

Technical Documentation: College Scorecard Data by Field of Study

Contents

Introduction	1
Data Methodology	2
Defining field of study	2
Number of credentials conferred	3
Post-completion earnings	4
Cumulative federal loan debt	5
Privacy protection	6
Data Sources and Limitations	7
Federally aided students	8
First-year earnings	11
Economic conditions	12
Aggregation of data	13
Monthly estimates	14
Nonemployment	15
Data quality	15
Matching fields of study reported to different data systems	21

Introduction

In response to the Improving Free Inquiry, Transparency, and Accountability at Colleges and Universities Executive Order and Secretary DeVos's Rethink Higher Education Initiative, the United States Department of Education (the Department) expanded and redesigned the College Scorecard (Scorecard) to include new data elements describing post-graduate earnings and cumulative loan debt of graduates by field of study. The primary purpose of the Scorecard is to provide data to help prospective postsecondary students make informed enrollment decisions. Since its last major redesign in 2015, Scorecard has provided information describing post-enrollment earnings for institutions as a whole. However, measures of earnings can vary substantially across graduates of different fields of study, which makes an institutional median difficult to interpret for prospective or current students. For example, among institutions offering more than two fields of study at the bachelor's degree level, the average difference between an institution's highest-earning bachelor's degree and lowest-earning bachelor's degree was \$32,866. In proportional terms, the average institution's highest-earning bachelor's degree was estimated to be 2.5 times higher than its lowest-earning bachelor's degree. Exhibit 1 describes the differences in earnings estimates by credential level.

Exhibit 1. Number of institutions with at least two reportable earnings estimates and the average difference and ratio between institutions' highest and lowest earnings estimates by credential level

Credential level	Number of institutions with at least 2 fields of study that met reporting requirements	Average difference between highest and lowest earnings estimates	Average ratio between highest and lowest earnings estimate (highest earnings estimate divided by the lowest earnings estimate)
Undergraduate Certificate/Diploma	806	\$12,792	1.7
Associate's Degree	1,029	\$26,207	2.2
Bachelor's Degree	1,397	\$32,866	2.5
Post-Baccalaureate Certificate	11	\$19,918	1.5
Master's Degree	887	\$42,287	2.2
Doctoral Degree	137	\$39,433	1.8
First Professional Degrees	144	\$47,892	1.9
Graduate/Professional Certificate	43	\$27,551	1.8

Note: The unit of analysis for this table is the six-digit Office of Postsecondary Education ID number (OPEID).

Source: U.S. Department of Education, College Scorecard data.

Scorecard's new earnings data by field of study provide not only more granular institution-level data than previously published, but also provide more clarity on two important aspects of earnings:

- **More clarity on earnings for graduates:** Since previously published institution-level earnings combined graduates and non-graduates, it was not possible to estimate the economic value of graduating. Newly published earnings data by field of study are based on only those who graduate.

- More clarity on the earnings measurement period: Since previously published institution-level earnings were evaluated 6, 8, and 10 years after students began their studies, there was no clarity on how much time had passed after leaving school because students may take shorter or longer to finish their studies for a variety of reasons (e.g., attending school part-time, enrolling in the summer). Newly published earnings data by field of study are consistently reported for salary earned in the first year after graduation.

The purpose of this document is to describe the data-generating process and data limitations of data elements disaggregated by field of study. For detailed data documentation describing the existing institution-level data metrics, see the [technical documentation for institution-level data files](#). In addition, data users are encouraged to review the [data dictionary](#), which provides information on each metric in the application programming interface (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, the data source, and high-level notes for each metric. The data dictionary contains a cohort map that identifies the group of students each metric corresponds to within the data files.

Data Methodology

The methodology used to calculate earnings and cumulative debt by field of study was based, in part, on discussions from the spring 2019 Scorecard [Technical Review Panel](#). The [Scorecard consumer-facing website](#) includes only a subset of data elements (e.g., number of completers per year, median earnings, median cumulative debt) disaggregated by field of study for currently operating undergraduate institutions.¹ A more comprehensive set of data elements, including counts and averages, is provided in the Scorecard downloadable data files and API (see <https://collegescorecard.ed.gov/data/>). In addition, the downloadable data files and API include calculations for fields of studies in all institutions that participate in the federal student financial assistance programs administered by the Department under Title IV of the Higher Education Act of 1965, as amended.

