Institutional Characteristics component. Costs associated with each program are available as well (see the Costs section of this documentation).

Admissions

This information describes the admissions rate and SAT/ACT scores of students.

Dev-category **admissions**

Open Admission Policy Integer

Some colleges accept any students that apply or have minimal requirements for admission (such as a high school diploma or equivalent). Such colleges are considered to have an open admission policy. Colleges that have an open admission policy for all or most entering first-time undergraduates report this policy to IPEDS and do

[13] Distance education program: A program for which all the required coursework for program completion is able to be completed via distance education courses. An institution may offer distance education courses without offering full programs. Such institutions are not identified in College Scorecard.

not report on other admissions considerations or test scores. OPENADMP captures this information for the Scorecard.

Admission Rate Float

Colleges report to IPEDS their Fall admissions rate, defined as the number of admitted undergraduates divided by the number of undergraduates who applied. For institutions with multiple branches, ADM_RATE includes the admissions rate at each campus, while ADM_RATE_ALL represents the admissions rate across all campuses, defined as the total number of admitted undergraduates across all branches divided by the total number of undergraduates who applied across all branches. ADM_RATE_SUPP provides the same information as ADM_RATE, but limited to institutions with cohort sizes of 30 or more. Please note these are also often referred to as *acceptance rates.*

SAT and ACT Scores Float

Test scores of admitted students are not reported for all institutions but may help students to find an institution that is a good academic match. Our files include the 25th and 75th percentiles of SAT reading (SATVR* for _25 and _75), writing (SATWR* for _25 and _75), math (SATMT* for _25 and _75); and of ACT English (ACTEN* for 25 and 75), writing (ACTWR* for 25 and 75), math (ACTMT* for 25 and 75), and composite scores (ACTCM* for 25 and 75). We also have derived midpoints[14] of the ACT (ACT*MID for CM, EN, MT, and WR) and SAT (SAT*MID for VR, MT, WR, but not composite) scores; however, note that institutions do not report those midpoints to IPEDS. SAT and ACT data are available from 2001-02 on; however, SAT writing scores are available only from 2006-07 to 2015-16, and ACT writing data are available only from 2008-09 to 2015-16. In March 2016 the College Board changed the format of the SAT. For the 2016-17 Scorecard data, the SAT scores reported were under the old (pre-March 2016) scoring system; institutions were instructed to convert any scores received under the new score ranges using concordance tables provided by the College Board. Starting with the 2017-18 Scorecard data, the SAT scores reported were under the new (post-March 2016) scoring system; institutions were instructed to convert any score received under the old score ranges using concordance tables provided by the College Board.

[14] Midpoints refers to the midpoint of the 25th and 75th percentile

In addition to the scores themselves, Scorecard also supplies an indicator of how test scores are considered in the admissions process. ADMCON7 indicates if test scores are required, recommended, considered but not required, or neither required nor recommended.

Costs

Information about the costs to students of an institution can provide important context for students and families as they seek to evaluate the tradeoffs of access, affordability, and outcomes. The elements in this category are elements from IPEDS.

Dev-category **cost**

Average Cost of Attendance, Tuition and Fees Integer

The average annual cost of attendance includes tuition and fees, books and supplies, and living expenses for all full-time, first-time, degree-/certificate-seeking undergraduates who receive Title IV aid. These data are available separately for academic year institutions (COSTT4_A) and for program-year institutions (COSTT4_P). For institutions primarily following an academic year-based calendar system, the element is representative of full-time, first-time undergraduate student costs for an entire academic year. For institutions primarily following a program-based or continuous enrollment calendar system, the cost of attendance for full-time, first-time students is provided for the program with the largest enrollment at the institution. Reported program costs cover the full length of the program, regardless of program length. However, when the full length of the program is greater than the institution's academic year (as recorded on their program participation agreement with FSA), COSTT4_P is annualized to correspond to the length of the institution's academic year. For both elements, expenses by living arrangement (on-campus, off-campus independent, or off-campus with family) are combined via a weighted average according to the distribution of full-time, first-time students utilizing those options at the institution. This combined figure is input into the overall calculation of these two elements. Data are not reported prior to 2009.

Additionally, the cost data include the tuition and required fees, books and supplies, and living expenses estimated by the institution. Tuition and required fees are provided for in-district students (TUITIONFEE_IN), out-of-state students (TUITIONFEE_OUT), and program-year institutions (TUITIONFEE_PROG). For academic year institutions, tuition and fees are

calculated for full-time, first-time students. For non-academic-year (program-year or continuous enrollment) institutions, tuition and fees represent those charged to full-time, first-time undergraduates for the largest program at the institution, regardless of program length. Some institutions have different tuition and fees for in-state students that are not reflected in this metric. Data are not reported prior to 2000. Books and supplies and living expense estimates are provided for full-time, first-time students at academic year institutions (BOOKSUPPLY, ROOMBOARD_ON, OTHEREXPENSE_ON, ROOMBOARD_OFF, OTHEREXPENSE_OFF, OTHEREXPENSE_FAM).

All cost elements are derived from data reported to the IPEDS Institutional Characteristics and Student Financial Aid (SFA) components. Institutional Characteristics data cover the academic year corresponding to the IPEDS collection year. SFA component data for institutions primarily following an academic year-based calendar system cover the academic year (the period generally extending from September to June; usually equated to 2 semesters or trimesters, 3 quarters, or the period covered by a 4-1-4 calendar system) prior to the IPEDS collection. For institutions primarily following a program-based or continuous enrollment calendar system, SFA data cover the 12-month period ending June 30 prior to the IPEDS collection year.

For institutions primarily following a program-based or continuous enrollment calendar system, additional data on the costs of the six largest programs are provided for the full program (CIPTFBS[1-6]) and estimated for a single year of study (CIPTFBSANNUAL[1-6]). Costs include tuition, fees, books, and other supplies. The annualized cost data are calculated from the full program costs, the length of each program, and the length of the institution's academic year as reported to IPEDS. Information on program identification can be found in the Academics section of this documentation.

Average Net Price

Integer

There are several elements in the dataset that describe the average net price, derived from the full cost of attendance (including tuition and fees, books and supplies, and living expenses) minus federal, state, and institutional grant/scholarship aid, for full-time, first-time undergraduate Title IV-receiving students. Average net price (NPT4_* for _PUB [public colleges; for public institutions, this metric is limited to those undergraduates who pay in-state tuition] and _PRIV [private colleges]) includes a weighted average of all full-time, first-time

undergraduate Title IV-receiving students[15]; whereas average net price by income quintile (NPT41_*, NPT42_*, NPT43_*, NPT44_*, and NPT45_* for _PUB and _PRIV, as well as _PROG [institutions primarily following a program-based or continuous enrollment calendar system] and _OTHER [other academic calendar institutions]) provides the net price separately for each income bracket for those students. Income quintiles[16] are: (1) $0-$30,000; (2) $30,001-$48,000; (3) $48,001-$75,000; (4) $75,001-$110,000; and (5) $110,000+. We also included several combinations of net price quintiles, which rely on a weighted average of those quintiles (i.e. NPT4_048_* is the net price for the first two income brackets, $0-$48,000). Net price data are not reported prior to 2009.

All net price elements are derived from data reported to the IPEDS Institutional Characteristics and Student Financial Aid (SFA) components. Institutional Characteristics data cover the academic year corresponding to the IPEDS collection year. SFA component data for institutions primarily following an academic year-based calendar system cover the academic year (the period generally extending from September to June; usually equated to 2 semesters or trimesters, 3 quarters, or the period covered by a 4-1-4 calendar system) prior to the IPEDS collection.

For institutions primarily following a program-based or continuous enrollment calendar system, this data point represents the largest program at the institution (instead of the institution as a whole). In cases where the program length is 1 academic year or shorter, the average annual cost for the largest program represents the full program costs. When the program length is longer than 1 academic year, the average annual cost for the largest program represents the average costs for a single academic year of study. In this situation, advertised prices are annualized using the school's reported academic year in credit/contract hours. Average grant and/or scholarship aid is reported for an annual period and is not adjusted.

[15] Scorecard derivation of net price based on this weighted average is different than the methodology used in other presentations of net price (e.g. College Navigator)

[16] Income values in nominal dollars (not adjusted for inflation)

Students and Staff

Several elements identify demographic and other details about the student body and the staff of the institution. Some of the elements are available through IPEDS and others were produced using NSLDS.

Dev-Category	**student**
Number of Undergraduate Students	Integer
	This element (UGDS) includes the number of degree/certificate-seeking undergraduates enrolled in the fall, as reported in the IPEDS Fall Enrollment component. In 2001-02, degree-/certificate-seeking status was not collected, so data for that year (UG) are reported as a separate element. IPEDS Fall Enrollment data are reported as of the official institutional census date, or October 15 of the IPEDS collection year, whichever is earlier.
Undergraduate Student Body by Race and Gender:	Float
	These data are reported by institutions to IPEDS in the Fall Enrollment component; and rely on students' self-reported race and gender data, as collected by the institution. This includes the total enrollment of undergraduate, degree-seeking students, based on fall enrollment, who are: men (UGDS_MEN), women (UGDS_WOMEN), white (UGDS_WHITE), black (UGDS_BLACK), Hispanic (UGDS_HISP), Asian (UGDS_ASIAN), American Indian/Alaska Native (UGDS_AIAN), Native Hawaiian/Pacific Islander (UGDS_NHPI), two or more races (UGDS_2MOR), non-resident aliens (UGDS_NRA), and race unknown (UGDS_UNKN). Note that prior to 2008-09, the categories were different, and are designated as separate elements. Additionally, in 2000-01, degree-/certificate-seeking status was not collected, so data for that year are also reported as separate elements. IPEDS Fall Enrollment data are reported as of the official institutional census date, or October 15 of the IPEDS collection year, whichever is earlier.

Undergraduate Students by Part-Time/Full-Time Status Float

This element includes the proportion of degree/certificate-seeking undergraduates enrolled part time in the fall term, as calculated from IPEDS Fall Enrollment data (PPTUG_EF). Note that this metric does not apply during 2000-01, when degree-/certificate-seeking status was not collected; data for that year are reported separately (PPTUG_EF2) and calculated from the number of part-time undergraduates divided by the total number of undergraduates. Both sets of data include non-first-time undergraduates.

IPEDS Fall Enrollment data are reported as of the official institutional census date, or October 15 of the IPEDS collection year, whichever is earlier.

Undergraduate Students by Family Income Float

The data files include the IPEDS count of students in each income bracket that are covered by the average net price by income quintile measurements (NUM[1-5]_[PUB or PRIV]). These elements report the number of full-time, first-time, degree/certificate-seeking undergraduates who received Title IV aid in each income bracket from the IPEDS Student Financial Aid (SFA) component. Separate metrics are calculated for public institutions and private institutions. This metric is not available prior to the 2009-10 academic year. SFA component data for institutions primarily following an academic year-based calendar system cover the academic year (the period generally extending from September to June; usually equated to 2 semesters or trimesters, 3 quarters, or the period covered by a 4-1-4 calendar system) prior to the IPEDS collection. For institutions primarily following a program-based or continuous enrollment calendar system, SFA data cover the 12-month period ending June 30 prior to the IPEDS collection year.

Undergraduate Student Body by Age Float

These data (UG25ABV) are reported by institutions every other year to IPEDS via the Fall Enrollment component. This element identifies the share of students enrolled as of the institutions official fall census date (or October 15 of the IPEDS collection year, whichever is earlier) who are ages 25 and over.

Staff by Race and Gender:

Float

These data are reported by institutions to IPEDS in the Human Resources component; and rely on staffs' self-reported race and gender data, as collected by the institution. This includes the share of instruction, research, and/or public service staff who are: men (IRPS_MEN), women (IRPS_WOMEN), white (IRPS_WHITE), black (IRPS_BLACK), Hispanic (IRPS_HISP), Asian (IRPS_ASIAN), American Indian/Alaska Native (IRPS_AIAN), Native Hawaiian/Pacific Islander (IRPS_NHPI), two or more races (IRPS_2MOR), non-resident aliens (IRPS_NRA), and race unknown (IRPS_UNKN). Note that prior to 2008-09, the categories were different, and the Asian and Pacific Islander category during that period is assigned to IRPS_ASIAN.