Defining field of study

The unit of analysis for calculations by field of study is the unique combination of an institution code (six-digit OPEID and UNITID), a four-digit Classification of Instructional Programs (CIP) code, and a credential level. Institutional identifiers² include

- Integrated Postsecondary Education Data System (IPEDS) assigned ID number (**UNITID**) and corresponding institution name (**INSTNM**), as reported by IPEDS.
- Office of Postsecondary Education ID number (OPEID). For purposes of Scorecard calculations, this is based on the six-digit OPEID (**OPEID6**), representing the main campus (and all of its branches) of a multi-branch institution. Scorecard uses this aggregated approach to calculate cumulative debt and post-completion earnings data because not all institutions report information needed to calculate these metrics at the more granular branch location level. For institutions with multiple UNITIDs (branches) that map to the same six-digit OPEID (main

¹ For more information on the universe of institutions provided on the consumer facing tool, see the technical documentation for institution-level data files.

² For more information on the relationship between UNITID and OPEID6, see the technical documentation for institution-level data files.

campus), data calculated at the aggregated OPEID6-level are repeated across any branches that offer those fields of study based on IPEDS Completions component reporting.

- CIP codes are a taxonomy of academic disciplines classified by a six-digit code, with the first two digits representing the broadest category for a discipline and the first four digits representing a more granular category; all six digits represent the most granular description of an academic discipline. Scorecard uses the first four digits of the CIP code in its calculations (**CIPCODE**) because four-digit CIP code descriptions are easier for prospective students to understand and because combining six-digit CIP codes together leads to larger cell sizes, which in turn leads to fewer data points that need to be privacy-suppressed.
- The credential level (**CREDLEV**) codes correspond to the following categories of programs:
 - 1: Undergraduate Certificates or Diplomas
 - 2: Associate's Degrees
 - 3: Bachelor's Degrees
 - 4: Post-Baccalaureate Certificates
 - 5: Master's Degrees
 - 6: Doctoral Degrees
 - 7: First Professional Degrees
 - 8: Graduate / Professional Certificates

Number of credentials conferred

Scorecard provides the number of recognized postsecondary credentials conferred by field of study for two consecutive award years (IPEDSCOUNT1 and IPEDSCOUNT2) as reported by institutions to IPEDS. This count of postsecondary credentials represents all credentials including first major credentials, second major credentials, etc.; therefore, a student who completed multiple credentials would be represented multiple times in these counts. IPEDS award levels do not exactly match the credential level classification in the National Student Loan Data System (NSLDS), the source of cohorts used in the calculation of cumulative loan debt and earnings by field of study data. In order to provide IPEDS counts that correspond to the NSLDS credential levels, some aggregation of IPEDS award levels was needed. Exhibit 2 indicates the recoding and aggregation scheme applied to IPEDS to calculate the number of awards provided in the field of study data.

Exhibit 2. Correspondence of NSLDS credential level to IPEDS award level

NSLDS Credential Level	IPEDS Completions Data Award Levels Aggregated
1: Undergraduate Certificates or Diplomas	Less-than-1-year
	At least 1- but less-than-2-year
	At least 2- but less-than-4-year certificates
2: Associate's Degrees	Associate's degrees
3: Bachelor's Degrees	Bachelor's degrees
4: Post-Baccalaureate Certificates	Not available
5: Master's Degrees	Master's degrees
6: Doctoral Degrees	Doctor's degrees – research/scholarship
	Doctor's degrees – other
7: First Professional Degrees	Doctor's degrees – professional practice
8: Graduate / Professional Certificates	Postbaccalaureate certificates
	Post-master's certificates

Post-completion earnings

While the long-term goal is to provide earnings data points for students several years after completion, the Department only recently began collecting information by field of study and Scorecard is presenting only first-year earnings with the acknowledgment that earnings in the early years after graduation may not be indicative of longer-term earnings. As of 2019, Scorecard provides one-year post-completion earnings by field of study (**MD_EARN_WNE**) for the cohort of students who received federal financial aid at any time in their academic careers (regardless of where aid was received) and graduated in award years 2014-15 and 2015-16 with earnings measured in calendar years 2016 and 2017, respectively. For example, if a student received federal financial aid as an undergraduate and subsequently completed a graduate-level program at a different institution, that student would be included in the graduate-level cohort even if that student did not receive federal financial aid for graduate-level enrollment. The cohort of evaluated excludes students who:

- were subsequently enrolled in school during the measurement year,
- died prior to the end of the measurement year,
- had loan deferments for military service during the measurement year, and
- did not work during the measurement year.