IPEDS Human Resources data are reported as of November 1 of the IPEDS collection year.

Financial Aid

Federal financial aid, including Pell Grants and federal student loans, helps many students, particularly low-income students, access and afford a higher education. Data on the amount of debt that students can expect to borrow, and the loan performance of former students, may help students find the best option in their price range.

Dev-Category **student**

Percent of Full-time, First-time Undergraduates Receiving Federal Loans

Float

This element (FTFTPCTFLOAN), as reported in the IPEDS Student Financial Aid (SFA) component, shows the share of full-time, first-time degree/certificate-seeking undergraduate students who received federal loans in a given year. It can provide important context to figures related to debt, repayment, and non-repayment. This figure may be influenced by the eligibility for federal loans and the extent to which students apply for federal loans, as well as by the cost of the programs. In particular, many community colleges are sufficiently low-cost to have low federal loan borrowing rates; and it may be difficult to compare borrowing behaviors for those institutions. These data are not available prior to 2009-10.

IPEDS SFA component data for institutions primarily following an academic year-based calendar system cover the academic year (the period generally extending from September to June; usually equated to 2 semesters or trimesters, 3 quarters, or the period covered by a 4-1-4 calendar system) prior to the IPEDS collection. For institutions primarily following a program-based or continuous enrollment calendar system, SFA data cover the 12-month period ending June 30 prior to the IPEDS collection year.

Percentage of Full-time, First-time Pell Students

Float

This element (FTFTPCTPELL), pulled from the IPEDS Student Financial Aid (SFA) component, shows the share of full-time, first-time degree/certificate-seeking undergraduate students who received Pell Grants in a given year. This is an important measure of the access an institution provides to low-income students. However, it may not capture all low-income students. Students who are undocumented immigrants or foreign nationals are not eligible to receive Pell Grants, and some low-income students may not have completed the Free Application for Federal Student Aid (FAFSA) to receive federal aid, but those students may have similar financial circumstances to Pell recipients, or may be just on the other side of Pell eligibility, creating a cliff effect. Additionally, in some states (such as California), state financial aid may be sufficient to cover costs at community colleges, in particular; so, those students may not seek or receive a Pell Grant. These data are not available prior to 2008-09.

IPEDS SFA component data for institutions primarily following an academic year-based calendar system cover the academic year (the period generally extending from September to June; usually equated to 2 semesters or trimesters, 3 quarters, or the period covered by a 4-1-4 calendar system) prior to the IPEDS collection. For institutions primarily following a program-based or continuous enrollment calendar system, SFA data cover the 12-month period ending June 30 prior to the IPEDS collection year.

Cumulative Median Student Debt[17]

Integer

[17] Prior to the 2017-18 merged data file, the count of students in the debt cohort was provided in two metrics (DEBT_N and CUML_DEBT_N). Starting with the 2017-18 data file, these data are only provided in DEBT_N.

This is the median loan debt accumulated at the institution[18] by all student borrowers of federal loans[19] who separate (i.e., either graduate or withdraw) in a given fiscal year, measured at the point of separation (DEBT_MDN)[20]. More specifically, the measure represents the sum of all undergraduate federal loans over students' college education at the institution for which the median debt is reported for —e.g., if a student receives a federal loan for $2,000 for each of eight semesters at one institution, their cumulative debt is recorded as $16,000 for that institution.[21]

These data are available for all borrowers at the institution, as well as disaggregated by completion status (GRAD_DEBT_MDN for students who completed and WDRAW_DEBT_MDN for students who withdrew without completing[22]); by FAFSA family income[23] (LO_INC_DEBT_MDN = $0-$30,000; MD_INC_DEBT_MDN = $30,001-$75,000; and HI_INC_DEBT_MDN = $75,001+); by dependent (DEP_DEBT_MDN) and independent (IND_DEBT_MDN) status; by Pell status (PELL_DEBT_MDN for students who ever received a Pell Grant and NOPELL_DEBT_MDN for students who never received a Pell Grant); by gender

[18] Note that a single borrower's loans could be in multiple institutional median debt calculations with only the loans accrued at each separate institution included in that respective institution's median debt calculation

[19] PLUS loans (federal loans to parents of undergraduate students) are not included in the cumulative debt metrics. The methodology for this metric changed starting with the merged 2017-18 data file. In contrast to the previous methodology, the updated version excludes Perkins loans in the calculation.

[20] Median debt is adjusted for any refunds that were posted to NSLDS by the date the data were drawn. Any refunds posted after that date would not be reflected.

[21] The methodology for this metric changed starting with the merged 2017-18 data file. In contrast to the previous methodology, the updated version excludes all undergraduate loans disbursed post-separation at the evaluated institution.

[22] The methodology for determining completion and withdrawal status for this disaggregation changed starting with the merged 2017-18 data file. In the previous version, this was based on the completion or withdrawal status upon separation from the school. In the updated version, a borrower is placed in the completed category if that borrower completed at *any time* prior to the separation at the evaluated institution (even if that separation event was a withdrawal event). In addition, if the student subsequently received a graduate-level loan, they would also be categorized as a completer. In some cases, the counts of borrowers who completed and withdrew may sum to a number higher than the aggregate number of borrowers because a borrower may have initially withdrawn and then completed in which case they would be counted in both disaggregated cohorts

[23] These calculations are based on nominal dollar values (not adjusted for inflation).

(FEMALE_DEBT_MDN and MALE_DEBT_MDN); and by first-generation status (FIRSTGEN_DEBT_MDN and NOTFIRSTGEN_DEBT_MDN).[24]

At institutions where large numbers of students withdraw before completion, a lower median debt level could simply reflect the lack of time that a typical student spends at the institution. Therefore, the Department uses the typical debt level for students who complete (GRAD_DEBT_MDN_SUPP or GRAD_DEBT_MDN10YR_SUPP for the debt level expressed in monthly payments[25]) on the consumer website. Additionally, this measure can be placed in context by looking at the borrowing rate of students at the institution (FTFTPCTFLOAN; see above); at institutions where few students borrow, the numbers may represent outliers.

Data are produced for rolling two-year pooled cohorts. For these variables, years refer to fiscal years (e.g., fiscal year 2018 begins on October 1, 2017, and ends September 30, 2018).

Cumulative loan debt data are also provided by field of study; more information on the field of study data is available in the technical documentation for the field of study data files.

Cumulative Median Parent PLUS Debt

Integer

This is the median parent PLUS loan debt accumulated at the institution associated with students who separate (i.e., either graduate or withdraw) in a given award year, measured at the point of separation. Cumulative PLUS loan debt is measured for loans that were originated at the institution where the debt is reported (PLUS_DEBT_INST_MD)[20] representing the sum of all PLUS loans associated with students over the students' college education at the institution for which the median debt is reported. Cumulative PLUS loan debt is separately measured for loans that were originated at any institution the student attended (PLUS_DEBT_ALL_MD)[20]. This metric represents the sum of all PLUS

[24] In the student's FAFSA, if at least one parent's status is "college" then this is coded as non-first generation. The methodology for determining first generation status changed starting with the merged 2017-18 data file. Previously, first generation status used imputation methodology to assign a first-generation or non-first-generation status to students where both parents had unknown statuses. Starting with data in the merged 2017-18 data file, all unknown statuses were combined with first-generation status (instead of imputed to both types of known statuses). In addition, for all other all calculations disaggregated by demographic categories, the methodology changed slightly to look across multiple demographic profiles (instead of just one) to minimize missing statuses subject to imputation.

[25] Calculation based on a 4.99% interest rate

loans associated with students over students' college education at any institution, For example, if a parent received a PLUS loan for $2,000 for each of four semesters their student was enrolled at college A, with the student separating in May 2016, and $1,000 for each of two semesters their student was enrolled at college B, with the student separating in May 2017, the debt is recorded as follows: $8,000 is included in the calculation of PLUS_DEBT_INST_MD and PLUS_DEBT_ALL_MD for college A metrics that cover the 2015-16 award year cohort, $2,000 is included in the calculation of PLUS_DEBT_INST_MD for college B metrics that cover the 2016-17 award year cohort, and $10,000 is included in the calculation of PLUS_DEBT_ALL_MD for college B metrics that cover the 2016-17 award year cohort.

These data are available for all PLUS loans associated with students at the institution, as well as disaggregated (PLUS_DEBT_INST_[*]_MD, PLUS_DEBT_ALL_[*]_MD) by completion status ([COMP]/[NOCOMP]), sex ([MALE]/[NOMALE]), Pell recipient status ([PELL]/[NOPELL]), Direct/Stafford loan recipient status at the institution the metric is associated with ([STAFFTHIS]/[NOSTAFFTHIS]), and Direct/Stafford loan recipient status at any institution/[STAFFANY]/[NOSTAFFANY]).

At institutions where large numbers of students withdraw before completion, a lower median PLUS debt level could simply reflect the lack of time that a typical student spends at the institution. Therefore, the Department uses the typical debt level for students who complete on the consumer website (PLUS_DEBT_INST_COMP_MD_SUPP/PLUS_DEBT_ALL_COMP_MD_SUPP, or expressed in monthly payments[26] as PLUS_DEBT_INST_COMP_MDPAY10_SUPP/ PLUS_DEBT_ALL_COMP_MDPAY10_SUPP). Additionally, this measure can be placed in context by looking at the estimated range for the proportion of students whose parents borrowed a PLUS loan. The range is calculated by dividing the number of Parent PLUS Loan recipients for an award year published in the FSA Data Center Title IV Loan Volume report[27] by the estimated number of degree- or certificate-seeking undergraduate students enrolled during the same award year. The student estimate comes from the IPEDS Fall Enrollment proportion of undergraduate students that are degree- or certificate-seeking multiplied by the IPEDS 12-Month Enrollment count of undergraduates. The resulting Parent PLUS percentage is rounded down to the nearest

[26] Calculation based on a 7.54% interest rate

[27] https://studentaid.gov/data-center/student/title-iv

5% to establish the lower bound (PPLUS_PCT_LOW) and rounded up to the nearest 5% to establish the upper bound (PPLUS_PCT_HIGH). If the start or end of the calculated range was within 1 percentage point of the estimated percentage, that bound was moved an additional 5% lower or higher.

Data are produced for rolling two-year pooled cohorts. For these variables, years refer to award years (e.g., award year 2017-18 begins on July 1, 2017, and ends June 30, 2018).

PLUS loan debt data are also provided by field of study; more information on the field of study data is available in the technical documentation for the field of study data files.

Institutional Loan Portfolio

Long

These metrics provide the outstanding loan balances owed (LP[TYPE]_AMT) and the corresponding number of borrowers associated with those outstanding balances (LP[TYPE]_CNT) at each institutions. The loan programs for which these metrics are calculated are Direct Loans ([TYPE]=STAFFORD), Parent PLUS ([TYPE]=PPLUS), and Grad PLUS ([TYPE]=GPLUS).

Completion and Retention

College completion is associated with other positive outcomes, like finding a job and successfully repaying student loans, and is an important metric for evaluating the experiences of students at the institution. However, both existing and new methods of measuring completion have limitations.

Dev-category **completion**

Completion rates for full-time, first-time students

IPEDS Completion Rates (100 and 150 Percent, and 6- and 8-years after entry) and 150 Percent Transfer Rate:

Float

Currently, institutions report (via the IPEDS Graduation Rates component) on the completion rates for full-time, first-time students who complete within 100 or 150 percent of the expected time to completion (C[100 or 150]_4 for four-year institutions and C[100 or 150]_L4 for less-than-four-year institutions). The 150 percent rates are available disaggregated by race (C150_4_* and C150_L4_*, for _WHITE

[white], _BLACK [black], _HISP [Hispanic], _ASIAN [Asian], _AIAN [American Indian/Alaska Native], _NHPI [Native Hawaiian/Pacific Islander], _2MOR [two or more races], _NRA [non-resident alien], and _UNKN [race unknown]) and by Pell Grant/Direct Subsidized Loan recipient status during the first year of college (C150_4_* and S150_L4_* for PELL [Pell Grant recipients], LOANNOPELL [Direct Subsidized Loan recipients that did not receive a Pell Grant], and NOLOANNOPELL [received neither a Direct Subsidized Loan nor a Pell Grant]). Note that the designation as four-year or less-than-four year for these metrics is based on the IPEDS level of institution (ICLEVEL), not on the predominant award (PREDDEG).

For institutions primarily following an academic year calendar system, the IPEDS completion rates are limited to full-time, first-time students beginning in the fall semester. For institutions primarily following a non-academic year calendar system (program or continuous enrollment), the IPEDS completion rates cover all full-time, first-time students. The exclusion of part-time students, transfer students, and students who do not start during the fall from IPEDS completion rates makes the rates less relevant for those populations of students. Full-time, first-time students make up fewer than half of all college students, or even less in some sectors of institutions (e.g. community colleges). Furthermore, although institutions have the option to report transfer outcomes for full-time, first-time students (TRANS_4 and TRANS_L4), many choose not to. In light of these limitations, in 2015-16, IPEDS began collecting student “outcome” status (completed [OMAWDP8_*], enrolled at another institution [OMENRAP8_*], still enrolled [OMENRYP8_*], or status unknown [OMENRUP8_*]) eight years after entering the institution for four cohorts of students: full-time, first-time (*=FTFT); part-time, first-time (*=PTFT); full-time, not first-time (*=FTNFT); and part-time, not first-time (*=PTNFT). IPEDS also collects award status at 6 years after entry (OMAWDP6_*) for the same four cohorts. Beginning with the 2017_18 data file, outcome data 8 years after entry are also disaggregated by Pell Grant recipient status (OMAWDP8_[PELL, NOPELL]_*, OMENRAP8_[PELL, NOPELL]_*, OMENRYP8_[PELL, NOPELL]_*, and OMENRUP8_[PELL, NOPELL]_*).

IPEDS Graduation Rate component data are cohort-based graduation rates. Four-year institutions[28] report on students that began six years prior to the IPEDS collection year, while less-than-four-year[28] institutions report on students beginning three years prior to the collection year. The varied cohort year starting points allow the measurement at 150% of expected time to occur at the same time for both cohorts (prior to August 31 of the collection year). For full-time, first-time, bachelor's degree-seeking undergraduates, 150 percent of expected time is typically 6 years, and for full-time, first-time, associate's degree-seeking undergraduates it is typically 3 years. For full-time, first-time, certificate-seeking undergraduates, the expected time period varies by the length of the program (for example, 9 months for a certificate with an expected completion time of 6 months).

To reduce variability from year to year, data are also available as pooled completion rates across two years on a rolling basis (C[100 or 150]_4_POOLED, C[100 or 150]_L4_POOLED, C150_[4 or L4]_PELL_POOLED, OM[AWD, ENRA, ENRY, or ENRU]P8_*_POOLED) and OM[AWD, ENRA, ENRY, ENRU]P8_PELL_*_POOLED); among institutions with fewer than 30 students in the combined cohorts, the measure was created based on a four-year cohort instead (see POOLYRS or POOLYRS100, POOLYRS150_PELL, POOLYRSOM_*, or POOLYRSOM_PELL_* for the number of years used in the rolling average). Data are also suppressed for institutions with fewer than 30 students, where the outcome of a single student could dramatically change the rate (C[100 or 150]_4_POOLED_SUPP, C[100 or 150]_L4_POOLED_SUPP, C150_[4 or L4]_PELL_POOLED_SUPP, OM[AWD, ENRA, ENRY, or ENRU]P8_*_POOLED_SUPP, and OM[AWD, ENRA, ENRY, ENRU]P8_PELL_*_POOLED_SUPP).

Pooled data are available only in the latest Scorecard data file. One-year (non-pooled) 150% measures are included for every year (C150_4 and C150_L4). One-year (non-pooled) 100% measures are included beginning with the 1997_98 data file for 4-year institutions (C100_4) and beginning with the 2009_10 data file for less-than-4-year

[28] Institutions are classified as four-year, two-year, and less-than-two year in IPEDS based on the highest level of *program offering*. This metric (ICLEVEL) is included in the College Scorecard data, and differs from the predominant degree or highest degree classifications (PREDDEG and HIGHDEG), which are based on the level of *awards conferred*. Students in the cohort at IPEDS 4-year institutions that are seeking a sub-baccalaureate award are included in the graduation rate calculation, with completion status measured at 150% of normal program completion time for the program they enrolled in (e.g., completion status for students seeking an associate's degree at a four-year institution is typically measured after three years [rather than 6 years]).

institutions (C100_L4). One-year (non-pooled) outcomes 8 years after entry are available beginning with the 2015_16 data.

200 Percent IPEDS Completion Rate

Float

Currently, institutions also report (via the IPEDS 200 Percent Graduation Rates component) on the completion rates within 200 percent of the expected time to completion for full-time, first-time students who begin school in the fall semester at institutions primarily following an academic year calendar system, or all full-time, first-time students from institutions primarily following a non-academic year calendar system (program or continuous enrollment). For four year institutions[28], the rate (C200_4) covers bachelor's degree-seeking students, and for less-than-four-year institutions, the rate (C200_L4) covers students seeking an any award. For full-time, first-time, bachelor's degree-seeking undergraduates, 200 percent of expected time is typically 8 years, and for full-time, first-time, associate's degree-seeking undergraduates it is typically 4 years. For full-time, first-time, certificate-seeking undergraduates, the expected time period varies by the length of the program (for example, 12 months for a certificate with an expected completion time of 6 months). Note that the designation as four-year or less-than-four year for these metrics is based on the IPEDS level of institution (ICLEVEL), not on the predominant award (PREDDEG).

These rates face the same limitations as the 150 percent completion rate above. To reduce variability from year to year, data are also available as pooled completion rates across two years on a rolling basis (C200_4_POOLED and C200_L4_POOLED); among institutions with fewer than 30 students in the combined cohorts, the measure was created based on a four-year cohort instead (see POOLYRS for the number of years used in the rolling average). Data are also suppressed for institutions with fewer than 30 students, where the outcome of a single student could dramatically change the rate (C200_4_POOLED_SUPP and C200_L4_POOLED_SUPP).

Pooled data (C200_4_POOLED and C200_L4_POOLED) are available only in the latest Scorecard data file.

Retention Rate

Float

Available through the IPEDS Fall Enrollment component, retention rate identifies (separately) the share of full-time and part-time students in the prior year, at four-year (RET_FT4 and RET_PT4) and less-than-four-

year institutions (RET_FTL4 and RET_PTL4), who return to the institution after the first year. For four-year institutions, the retention rate covers bachelor's degree-seeking students only; at less-than-four-year institutions, it covers all degree/certificate-seeking students. Note that the designation as four-year or less-than-four year for these metrics is based on the IPEDS level of institution (ICLEVEL), not on the predominant award (PREDDEG).

To reduce variability from year to year, data are also available as pooled retention rates across two years on a rolling basis (RET_FT4_POOLED, RET_PT4_POOLED, RET_FTL4_POOLED, and RET_PTL4_POOLED); among institutions with fewer than 30 students in the combined cohorts, the measure was created based on a four-year cohort instead (see POOLYRSRET_FT and POOLYRSRET_PT for the number of years used in the rolling average). Data are also suppressed for institutions with fewer than 30 students, where the outcome of a single student could dramatically change the rate (RET_FT4_POOLED_SUPP, RET_PT4_POOLED_SUPP, RET_FTL4_POOLED_SUPP, and RET_PTL4_POOLED_SUPP).

Data are not available prior to 2004-05; pooled data (C200_4_POOLED and C200_L4_POOLED) are available only in the latest Scorecard data file.

Outcomes for Title IV Students

NSLDS Completion and Transfer Rates Float

Using data from NSLDS, the Department also produced completion rates that track institutional outcomes for students who receive federal financial aid. Each institution has all possible outcomes reported: share of students who died (DEATH_YR*_RT), completed at the original institution (COMP_ORIG_YR*_RT), transferred and completed at a four-year institution (COMP_4YR_TRANS_YR*_RT), transferred and completed at a two-year institution (COMP_2YR_TRANS_YR*_RT), withdrew from the original institution (WDRAW_ORIG_YR*_RT), transferred and withdrew from a four-year institution (WDRAW_4YR_TRANS_YR*_RT), transferred and withdrew from a two-year institution (WDRAW_2YR_TRANS_YR*_RT), still enrolled at the original institution (ENRL_ORIG_YR*_RT), transferred and is still enrolled at a four-year institution (ENRL_4YR_TRANS_YR*_RT), transferred and is still enrolled at a two-year institution (ENRL_2YR_TRANS_YR*_RT),

status unknown at the original institution (UNKN_ORIG_YR*_RT), transferred to a four-year institution and status is unknown (UNKN_4YR_TRANS_YR*_RT), and transferred to a two-year institution and status is unknown (UNKN_2YR_TRANS_YR*_RT)[29].

Each of those rates is available at two (_YR2_RT), three (_YR3_RT), four (_YR4_RT), six (_YR6_RT), and eight (_YR8_RT) years after entering the institution.

Additionally, each rate is disaggregated for low-income students (LO_INC_* = $0-$30,000 in FAFSA family income), middle-income students (MD_INC_* = $30,001-$75,000 in FAFSA family income), and high-income students (HI_INC_* = $75,001+ in FAFSA family income) [30]; dependent (DEP_*) and independent (IND_*) students; male (MALE_*) and female (FEMALE_*) students; students who ever received a Pell grant (PELL_*) and students who never received a Pell grant (NOPELL_*); students who ever received a federal loan at the measured institution (LOAN_*) and students who never received a loan there (NOLOAN_*); and for first-generation students (FIRSTGEN_*) and not-first-generation students (NOT1STGEN_*).[24]

NSLDS records the first-time students received aid, rather than the first time they enroll. Therefore, students were placed in cohorts based on students' responses to a question on the FAFSA about their grade level; students indicating they were first-year students were placed in the completion cohort for that award year; second-year placed in the previous cohort; and third- and fourth-year students were placed two cohorts prior to reduce the risk of misreporting on the FAFSA.

Based on those cohorts, the Department used the enrollment records reported by institutions to determine the percentage of students completed within a given period. Where institutions had not provided any information, including a completion status, students were recorded as non-completers. Additionally, students who transferred to another institution (i.e., separated from the original institution and subsequently received Title IV aid at another institution) were reported separately.

The Department identified several limitations that led us to exclude these rates from our consumer tool until institutions have another

[29] Four-year institutions and two-year institution designations for these calculations are based on the SCH_DEG variable.

[30] Calculations are based on nominal dollar values (not adjusted for inflation).

opportunity to improve their reporting. Some institutions seem to report completion and withdrawal interchangeably, since both indicate the student is entering repayment. Similarly, because NSLDS is used for administering financial aid and for those purposes separating from the institution only mattered for students with loans who then entered repayment, institutions were not required to report completion status for students who receive only grants prior to 2012, so historical rates reflect exceptionally low completion rates for Pell-only students. Reporting of part-time and full-time status is exceptionally spotty, so it is not currently possible to separate out those students; and, of course, the rates do not include non-Title-IV students and may therefore be unrepresentative of the outcomes of some institutions.

For these variables, years refer to award years (e.g., award year 2017-18 begins on July 1, 2017, and ends June 30, 2018).

Earnings

One of the most common reasons students cite in choosing to go to college is the expansion of employment opportunities. To that end, data on the earnings and employment prospects of former students can provide key information. To measure the labor market outcomes of individuals attending institutions of higher education, data on cohorts of federally aided undergraduate students were linked with earnings data from de-identified tax records and reported back at the aggregate, institutional level. Mean earnings data elements at the institution-level were last updated in the fall of 2018. See Appendix B for information about those metrics.

There are two notable limitations that researchers should keep in mind for all of these metrics. First, research suggests that the variation across programs within an institution may be even greater than aggregate earnings across institutions. For information related to more recent earnings calculations by field of study, please see the technical documentation for field of study data files. Second, the data include only Title IV-receiving students, so figures may not be representative of institutions with a low proportion of Title IV-eligible students. Additionally, the data are restricted to students who are not enrolled (enrolled means having an in-school deferment status for at least 30 days of the measurement year), so students who are currently enrolled in, for example, graduate school at the time of measurement are excluded.

Dev-category	**earnings**
Median Earnings	Integer There are two different cohorts for median earnings measurements. The “entry” cohort, which consists of students

that began their studies at the institution at the same time, and the "exit" cohort, which consists of students that completed a credential at the institution at same time. In both cohorts, median earnings are measured for individuals where were federally aided and who are employed but not enrolled. Earnings are defined as the sum of wages and deferred compensation from all non-duplicate W-2 forms received for each individual (from both full- and part-time employment), plus positive self-employment earnings from Schedule SE.

Median entry cohort earnings (MD_EARN_WNE_P*) are available for each year starting six years after a student enrolls in college, up to 10 years after the student enrolls; enrollment dates are estimated based on FAFSA self-reporting, as with the completion rate cohort construction described above.

Beginning in the 2018_19 data file, median entry cohort earnings are also available disaggregated by FAFSA family income[31] (MD_EARN_WNE_INC1_P*, MD_EARN_WNE_INC2_P*, and MD_EARN_WNE_INC3_P*); by dependent status (for dependents, MD_EARN_WNE_INDEP0 _P*; and for independents, MD_EARN_WNE_INDEP1_P*); and by gender (for female, MD_EARN_WNE_MALE0_P*; and for male, MD_EARN_WNE_MALE1_P*).

Median exit cohort earnings (MD_EARN_WNE_*YR) are available 1 and 4 years after completion, with the measurement period beginning in the calendar year following completion.

Earnings included in the 2011_12 and prior Scorecard data files are inflation adjusted to 2014 dollars using the Consumer Price Index for all Urban Consumers (CPI-U). Earnings included in the XXXX_YY data file are inflation adjusted to XXXX+3 dollars using the CPI-U for the 2012_13 to 2014_15 Scorecard data files. Beginning with the 2018_19 data file, earnings included in the XXXX_YY data file are adjusted to XXXX+2 dollars. For example, earnings included in the 2014_15 Scorecard data file are

[31] For earnings variables that are disaggregated by FAFSA family income tercile (low-income: $30,000 or less; middle-income: $30,001-$75,000; and high-income: $75,001+), family income was adjusted for inflation prior to grouping by tercile for calculations measured in calendar years 2014 and prior (i.e. values in the merged_2013-14 data file and prior data files).

inflation adjusted to 2017 dollars, and the earnings included in the 2018_19 data file are adjusted to 2020 dollars.

While earnings are measured based on calendar years, cohort years for earnings variables are based on award years.

Threshold Earnings Float

In the 2017_18 and 2018_19 data files, threshold earnings calculations are included in the Scorecard data that describe the number of undergraduate completers earning more than 150% of the poverty threshold for a single individual (CNTOVER150_xYR, where x=number of years after completion). These data are calculated for pooled award year cohorts of completers at the institution, and the earnings values compared to 150% of the poverty threshold have been inflation adjusted to correspond to the year after closing year of the data file they are included in (e.g, adjusted to 2020 dollars if included in the 2018_19 data file). The cohort of completers that were working and not enrolled (COUNT_WNE_xYR) is also provided for context. These data are calculated 1 and 3 years after completion.

Threshold earnings are also calculated to provide the fraction of former undergraduate students who earned more than the median wage of workers ages 25 to 34 ($32,000 in 2021 dollars) that self-identify as a high school graduate (by indicating high school completion was their highest level of education), as measured 6, 8, and 10 years after entering the evaluated institution (GT_THRESHOLD_P[6, 8, or 10]) and also measure 1 and 4 years after completing a credential (GT_THRESHOLD_[1 or 4]YR. The denominator of this fraction for each measure only includes members of the cohort who were working and not enrolled in school during in the measurement year.

This comparison group of self-identified high school graduates should be viewed with caution - the group may include individuals who have completed postsecondary education and/or training including apprenticeships and/or industry certifications. In addition, observations of self-identified high school graduates are not perfectly aligned with the entry or exit cohorts of students measured. Some high school graduates may have much more work experience than students who recently

were enrolled in college. For example, high school graduates ages 25 to 34 may have been in the workforce for as many as 16 years, whereas students completing a four-year degree within 150 percent of time may just be looking for entry-level work 6 years after initial enrollment or have no more than 4 years of post-completion experience 10 years after initial enrollment.

Beginning with the threshold earnings values in the 2018_19 data file and median earnings values in the 2019_20 data file, to protect the confidentiality of tax-payer information, the IRS perturbed these data with a differentially private algorithm to add noise to the data .[32] Perturbed data include the count of non-enrolled workers in an earnings cohort, and the count of non-enrolled workers earning above 150 percent of the poverty threshold, and the count used to calculated the fraction of former students earning more than the median high school graduate. The amount of noise introduced for the most recent cohort calculations is similar to the amount in prior year calculations where most perturbed counts do not differ from actual counts, less than a quarter of perturbed counts differ by actual counts by an absolute value of one, and less than one percent differ by an absolute value of two or three.

Perturbed data also include median calculations. For the entry cohort median earnings, percentage differences compared to undisclosed IRS data vary by the amount of time elapsed since entry:

Table 1. Percentage distribution of calculated median, by years after entry and percentage difference range for cohorts with earnings values measured in calendar years 2019 and 2020.

Difference compared to unperturbed value	Years after entry		
	6 years	8 years	10 years
Less-than 1 percent	73%	75%	73%
1 to 4 percent	23%	22%	24%
5 to 9 percent			
Percent	4%	3%	3%
Median difference	$ 1,339	$ 1,434	$ 1,598

[32] For a small number of observations, perturbation may result in inconsistencies such as the count of workers earning above the threshold exceeding the count of all workers in the cohort, or the median earnings of all subcategories being less (or greater) than the overall median earnings.

Table 2. Percentage distribution of calculated median, by years after completion and percentage difference range for cohorts with earnings values measured in calendar years 2019 and 2020.		
Difference compared to unperturbed value	Years after completion	
	1 year	4 years
Less-than 1 percent	72%	73%
1 to 4 percent	25%	24%
5 to 9 percent		
Percent	3%	3%
Median difference	$ 1,121	$ 1,366

Due to the noise in these data, users should exercise caution in calculating and using provided percentages, particularly in smaller cohorts where absolute differences of just one or two in either the numerator and/or denominator can more substantially change percentage estimates.

Earnings data for completers, including median earnings values, are also provided by field of study; more information on the field of study data is available in the technical documentation for the field of study data files.

Repayment

To provide a sense for the debt burden of attending college and the loan performance metrics for each institution, we produced several elements using NSLDS. These elements can provide useful information for students and families concerned about borrowing for college and interested in seeing borrowers' behavior after they leave the institution.

Dev-category **repayment**

Cohort Default Rate Float

Cohort default rates are produced annually[33] as an institutional accountability metric; institutions with high default rates may lose access to federal financial aid. The three-year cohort default rate (CDR3) represents a snapshot in time. For example, FY 2016 rates were calculated using the cohort of borrowers who entered repayment on their federal student loans between October 1, 2015 and September 30, 2016, and who defaulted before September 30, 2018. Three-year CDR data are not

[33] Note that each annual release of the cohort default rate contains default information from three cohorts – the most recent (new) cohort, and updates/revisions to the immediately prior two cohorts. For example, the FY 2014 cohort default rate release contained the FY2014 3-year default rate and updates to the FY2013 and FY2012 3-year default rates. These updated default rates are incorporated into the appropriate Scorecard data file.

available prior to the 2011_12 Scorecard data file. The three-year CDR replaced the two-year CDR as the Department's accountability metric and the two-year CDR is available in the 2012_13 Scorecard data file and prior data files.

Borrower-Based Repayment Rate on Federal Loans[34]

Float

This element depicts the fraction of borrowers at an institution who entered repayment at any point during a two-award-year period that belong to one of eight repayment status categories, measured after the specified time from entry had elapsed (e.g., 2 years after entering repayment). The rates (BBRR[YR]_[LOAN]_[GROUP]_[STATUS]) are available for 1, 2, 3, and 4 years[35] ([YR]=1, 2, 3, or 4) after entering repayment. These data aim to allow the user to examine the full picture of what happens to borrowers as they attempt to satisfy their loan obligations. The repayment statuses are assigned based on a hierarchical set of numbered categories. If a borrower meets the criteria for a specific category, the borrower is classified in that category and not evaluated for any other categories later in the numbered sequence. The repayment rate categories and their numbered sequence is as follows:

1. *Default*: failure to pay as outlined in the promissory note for more than 360 days ([STATUS]=DFLT). If at least one of the borrower's loans is in default, the borrower is classified in this category.
2. *Delinquent*: failure to pay as outlined in the promissory note for between 31 and 360 days ([STATUS]=DLNQ). If at least one of the borrower's loans is delinquent (and none of the loans are in default), the borrower is classified in this category.
3. *Forbearance*: a period of time when monthly loan payments are temporarily stopped or reduced, with interest continuing to accrue ([STATUS]=FBR). If at least one of the borrower's loans is in forbearance (and none of the loans are in default or delinquent), the borrower is classified in this category.
4. *Deferment*: a temporary postponement of payment of a loan allowed under certain conditions and during which

[34] Perkins loans are not included in the repayment rate metrics.

[35] Status evaluated at [YR]*365 days after entering repayment; e.g., 2-year rates are evaluated at 730 days after entry into repayment.

interest generally does not accrue on subsidized loans ([STATUS]=DFR). If at least one of the borrower's loans is in deferment (and none of the loans are in any of the previous categories) the borrower is classified in this category.

5. *Not making progress*: making regular payments but the sum of all outstanding loan balances exceeds the sum of the original loan balances and none of the prior categories apply. ([STATUS]=NOPROG)
6. *Making progress*: making regular payments and the sum of all outstanding loan balances is less than the sum of the original loan balances and none of the prior categories apply. ([STATUS]=MAKEPROG)
7. *Paid in full*: all of the loans considered are paid in full. ([STATUS]=PAIDINFULL)
8. *Discharged*: the obligation to repay has been removed, typically due to death, disability, bankruptcy, fraud, or identity theft. ([STATUS]=DISCHARGE)

Repayment rates are based on the set of federal loan borrowers who enter repayment in a given two-award-year period, so the 2016-2017 repayment cohort is based on students entering repayment from July 1, 2015 – June 30, 2017. In terms of measurement, repayment status is assigned at the borrower level and is measured after the exact measurement period indicated by the metric – e.g., a 1-year rate records the status of each borrower 365 days after that borrower entered repayment.

Repayment variables include several loan types and demographic disaggregations. Data related to repayment are reported for undergraduate student loans (Direct Loans, [LOAN]=FED) for individuals who were undergraduate students, overall and disaggregated by undergraduate completion status ([GROUP]=UG, UGCOMP, UGNOCOMP, UGUNK); graduate student loans (Direct Loans and Graduate PLUS Loans. [LOAN]=FED) for individuals who were graduate students, overall and disaggregated by graduate completion status ([GROUP]=GR, GRCOMP, GRNOCOMP); for all student loans by dependent status, Pell Grant recipient status, and sex ([LOAN]=FED, [GROUP]=IND, DEP, PELL, NOPELL, MALE, NOMALE); and for Parent PLUS loans ([LOAN]=PP) associated with individuals who were undergraduate students, overall, disaggregated by undergraduate completion status, and disaggregated by dependent status, Pell Grant recipient status,

and sex ([GROUP]=UG, UGCOMP, UGNOCOMP, UGUNK, IND, DEP, PELL, NOPELL, MALE, NOMALE)).[36]

Since students who graduate may not immediately enter repayment due to either their 6-month grace period, or being granted deferment because of hardship or upon entering graduate school, students may enter repayment in a different year than when they separate from the institution (and are captured in the institution-level median cumulative debt metrics).

Data included in the 2019_20 data file were subject to a new privacy protocol that combined suppression with rounding and reported some of the rounded values in ranges. This new protocol allowed the Department to decrease some of the suppression and report more data in comparison to prior years.

Data included in the 2019-20 and 2020-21 data files, which include cohorts of borrowers evaluated in award years 2018-19, 2019-20, and 2020-21 should be interpreted with the understanding the two-year evaluation period partially overlaps with the coronavirus pandemic and measurements for some borrowers may have been affected by administrative forbearances and other pandemic -related changes in borrower status.

The two-year repayment rate data included on the consumer tool is additionally suppressed for institutions with fewer than 30 borrowers in the cohort (BBRR2_FED_UG_[STATUS]_SUPP and BBRR2_FED_UGCOMP_[STATUS]_SUPP).

Dollar-Based Repayment Rate on Federal Loans[34]

Float

This element depicts the outstanding balance (principal plus interest), amount originally disbursed, ratio of outstanding balance to amount disbursed, and number of borrowers at each institution who entered repayment at any point during a two-award-year period. The metrics

[36] Cohort placement for PLUS loan calculations is based on the date entered repayment for associated Stafford loans of the same borrower. If there was no corresponding Stafford loan, cohort placement is based on the date after all continuous in-school deferments have ended for the borrower.

(DBRR[YR]_[LOAN]_[GROUP]_[METRIC]) are available for 1, 4, 5, 10, and 20 years ([YR]=1, 4, 5, 10, 20) after entering repayment.

These metrics are based on the set of federal loan borrowers who enter repayment in a given two-award-year period, so the 2017-2018 repayment cohort is based on students entering repayment from July 1, 2016 – June 30, 2018. In terms of measurement, the repayment entry point is assigned at the borrower level and measurement takes place the specified number of years following entry into repayment. For example, a rate of 1.13 at the 1-year measurement point indicates that the outstanding balance 365 days after each borrower entered repayment is 113% of the amount originally disbursed.

Repayment variables include several loan types and demographic disaggregations. Data related to repayment are reported for undergraduate student loans (Direct Loans, [LOAN]=FED) for individuals who were undergraduate students, overall and disaggregated by undergraduate completion status ([GROUP]=UG, UGCOMP, UGNOCOMP, UGUNK); graduate student loans (Direct Loans and Graduate PLUS Loans. [LOAN]=FED) for individuals who were graduate students, overall and disaggregated by graduate completion status ([GROUP]=GR, GRCOMP, GRNOCOMP); and for Parent PLUS loans ([LOAN]=PP) associated with individuals who were undergraduate students, overall and disaggregated by undergraduate completion status ([GROUP]=UG, UGCOMP, UGNOCOMP, UGUNK). [37]

Since students who graduate may not immediately enter repayment due to either their 6-month grace period or being granted deferment because of hardship or upon entering graduate school, students may enter repayment in a different year than when they separate from the institution (and are captured in the institution-level median cumulative debt metrics).

[37] Cohort placement for PLUS loan calculations is based on the date entered repayment for associated Stafford loans of the same borrower. If there was no corresponding Stafford loan, cohort placement is based on the date after all continuous in-school deferments have ended for the borrower.

Appendix A: Highlights and Excerpts from the 2015 Technical Paper

Each year, tens of millions of Americans make choices about higher education. To address the limited information about college quality and costs, the new College Scorecard was created to provide reliable and unbiased information about college performance. The College Scorecard data provides students and families with tools to make informed decisions about postsecondary education based upon an institution's prior performance, institution and student outcomes, and financial costs. The College Scorecard data empowers consumers and policymakers involved in the college decision making process, bringing transparency to the outcomes and costs faced by students in colleges across the country. However, to make well-informed decisions it is essential for users to understand the strengths and limitations of the outcome and cost metrics available.

This appendix provides highlights and excerpts adapted from the technical report "Using Federal Data to Measure and Improve the Performance of U.S. Institutions of Higher Education" released in September 2015, offering users of this data documentation report more detail about the characteristics of the data and factors data users might find useful prior to using the data. In adapting this appendix from the technical report, minor changes have been made primarily to account for the removal of specific analyses of data that were interspersed with the general information relevant to this data documentation report.

Data Sources and Performance Metrics

Integrated Postsecondary Education Data System (IPEDS)

Collected annually through surveys administered by the Department of Education's National Center for Education Statistics (NCES), IPEDS is the primary source of data on postsecondary education institutions in the United States. The data include important performance information such as graduation rates for student subgroups; retention rates; tuition, cost of attendance, and net prices; and enrollment of low-income students (i.e., the number of Pell Grant recipients). The data also include a host of institutional information including location, type of control, level of degrees offered, enrollment size, institutional resources (including student-per-faculty ratios and faculty salaries), institutional selectivity, and characteristics of enrolled students, such as the 25th and 75th percentiles of admissions test scores for students who submitted their results.

Under the Higher Education Act, all institutions that participate in Title IV federal student aid programs must complete the IPEDS questionnaires. As such, IPEDS provides a broad starting point to define a universe of institutions on which to focus attention.

IPEDS data have several important limitations for measuring institutional performance. Perhaps the most significant is that many outcomes are recorded for a limited subset of students. Most importantly, graduation rates are only reported for cohorts of full-time, first-time students, so graduation rate information is not available for students who may have previous higher education experience or for

part-time students.[38] Another limitation is that outcomes are not recorded for students who transfer from the institution. Thus, information on graduation rate outcomes is limited.

Below are some measures of interest related to institutional performance using IPEDS data:

1. Cohort graduation rates for full-time, first-time students. This is the official measure of graduation rates mandated by the Higher Education Act, measuring the fraction of full-time, first-time students who complete their program of study within 100, 150, or 200 percent of the 'normal' completion time—e.g., the 150 percent completion rate measures the fraction of the cohort that graduates within six years for students pursuing a four-year degree or three years for students pursuing a two-year degree.

2. Average net price and cost of attendance information for full-time, first-time students who receive federal financial aid. IPEDS requires institutions to report information about tuition, fees, and other living costs that together constitute an estimate of the total cost of attending the institution. Institutions also report the average grant or scholarship aid awarded to any student receiving such aid from federal, state or local governments, or from the institution. Average net price can then be computed for both this group overall and for Title IV students only. More accurate affordability information is available by measuring the average net price for students across five income categories for Title IV recipients.

3. The fraction of enrolled students who receive a Pell grant. This is one of the most commonly used measures of the degree to which institutions provide access to low-income students. A limitation of this measure in reflecting the extent to which the institution serves low-income students is that many low-income students never apply for aid or may not be eligible (e.g., undocumented and international students), and so are not captured in the measure

National Student Loan Data System (NSLDS)

The National Student Loan Data System (NSLDS) is the Department of Education's central database for monitoring federal student aid—primarily federal student loans and Pell grants. The NSLDS contains data exclusively on federal borrowers and grant recipients dating back to the 1960s. It is used primarily for operational purposes, such as tracking federal grant and loan disbursements, the enrollment status of aid recipients to determine repayment status, payments on federal loans, borrower status (e.g., deferment, forbearance, or default), and remaining loan balances.

These data can be used to produce a variety of new institutional performance metrics that are described and presented below:

[38] There is information on total numbers of degrees awarded to students of various types, but this cannot be used to compute outcomes for enrollment cohorts. Note that since publication of the technical paper, IPEDS began collecting the Outcome Measures component, which expands the collection of cohort completion rates to cohorts beyond full-time, first-time students.

1. The median cumulative loan debt originated at the institution for all student borrowers of federal loans who leave the institution (i.e., either graduate or withdraw) in a fiscal year, measured at the point of separation. This gives a measure of how much a typical borrower borrows to attend the institution and thus can provide students with a sense for how much to save in preparation to attend and how much they will need to earn to cover their debt service payments after they leave college. Note that this measure does not reflect borrowing for the typical student (including non-borrowers), since the fraction of students who borrow varies greatly across institutions. Thus, information from IPEDS on the fraction of undergraduate students who borrow thus provides important context for inferring the level of borrowing for the typical student at the institution overall. Importantly, this measure is also available for the set of borrowers who complete their degree. This allows a more accurate characterization of the total borrowing a student should expect if they complete their studies at an institution.

2. Institutional cohort repayment rate, or the fraction of student borrowers who are making at least some progress paying down their loans (i.e., their balance is declining) and are not in default. This measure is conceptually like a cohort default rate but is intended to be less susceptible to artificial manipulation, which may occur if an institution pushes students into deferment or forbearance until the measurement window expires. The repayment rate is measured at one, three, five, and seven years after entering repayment.

3. Cohort completion and transfer rates for all federally-aided students who begin their education at an institution in an award year. Since aid disbursements are associated with institutions, NSLDS data can be used to identify the set of students who are first aided in an award year at each institution. While some students may first receive federal aid after their first year of enrollment, student-reported data on grade level from the student's Free Application for Federal Student Aid (FAFSA) can be used to place students in a cohort based on the year in which they first enrolled in a college. Since institutions report when students graduate or leave school, cohort completion rates can be estimated for each institution. Moreover, transfer to and completion at other institutions can be measured for the same cohorts by tracking whether students enroll and/or graduate at other institutions. Since NSLDS has not traditionally been used to measure these outcomes, there are several important caveats to these measures. Currently these data are intended to be used by researchers and institutions to gauge and benchmark their performance relative to other institutions and to help generate better understanding of the validity of these metrics. Until their validity are better understood, the data should not be used for consumer information applications, as they have the potential to mislead.

4. Student demographic information taken from FAFSA forms. While IPEDS provides information about prior achievement, racial and ethnic composition, and other aspects of enrolled students, data from the FAFSA complement that information in several ways. First, since the data cover all federal aid recipients, the FAFSA characterizes the same students whose outcomes are measured in NSLDS or in the earnings information from the Department of Treasury. Additional information on students' family income and parental education can be gleaned from the FAFSA to help provide

context for differences in outcomes for students across institutions. These FAFSA data allow us to report both NSLDS and earnings information for some subgroups based on gender, family income, financial dependency status, and parents' educational levels.

Administrative Earnings Data from Tax Records

To gain insight into the labor market outcomes of individuals attending institutions of higher education, data on federally aided students have been linked to earnings data from administrative tax records maintained by the Department of the Treasury. These linked data are used to produce aggregated and de-identified estimates of institution-level statistics, such as the mean and median of the earnings distribution of federally-aided students in a year. Earnings are defined as the sum of wages and deferred compensation[39] from all W-2 forms received for each individual, plus self-employment earnings from Schedule SE. Importantly, because W-2 forms are filed by employers, the estimates of labor-market outcomes cover the population of employees in the Social Security system and the self-employed individuals who file tax returns. Moreover, the earnings information offers national coverage, in contrast to some commonly used information from statewide programs. For example, state unemployment insurance administrative data is limited to the subset of students who work in the same state after leaving college.

Administrative earnings records are used to estimate inter alia the following measures of student labor market success for cohorts of federally-aided students beginning their studies in various years:

1. Mean and median earnings among workers (i.e., among those with positive yearly earnings) 6 to 10 years after first enrolling in an institution. These simple measures provide an overall assessment of the degree to which past attendees of an institution can find jobs that pay well. Individuals who are identified as currently enrolled (determined by having a federal loan that is in in-school deferment) are excluded from the calculation.

2. The fraction of former students earning over $25,000 (in constant dollars). This alternate measure of labor market success is designed to measure the extent to which former students find at least a minimal level of employment. Similar to the estimates of earnings percentiles and average earnings, this measure is based on all non-enrolled individuals with positive annual earnings. The $25,000 threshold was chosen since it approximately corresponds to the median wage of workers age 25 to 34 with a high-school degree only.[40]

[39] More precisely, the sum of earnings reported in Box 1 and Box 12 on the W-2 form. This includes both cash and noncash payments. Income related to workers' compensation, small employee achievement awards, adoption assistance, low levels of educational assistance, and small transportation benefits are not included in the earnings calculation (IRS, 2014).

[40] This figure was calculated by using the median earnings from the personal income tables from the 2014 CPS ASEC for high school graduates ages 25-34 for all races. These earnings ($24,835) were reported in 2013 dollars, and were converted to 2014 dollars using the CPI-U (multiplying by 236.712/232.962), giving a final median of $25,234.

3. Percentiles of the earnings distribution for workers. To give a broader sense for the range of earnings outcomes of prior students, for larger institutions the data contain information on the 10th, 25th, 75th, and 90th percentiles of the earnings distribution.

Important Properties and Limitations of Federal Data

As described above, the data contained in NSLDS, together with earnings information from de-identified tax records, have the potential to significantly expand our understanding of the performance of higher education institutions. There are, however, important aspects of how the data are derived that users of the data should be aware of to prevent inaccurate or misleading conclusions about institutions and the higher education community at large.

Students Covered

Most of the metrics of institutional performance described in this appendix—those based on data in the NSLDS, or matched earnings data for students appearing therein—are based on undergraduate students receiving federal aid. Moreover, institution-specific measures of debt, default, and repayment are based on the subset of students with federal loans.

The lack of data on students who do not receive Title IV aid may result in somewhat biased estimates of collective student outcomes at institutions with low proportions of Title IV aided students.[41] The data thus may not serve as a comprehensive indicator for how well institutions serve all the students they enroll. However, relative to other publicly available data sources that have their own limitations, the data are still likely to be a significant improvement. Again, the question of bias depends on the statistic desired—the average outcome of federally-aided or all students—and the differences between those groups within each institution.

Institutions Covered and the Level of Aggregation of Information

The IPEDS definition of institution is used as the basic level for reporting data, though for some metrics a higher level of aggregation (i.e., six-digit OPEID) is used due to limitations. Postsecondary institutions are complex organizations, often comprising many separate campuses, including online programs, and sometimes operating under different names entirely. Each organization chooses which of its several branches it would like to report information for separately, as opposed to jointly with the 'main campus,' generally aligning with how the institution represents its collection of sub-entities to the public.

These institutions vary dramatically in size across sectors. Some complications in comparing institutions arise when considering branch locations. The IPEDS data and NSLDS data have different reporting

[41] Certain measures like repayment or debt to earnings ratios require students to have Title IV aid, so restricting to the Title IV population does not bias estimates for these measures.

requirements for branch locations, and institutions also have latitude in how they aggregate campuses and programs in reporting.[42]

While some large institutions assign only one six-digit OPEID and differentiate at the eight-digit level (e.g., Pennsylvania State University), it is worthwhile to note inconsistent organization at the six-digit OPEID level, as some large public and for-profit systems assign different six-digit OPEIDs for each campus. For example, the California State system has twelve different six-digit OPEIDs assigned to it, and the University of Wisconsin system has thirteen. Some for-profit institutions also have many six-digit OPEIDs.

Cohort Definitions Are Imperfect and Vary for Different Metrics

The data contain diverse measures of institutional performance constructed both with an eye towards the type of information that would be most useful to prospective students, as well as towards how the measures might promote accountability for institutions. The measures require different definitions of cohorts. Users of the data should be aware of this, particularly when constructing analyses of the relationship between different measures. Moreover, reporting inaccuracies in some data elements used for cohort definitions are also important.

Data for both earnings and NSLDS-constructed completion rates are based on cohorts of students who are estimated to have begun their studies in the same "entry year"—measured in terms of "award years," which run from July 1st to June 30th (e.g., award year 2002 is from July 1, 2001 to June 30, 2002). While this construction is similar to the completion cohorts reported in IPEDS, limitations in the data create measurement error, the extent of which is difficult to assess. The NSLDS data do not directly record the date when students first enter an institution. That date is estimated based on a combination of when the student is first observed receiving federal aid at an institution, and the student's self-reported grade level on the FAFSA form associated with that record. If students report they are entering their second undergraduate year on the FAFSA associated with their first receipt of aid at a university,

[42] The Office of Federal Student Aid (FSA) recognizes the separate branch 'locations' for all Title IV-eligible institutions that have a Program Participation Agreement (PPA)—the formal document establishing their eligibility to participate in Title IV programs—with the Office of Postsecondary Education (OPE), and assigns each location an 8-digit OPE identification number (OPEID). Related institutions all share a common 6-digit root tied to the "main campus" that is the entity that enters into the PPA with FSA. Institutions may choose to list additional locations as a separate entity (or branch campus) in IPEDS so long as the institution is permanent, is physically removed from the parent (main campus) institution, and offers organized programs of study (e.g., rather than a location for offering courses only). Aside from closures, mergers, and new schools, institutions of higher education may revise their reporting structures in IPEDS. For instance, the University of Phoenix requested to change its reporting beginning in 2014-15 from campus-level to state-level. Therefore, over time, UNITIDs for the institution are not always consistent. While NSLDS is increasingly moving towards reporting student enrollments and aid-receipt at the 8-digit OPEID level, this is not yet universal and many institutions report information on where students enroll or receive aid only at the 6-digit OPEID level (since the main campus is often where aid is managed for all branches of an institution). At present, therefore, data derived from the NSLDS and earnings measures that take the universe of Title IV students at each institution as a base can be reliably calculated only at the 6-digit OPEID level, a limitation shared by the familiar cohort default rate measure. Thus, IPEDS institutions sharing a common 6-digit OPEID are all assigned the (student-weighted) average outcome or median outcome for students across all branches of the institution for NSLDS or tax-data derived measures.

they are assigned to an entry year one year prior to when we observe them first receiving aid. For students reporting that they are entering their third through fifth undergraduate year, they are assigned an entry year two years prior. The adjustment is capped at two years, since it appears that a non-trivial portion of respondents who report entering their fourth or fifth undergraduate years are misreporting their secondary school grade level on the FAFSA form.[43] In sum, the NSLDS completion and earnings cohort measures are based on the universe of Title IV students in a given entry cohort at each institution. IPEDS completion measures are based on the set of full-time first-time students in a given entry cohort at each institution. While in principle the NSLDS cohorts contain only undergraduate students, in practice it appears that inaccurate reporting may lead to graduate students being included in the cohorts in some cases. Students are included in the cohort if they receive either a Pell Grant or an undergraduate federal loan, where an "undergraduate loan" is identified as those where an institution does not report the academic level of student as a graduate student on the loan record. There seems to be some evidence of misreporting by academic level, for example, as we find undergraduate students enrolled in some institutions known to enroll only graduate students. Another way in which this might occur is if students receive an undergraduate loan at an institution, but then begin a graduate program at the same institution with no changes recorded for their academic level in NSLDS. Both types of error appear limited in scale but might affect results for some institutions. To a certain extent, we address this type of error by eliminating institutions that report no degree-seeking undergraduates or undergraduate awards in IPEDS.

Student debt measures are based on cohorts of federal loan borrowers who exit from institutions in a given year. That is, cumulative student debt is measured both for students who withdraw and those who graduate from an institution in a given year, again based on award years. The data are provided as a single metric and then disaggregated by completion status.

Finally, repayment rates are based on the set of federal loan borrowers who enter repayment in a given repayment year, corresponding to the federal fiscal year. For example, the 2011 repayment cohort is based on students entering repayment from October 1, 2010 to September 30, 2011. Since students who graduate may not immediately enter repayment either due to their 6-month grace period or to a loan deferment, many students are likely to enter repayment in a different year than when they exit the institution.

Since different metrics are based on different groups of students and organized by different event dates, users of the data should exercise caution when linking data for different metrics for the same institution. For example, an analyst may be interested in how changes in completion rates affect the likelihood of loan repayment across a set of institutions. This analysis is complicated because neither NSLDS nor IPEDS measures of completion are presented only for borrowers (the relevant population for the repayment measure), and because the set of students in a given entry year cohort may enter repayment across several different repayment years. To partially facilitate linking measures for such analyses, the

[43] For example, a student who is first aided in 2003 who reports entering his second undergraduate year would be assigned to the 2002 entry cohort. An exception to this general rule is that students who are observed transferring from a previous institutions are assigned to the entry cohort when they are observed first-aided at the institution.

completion and earnings data contain the median dates at which individuals in each entry cohort exited the institution (relevant for debt) and entered repayment (relevant for repayment rates).

Using Institutional Data to Explore Sector-level and Student Outcomes

The College Scorecard data can be organized by institution, and in some cases can be used to inform student outcomes at various levels of aggregation such as for a specific institution type or sector, or student subgroup (e.g., low-income students). For many institutional outcome measures reported, including completion or repayment rates, the data include the number of students in the institution and the subgroup upon which each measure is based (i.e., the number in the denominator of the rate) so that a weighted student level average can be completed by aggregating outcomes across any relevant set of institutions.

There are, however, some limitations to subgroup calculations that are important to understand. First, as mentioned above, the data are only for Title IV aid recipients. This makes it difficult to calculate what fraction of the relevant subgroup is represented by the data because IPEDS and other data sources do not contain the total of all Title IV students or all undergraduate students in a given entry or exit cohort. Moreover, some data elements are suppressed for institutions or subgroups with few observations, and so a small fraction of students may not be represented in the data. Additionally, to reduce the influence of outlier observations and provide students with more reliable information, many measures of performance such as cumulative debt levels and earnings information are presented for the median (rather than mean) student in a cohort. Medians cannot be aggregated in the same way as means to calculate, for example, median earnings for students in a particular sector. In some cases, both means and medians are presented, so users can select the measure that best suits their purpose.

Additional Notes on the Earnings Measures

Beyond the data issues described above, users of the earnings data should be aware of several details that give context to the interpretation of particular student outcome metrics. Aggregate earnings statistics are presented for students measured at various points in time following the year they enter an institution. Data users should be aware of the following important aspects of these data:

1) Earnings are estimated for undergraduate Title IV recipients only, because we match student-level data in NSLDS to administrative earnings records. The share of each institution's entering class represented by Title IV students can vary substantially due to both differences in family income of students attending those institutions and to state and institutional aid policies. In institutions where this share is low, results may be less representative of the entire student body.

2) Mean and median earnings, as well as percentiles of the earnings distribution, are presented for non-enrolled workers (where "worker" is defined as an individual with positive earnings in the calendar year). Enrolled individuals are omitted from the calculation to avoid having earnings appear low for institutions where a high fraction of students go on to graduate studies (though some enrolled individuals are in different undergraduate institutions). These students are

identified as enrolled if they have an undergraduate loan that is in in-school deferment status for at least 30 days during the measurement year. A limitation of this methodology is that there is no way of identifying whether former Title IV recipients who received only Pell grant aid are enrolled. Although further work is required to fully understand the measurement gaps, preliminary results suggest we may be mistakenly including some enrolled students in the earnings measure. This could lead to an underestimation of mean and median earnings of non-enrolled students at institutions where students significantly reduce their work hours while enrolled.

3) Earnings are measured for all Title IV students who attended an institution, regardless of completion status. Thus, variation in earnings across institutions will partially reflect differences in completion rates (to the extent completion affects earnings at an institution). This differs from other commonly reported earnings measures that often present earnings information for graduates only, ignoring the outcomes of the often numerous students who fail to complete their degree.

4) The data include information on the fraction of individuals who did not work for pay among those who are not currently enrolled. This is based on information about the number of individuals with no reported earnings over the course of the full year. Data users should be careful in interpreting this as a measure of unemployment, meaning the fraction of workers in the labor force (actively searching for a job) who are unable to find employment. The fraction of those not working who are likely to be "unemployed" is likely to differ across institutions. For example, at institutions that specialize in vocational training offerings, it is likely that a significant share of non-employment does in fact represent unemployment, given the goals of students enrolling in such programs.

Additional Notes on NSLDS Completion and Transfer Rate Measures

The Scorecard data relies primarily on the well-known IPEDS completion measures and also include preliminary, NSLDS-derived measures of completion and transfer rates that appear promising relative to currently available measures and are likely to improve as institutional reporting becomes more accurate. Since NSLDS is not designed to measure these outcomes, however, it is not surprising that there are some limitations to using the data to measure completion and transfers. These data are currently not appropriate for consumer information purposes, pending further quality review. Researchers should therefore be careful in their use and especially cognizant of the following limitations:

1) While many institutions appear to report completion or withdrawal outcomes very accurately, many struggle to do so for their students who do not take out loans. This is primarily attributable to the fact that NSLDS is an administrative financial aid database, the main purpose of which is tracking such information to determine when students enter repayment. This decision does not depend on whether students separate from the institution through withdrawal or graduation and is irrelevant for students who do not borrow to attend school. As a result, reporting for Pell-only recipients is inconsistent, leading to low estimated completion rates in many institutions with high fractions of Pell-only recipients. Completion rates are particularly low for community colleges at which many students receive only Pell grants, and appear low for some elite institutions that have

adopted "no-loan" financial aid policies for students below some family income threshold (so Title IV recipients are predominantly Pell only recipients). More generally, some institutions appear not to reliably report completion outcomes for any students (those receiving loans and/or Pell grants) and therefore have (implausibly) low estimated completion rates.[44]

2) The data does not directly report start dates for students. Rather, start dates are estimated from the dates we observe students first receiving federal aid at an institution in conjunction with their self-reported grade level at that time. As a result, there may be errors in assigning students to cohorts. Validation analyses using state administrative data suggested fewer than five percent of students overall were wrongly assigned to a more recent cohort.

3) Finally, the data in NSLDS on enrollment intensity and transfer status are both of poor quality for Pell-only students prior to 2012. Because of this, the data do not support reporting completion rates disaggregated by full-time and part-time status, or first-time and not-first-time status. Moreover, since transfers can only be identified if the student receives Title IV aid at the transfer-in institution, NSLDS cannot reliably identify all transfer students.

Overview of the Measures Used

Five areas of college performance measures are discussed below: earnings, completion, cost, debt and repayment and access. Each content area section explores different topics related to measurement. First, the earnings section outlines the benefits of using Scorecard data rather than Unemployment Insurance (UI) data from states to measure earnings. Second, the completion section discusses the National Student Loan Data System (NSLDS) data as a source for completion outcomes. Third, a section on costs describes the components of college cost, with attention paid to the differences between various cost measures. Fourth, the debt and repayment section presents details about the NSLDS data used to track student borrowing. Finally, the access section presents a number of metrics that may be constructed with existing data to measure accessibility for disadvantaged students.

Earnings

While post-college earnings represent only one dimension of labor market success, they are an important indicator for students and a signal of institution quality. Post-enrollment earnings are one of few meaningful indicators available for nearly all institutions that are reported on a common scale. Even though students enroll in diverse programs of study, their earnings reflect the labor market's valuation of the human capital acquired in school.

The simple measures outlined in the "Administrative Earnings Data from Tax Records" section of this Appendix—mean earnings, median earnings and fraction not working—provide an overall assessment of

[44] NSLDS data were compared with data from other sources, such as data from the National Student Clearinghouse (NSC), the State Council of Higher Education for Virginia (SCHEV), and the Beginning Postsecondary Survey (BPS). These data were used to conduct national-, sector-, and student-level comparative analyses, using NSC supplemented data where appropriate.

the degree to which past attendees of an institution are able to find good jobs that pay well. These measures are calculated using administrative IRS tax records, a resource that has several benefits when compared to other earnings data sources, including state UI records.

It is important to note how well IRS data compare to administrative state unemployment insurance (UI) data on quarterly earnings. The state UI data have some advantages and disadvantages relative to the IRS data. An advantage—though not related to the underlying data—is that state UI based earnings are reported for all students rather than Title IV students only. On the other hand, state UI data are limited in that only students who work in the same state as the institution can generally be matched (though efforts to match UI records across states are underway), and certain kinds of earnings (e.g., self-employment earnings) and occupations not covered by the UI system are excluded.

Overall, there are large gaps in the number of students covered by UI data, but despite these differences, estimates of average and median earnings based on UI and national tax data are quite similar. Still, for a small fraction of institutions, quite sizeable differences exist. Further work is necessary to characterize the sources of these differences, but the fraction of students who work out-of-state and the level of earnings overall drive some of the bias in UI data-based estimates.

Completion

The newly developed NSLDS completion data hold the promise of being able to measure outcomes for a broader set of students—including part-time and transfer students—and to provide more granular outcome estimates for various student subgroups. While NSLDS data is expected to improve over time, there are several limitations and concerns over the data quality behind the NSLDS rates that recommend caution in their use, especially for 'consumer information' applications. These data are being made available to the field to facilitate dialog about how these new data, in conjunction with other existing information, can best enhance estimates of institutional progression and completion related outcomes. For students with loans, NSLDS completion measures tend to look similar to other data sources,[45] particularly in more recent years. However, reported data for students who only receive grants remains challenging to find.

Cost

Understanding the costs of higher education is critical to both consumers, choosing whether and where to attend college, and to policymakers, examining whether scarce resources are being deployed in an efficient manner. Complications arise, however, because there are a variety of different ways to measure costs and the price students pay often differs from the `sticker price' they see. Various cost measures are outlined below. It is also worthwhile for interested parties to consider the ways in which institutional spending and cost relate.

- Tuition, sometimes called the `sticker price,' is what institutions generally advertise as the price charged to attend. The tuition measure reported in the College Scorecard is the sum of tuition

[45] E.g., data from the Beginning Postsecondary Students Longitudinal Study.

and fees for full-time students, reported by institutions in IPEDS. Tuition and fees can vary by residency status, as many public academic-year institutions report costs separately for in-state and out-of-state students. Some colleges also have a separate tuition and fee schedule for 'in-district' students who reside in the same locality as the institution. Many other institutions, especially private colleges, have the same tuition and fees for all students, regardless of residency status.

- The broadest measure of costs to students is cost of attendance, also reported to IPEDS by institutions for students paying the in-state or in-district tuition rate.[46] In the cost of attendance measure, living expenses are calculated by weighting the expenses faced by students living on-campus, students living off-campus with their families, and students living off-campus without families by the number of students in each category.

  Though this measure provides students the most comprehensive view of the total cost of attending college (assuming they receive no additional aid), research by Kelchen, Hosch, and Goldrick-Rab (2014) find considerable variability in the accuracy of reported living expenses. They find that public institutions tend to underestimate actual living costs for students, while for-profits tend to provide the most inaccurate estimates, with both over and underestimation of expenses. Such inaccuracies can be costly for students deciding if and where to attend college.

- A final cost measure is the average net price to students at each institution, defined as the cost of attendance less all grant aid for federal aid recipients paying the in-state or in-district tuition rate. The net price measure captures the fact that many students do not need to pay the full cost of attendance, as they receive grant aid from the state or federal government or the institution itself.

Net prices paid by students can differ substantially from the full cost of attendance on average, but these data mask even greater differences for students from different family income backgrounds. Providing these net price data by family income categories is important in order to give students the most accurate data possible.

Debt and Repayment

As students increasingly finance their education with loans, there is growing concern that a) financing costs unduly add to the already rising cost of college, b) some individuals may struggle to repay their loans after college and the resulting damage to credit records may hinder the ability to build wealth, and c) the burden associated with repaying loans may affect student choices after they exit college in undesirable ways.

At institutions where a large number of students withdraw before completion, the total debt level is likely to be influenced by the fact that the typical student spends only a fraction of the time it takes to complete a degree at the institution. Using the median debt level for students who complete a degree avoids this distortion. It is important to note, however, that this measure (appropriately) is still

[46] There is some debate among education researchers as to whether living costs should be included in the cost of attendance, as estimates of living costs may not represent the additional cost relative to what an individual would need to spend if they were not in college (Johnson, 2009; Romano, Losinger, and Millard, 2011; College Board, 2014a).

influenced by time-to-degree differences across institutions: institutions where students take longer to complete are likely to have higher debt levels.

Another critical consideration for consumers and policymakers is the repayment rate, which shows the fraction of students at an institution who are able to pay down their loans. This measure is similar to the well-known cohort default rate metric (CDR) but is designed to be a more robust measure of the fraction of borrowers struggling to repay their loans and is meant to be less susceptible to gaming behavior by institutions.[47] In addition to default, individuals count as a 'failure' if their loan balance fails to decline between repayment entry and the measurement date. As such, forbearance or deferments alone do not improve the institutions' rate since loan interest will still generally accrue even if this prevents a loan from entering default. One flaw with treating negative amortization of the loan as a failure is that students with unsubsidized loans who might defer for 'good reasons'—e.g., to attend graduate school—may see their loan balances increase between measurement periods as interest accumulates even if their loan is always in good standing. This problem is mitigated somewhat by excluding students in in-school loan deferment at the point of measurement from the calculation. But the measure would still record as a 'failure,' for example, a student who entered repayment, saw their balance increase while in a master's program, and then entered repayment again just prior to the three-year measurement date.

Access

Improving outcomes and reducing costs is especially important for disadvantaged students. To understand and improve higher education opportunities for disadvantaged students, it is first necessary to measure enrollment of these students at institutions. In this section, we outline four ways that enrollment of disadvantaged students in higher education could be measured. First, IPEDS data can be used to determine the fraction of Pell students at an institution. This measure is simple to calculate but obscures differences in financial circumstances among students who receive Pell grants and those who do not and miscounts the fraction of low income students in places where low tuition and alternative grants reduce the need for Pell.

IPEDS data can also be used for measuring enrollment of disadvantaged students by examining the fraction of students in different family income bands, where low-income status may be reasonably defined as students with family income less than $48,000. Using the IPEDS data to construct a family income measure requires restricting analysis to full-time, first-time students who receive Title IV aid, while the percentage Pell measure relates to all students at an institution.

Alternatively, parental financial and parental education data[48] from the NSLDS can be used to construct enrollment rates of students with different demographic backgrounds. Though this data is only available for Title IV students, the demographic information in the NSLDS can be used to calculate the fraction of first-generation students.

[47] From the late 1990s through the 2000s, an increased length of the delinquency period of a loan (180 to 270 days), coupled with a measurement period of 2 years, better allowed for deferment and forbearance periods to mask student default, leading to significantly lower default rates (Lederman, 2007).
[48] Derived from the student's FAFSA.

A final enrollment measure could be defined by the poverty rate in the students' zip code using U.S. Census data merged to individual student records in the NSLDS. As with other NSLDS calculations, this enrollment rate is restricted to Title IV students which may decrease the comprehensiveness of the measure.

Colleges face a unique challenge in serving disadvantaged students. The measures described above can be a useful starting point to examine how well colleges serve these students.

References

College Board. 2014a. "Trends in College Pricing: 2014." *Trends in Higher Education.*

Internal Revenue Service (IRS). 2014. "Wages, Salaries, and Other Earnings: Special Rules for Certain Employees." Chapter in Publication 17: Your Federal Income Tax. Retrieved from http://www.irs.gov/publications/p17/index.html.

Johnson, Nate. 2009. "What Does a College Degree Cost? Comparing approaches to measuring 'cost per degree.'" Delta Cost Project: White Paper Series.

Kelchen, Robert, Braden J. Hosch, and Sara Goldrick-Rab. 2014. "The Cost of College Attendance: Trends, Variation, and Accuracy in Institutional Living Cost Allowances." Wisconsin Hope Lab.

Lederman, Doug. 2007. "A More Meaningful Default Rate." *Inside Higher Ed.*

Romano, Richard M., Regina Losinger, and Tim Millard. 2011. "Measuring the Cost of a College Degree: A Case Study of a SUNY Community College." *Community College Review* 39(3): 211-234.

Appendix B – Metrics no longer updated

College Scorecard includes in its downloadable data files and API, previously published data elements even if those data elements are not updated with more recent cohorts of students or more recent measurement periods on an annual basis. This appendix provides information on data elements that were previously published (and still available), but no longer updated on an annual basis.

Student Body

Several elements identify demographic and other details about the student body of the institution. Some of the elements are available through IPEDS and others were produced using NSLDS.

Undergraduate Students by Family Income — Float

Last updated in the fall of 2018 (released in the 2016-17 data file).

Using data from NSLDS, these elements calculate the percentage of Title IV-receiving students who had family income data from the Free Application for Federal Student Aid (FAFSA) for each of five income quintiles[49] (INC_PCT_LO = $0-$30,000; INC_PCT_M1 = $30,001-$48,000; INC_PCT_M2 = $48,001-$75,000; INC_PCT_H1 = $75,001-$110,000; and INC_PCT_H2 = $110,001+) in an entry cohort. These data are also produced separately for dependent (DEP_INC_PCT_**) and independent (IND_INC_PCT_**) students. Data are produced for rolling two-year pooled cohorts.[50] For these variables, years refer to award years (e.g., award year 2016-17 begins on July 1, 2016, and ends June 30, 2017).

Number of Institutions to Which Students Sent FAFSAs — Float

Last updated in the fall of 2018 (released in the 2016-17 data file).

When students complete their FAFSA applications for federal student aid, they must list at least one institution (and may elect to list several) to receive their information. This element depicts the percentage of students who send their FAFSA report to at least 2, 3, 4, or 5 institutions

[49] NSLDS calculations of the percentage of students in each income groups were based on nominal dollar values (not adjusted for inflation). This methodology differs from the way the Treasury Department identified income groups (with inflation adjustments) for disaggregated earnings calculation values in calendar years 2014 and prior (i.e. provided in the merged_2013-14 data file and prior data files).

[50] More information on how these entry cohorts were constructed is available in the "NSLDS Completion and Transfer Rates" section of this document

(APPL_SCH_PCT_GE* for 2, 3, 4, or 5). Data are produced for rolling two-year pooled entry cohorts.[51]

Undergraduate Student Demographics for Earnings Cohorts

Float and Integer

Last updated in the fall of 2018 (released in the 2016-17 data file).

NSLDS derived data elements describing entry cohorts include: the share of married students (MARRIED); the share of dependent students (DEPENDENT); the share of veteran students (VETERAN); and the share of students at the institution who report that they are first-generation on the FAFSA (PAR_ED_PCT_1STGEN[52]). Data are also available separately for the share of students who report that the highest level of education their parents completed is middle school (PAR_ED_PCT_MS), high school (PAR_ED_PCT_HS), or postsecondary education (PAR_ED_PCT_PS). Data elements also include the average and median family income of students (FAMINC and MD_FAMINC) and separately of independent students (FAMINC_IND).[53] Data are produced for rolling two-year pooled entry cohorts.[54] For these variables, years refer to award years (e.g., award year 2016-17 begins on July 1, 2016, and ends June 30, 2017).

In addition, the Treasury Department calculated the percentage of the population from students' home ZIP codes[55] by race (PCT_* for WHITE, BLACK, ASIAN, HISPANIC), education level (PCT_* for BA [bachelor's degree over the age of 25] and GRAD_PROF [professional degree over the age of 25]), nationality (PCT_BORN_US), poverty (POVERTY_RATE), household income (MEDIAN_HH_INC), and employment status (UNEMP_RATE) using Census data. These Treasury calculated data elements are available for the 2004-05 and 2005-06 pooled cohort (see 2005_06 file); the 2002-03 and 2003-04 pooled cohort (see 2003_04

[51] More information on how these entry cohorts were constructed is available in the "NSLDS Completion and Transfer Rates" section of this document

[52] The variable FIRST_GEN contains the same information and PAR_ED_PCT_1STGEN. In the student's FAFSA, if at least one parent's status was "college" then this was coded as non-first generation. First generation status used imputation methodology to assign a first-generation or non-first-generation status to students where both parents had unknown statuses.

[53] Calculations were based on nominal dollar values (not adjusted for inflation).

[54] More information on how these entry cohorts were constructed is available in the "NSLDS Completion and Transfer Rates" section of this document

[55] ZIP Codes based on students' addresses described on FAFSA forms when students apply for aid, rather than addresses collected on tax forms when earnings were measured

file); the 2000-01 and 2001-02 pooled cohort (see 2001_02 file); the 1998-99 and 1999-00 pooled cohorts (see 1999_00 file); and the 1995-96 and 1996-97 pooled cohort (see 1997_98 file). For these variables, years refer to award years (e.g., award year 2005-06 begins on July 1, 2005, and ends June 30, 2006).

Earnings

There are two notable limitations that researchers should keep in mind for all of these metrics. Research suggests that the variation across programs within an institution may be even greater than aggregate earnings across institutions; for instance, STEM and health majors frequently earn more than students who study in other fields. Second, the data include only Title IV-receiving students, so figures may not be representative of institutions with a low proportion of Title IV-eligible students. Additionally, the data are restricted to students who are not enrolled (enrolled means having an in-school deferment status for at least 30 days of the measurement year), so students who are currently enrolled in graduate school at the time of measurement are excluded.

For earnings variables that are disaggregated by family income tercile (low-income: $30,000 or less; middle-income: $30,001-$75,000; and high-income: $75,001+), family income was adjusted for inflation prior to grouping by tercile for calculations measured in calendar years 2014 and prior (i.e. values in the merged_2013-14 data file and prior data files).

Dev-category	**earnings**
Mean and Median Earnings	Integer
	Last updated in the fall of 2018 (released in the 2014-15 data file).
	Mean (MN_EARN_WNE_P*) earnings are for the institutional aggregate of all federally aided students who enroll in an institution each year and who are employed but not enrolled. Earnings are defined as the sum of wages and deferred compensation from all non-duplicate W-2 forms received for each individual (from both full- and part-time employment), plus positive self-employment earnings from Schedule SE. Data are available for each year starting six years after a student enrolls in college, up to 10 years after the student enrolls; enrollment dates are estimated based on FAFSA self-reporting, as with the completion rate cohort construction described above.

Mean earnings are also available disaggregated by FAFSA family income[56] (MN_EARN_WNE_INC1_P*, MN_EARN_WNE_INC2_P*, and MN_EARN_WNE_INC3_P*); by dependent status (for dependents, MN_EARN_WNE_INDEP0 _P*; and for independents, MN_EARN_WNE_INDEP1_P*); and by gender (for female, MN_EARN_WNE_MALE0_P*; and for male, MN_EARN_WNE_MALE1_P*).

Earnings included in the 2011_12 and prior Scorecard data files are inflation adjusted to 2014 dollars using the Consumer Price Index for all Urban Consumers (CPI-U). Beginning with the 2012_13 Scorecard data file, earnings included in the XXXX_YY data file are inflation adjusted to XXXX+3 dollars using the CPI-U. For example, earnings included in the 2014_15 Scorecard data file are inflation adjusted to 2017 dollars.

Earnings are based on measurement periods with years after cohort entry labeled in the variable names with a *pZ at the end. While earnings are measured based on calendar years, cohort years for earnings variables are based on award years.

Threshold Earnings

Float

Last updated in the fall of 2018 (released in the 2014-15 data file).

This metric provides the fraction of former students (including non-completers as well as completers) who earned more than the median wage of workers ages 25 to 34 that self-identify as a high school graduate (by indicating high school completion was their highest level of education), as measured 6, 8, and 10 years after entering the evaluated institution. The denominator of this fraction includes only former students who were working and not enrolled in school during in the measurement year.

This comparison group of self-identified high school graduates should be viewed with caution - the group may include individuals who have completed postsecondary education and/or training including apprenticeships and/or industry certifications. In addition, observations of self-identified high

[56] Categorization was based on inflation-adjusted values for family income.

school graduates are not perfectly aligned with the cohorts of students who entered college 6, 8, and 10 years prior. Some high school graduates may have much more work experience than students who recently were enrolled in college. For example, high school graduates ages 25 to 34 may have been in the workforce for as many as 16 years, whereas students completing a four-year degree within 150 percent of time may just be looking for entry-level work 6 years after initial enrollment or have no more than 4 years of post-completion experience 10 years after initial enrollment.

Metrics included in the 2013_14 and prior Scorecard data files use $25,000 as the threshold (GT_25K_P*) earnings for high school graduates, while the 2014_15 data file uses $28,000 as the threshold (GT_28K_P*) for high school graduate earnings.

For threshold earnings metrics included in the 2011_12 and prior Scorecard data files, earnings are adjusted to 2014 dollars using the Consumer Price Index for all Urban Consumers (CPI-U) prior to comparison with the threshold. Beginning with the 2012_13 Scorecard data file, threshold metrics included in the XXXX_YY data file are created by comparing earnings that have been inflation adjusted to XXXX+3 dollars using the CPI-U to the threshold. For example, threshold earnings metrics included in the 2014_15 Scorecard data file compare earnings that have been inflation adjusted to 2017 dollars to the threshold.

Repayment

To provide a sense for the debt burden of attending college and the loan performance metrics for each institution, we produced several elements using NSLDS. These elements can provide useful information for students and families concerned about borrowing for college and interested in seeing borrowers' behavior after they leave the institution.

Dev-category	**repayment**
Repayment Rate on Federal Student Loans[57]	Float
	Last updated in the fall of 2018 (released in the 2016-17 data file).

[57] Perkins loans and PLUS loans were not included in these repayment rate metrics.

This element depicts the fraction of borrowers at an institution who are not in default on their federal loans[58] and who are making progress in paying them down (i.e. have paid down at least $1 in the initial balance on their loans[59]) after entering repayment (RPY_*YR_RT). The rates are available for 1 (_1YR_RT), 3 (_3YR_RT), 5 (_5YR_RT), and 7 (_7YR_RT) years after entering repayment. Repayment rates are generally considered more sensitive than default rates, which measure only the worst-case scenario for repayment outcomes, and which can be manipulated through the use of allowable nonrepayment options like deferments and forbearances.

Repayment rates are based on the set of federal loan borrowers who enter repayment in a given fiscal year, so the 2013 repayment cohort is based on students entering repayment from October 1, 2012 to September 30, 2013. In terms of measurement, repayment rates are measured at the end of each respective fiscal year for which rates are calculated. For example, the three-year repayment rate for fiscal year 2013 cohort is measured at the end of fiscal year 2016 (September 30, 2016).

Repayment variables are based on only undergraduate debt acquired at the institution for which the repayment rate is reported.[60] Since students who graduate may not immediately enter repayment due to either their 6-month grace period, or being granted deferment because of hardship or upon entering graduate school, students are likely to enter repayment in a different year than when they exit (and are captured in the median cumulative debt metric). Students who have received either an in-school or military deferment during the time of

[58] This repayment rate methodology evaluated default at the time of measurement. As such, any loan that went into default but was then paid in full during the time of measurement was considered *in repayment.*

[59] *Initial balance* for this repayment rate metric refers to the entire balance upon entering repayment including loan disbursements, capitalized interest, and prior payments made prior to entering repayment. A borrower was considered *in repayment* if his or her loan balance, at the time of measurement, was at least $1 less than the initial balance.

[60] Note the same student may have been included in this repayment rate calculation for multiple institutions with the share of loans attributed to one school in that institution's calculation, and the share of the loans attributed to another institution in that other institution's calculation. In cases of consolidation, the outstanding balance of a consolidated loan was allocated among the represented schools on the basis of the relative share of each underlying loan's payoff amount.

measurement are excluded from the calculation.[61] Forbearances and other types of deferments do not impact the inclusion of a loan in the repayment rate calculation, i.e., such loans are included in both the numerator and denominator of the calculation. Loans that are discharged for death or permanent disability at the time of measurement are not included in the rate.

These data are available for all borrowers at the institution, as well as disaggregated by completion status (COMPL_RPY_* for students who completed and NONCOM_RPY_* for students who withdrew without completing); by FAFSA family income[62] (LO_INC_RPY_* = $0-$30,000; MD_INC_RPY_* = $30,001-$75,000; and HI_INC_RPY_* = $75,001+); by dependent (DEP_RPY_*) and independent (IND_RPY_*) status; by Pell status (PELL_RPY_* for students who ever received a Pell Grant and NOPELL_RPY_* for students who never received a Pell Grant); by gender (FEMALE_RPY_* and MALE_RPY_*); and by first-generation status (FIRSTGEN_RPY_* and NOTFIRSTGEN_RPY_*).[63] The repayment rates are produced in rolling two-year averages to reduce variability from year to year; and the three-year repayment rate, which is included on the consumer tool, is suppressed for institutions with fewer than 30 borrowers in the two cohorts to produce more stable measures (RPY_3YR_RT_SUPP).

[61] If a borrower was not in an excluded status on the measurement date, they were included in the calculation for this metric.

[62] Calculations were based on nominal income dollar values (not adjusted for inflation).

[63] In the student's FAFSA, if at least 1 parent's status was "college" then this was coded as non-first generation. First generation status used imputation methodology to assign a first-generation or non-first-generation status to students with unknown statuses if both parents had unknown statuses.