In addition, Scorecard provides the count of completers (who received federal financial aid)

- excluding students due to postsecondary enrollment, military service, or death (**TITLEIVCOUNT**), and
- excluding students due to postsecondary enrollment, military service, or death, who were working in the measurement year and whose earnings are included in the calculated median value (**EARNINGS COUNT**).

Students who completed multiple awards (e.g., double major) were measured multiple times only if multiple awards were completed in different four-digit CIP codes and/or credential levels. Students who completed multiple awards within the same institution at the same four-digit CIP code and credential level were measured only once. Students who completed in multiple institutions were measured multiple times.

The earnings measurement represents the sum of wages and deferred compensation from all non-duplicate W-2 forms and positive self-employment earnings from IRS Form 1040 Schedule SE (Self-Employment Tax) for each student measured. Earnings values are presented in 2018 inflation-adjusted dollars.

Unlike previously published Scorecard institution-level earnings, the earnings calculations by field of study are based on only students who complete (calculations exclude those who withdrew or transferred). This completers-only methodology stems, in part, from the difficulty in assigning non-completers to fields of study due to students' changing majors throughout their academic careers. In addition, users have indicated a desire to evaluate outcomes based on a completers-only cohort to better understand the value of completing.

Cumulative federal loan debt

Scorecard provides the mean (**DEBTMEAN**) and median (**DEBTMEDIAN**) cumulative federal loan debt of federal borrowers by field of study for cohorts of students who complete in the same two-year timeframe (two consecutive award years). In addition, Scorecard provides the borrower count associated with each mean and median federal loan debt value (**COUNT**).

The cumulative federal loan debt includes only loan disbursement amounts and does not capture any accrued interest, even if that interest accrued prior to completion. The cumulative loan debt only includes loans disbursed at the same academic level (i.e., graduate, undergraduate) as the evaluated credential level. For example, calculated debt values for a master's degree-level field of study include all graduate-level loans disbursed by the evaluated institution and exclude all undergraduate-level loans. Academic levels are assigned to credential levels based on the following categorizations:

- Undergraduate credentials
 - Undergraduate Certificates or Diplomas
 - Associate's Degrees
 - Bachelor's Degrees
 - Post-Baccalaureate Certificates
- Graduate credentials
 - Master's Degrees
 - Doctoral Degrees
 - First Professional Degrees
 - Graduate / Professional Certificates

The cumulative federal loan debt includes all debt (at the same academic level of the evaluated credential level) dispersed by the evaluated institution regardless of whether the evaluated student switched fields of study at the evaluated institution. For example, if a student began as an undergraduate general studies student and then earned a bachelor's degree in agriculture at the same institution, the debt calculated for the agriculture bachelor's degree would include debt used to enroll in the prior general studies field of study.

All Direct Subsidized and Unsubsidized Loans (also known as Stafford Loans)³ and Graduate Direct PLUS Loans are included in the sum of cumulative federal debt. Cumulative federal debt values exclude parent PLUS loans and Perkins Loans.

Students who completed multiple awards (e.g., double major) were measured multiple times only if multiple awards were completed in different four-digit CIP codes and/or credential levels. Students who completed multiple awards within the same institution at the same four-digit CIP code and credential level were measured only once. Students who completed in multiple institutions were measured multiple times.

The estimated monthly payment (**DEBTPAYMENT10YR**) is based on the median cumulative loan debt amortized over 10 years using current interest rates. While borrowers have many payment plan options beyond a 10-year fixed payment plan, 10 years is considered a standard payment plan and provides a higher estimate for a monthly payment. For undergraduate programs, calculations are based on the current interest rate for undergraduate Direct Subsidized Loans and undergraduate Direct Unsubsidized Loans (4.53 percent). For graduate programs, calculations are based on the current interest rate for Direct PLUS Loans for graduate and professional students (7.08 percent). While the cumulative median debt calculations include both Direct Unsubsidized Loans for graduate and professional students and Direct PLUS Loans, the Direct PLUS Loan interest rate is used here because it is currently higher. This scenario provides users with a higher estimate of a monthly payment based on a standard repayment plan.

Privacy protection

Data presented by field of study have been treated with cell suppression methods to reduce the risk of disclosure. To further reduce disclosure risk, information about the specific suppression rules is not available to the public. Data points suppressed for privacy are indicated by the “PrivacySuppressed” data code.

Exhibit 3 shows that while the majority of the fields of study offered by institutions are not reportable (privacy-suppressed) for cumulative loan debt and post-completion earnings, the fields of study for which there are reportable data values represent a large share of students who complete.

³ Stafford Loans were available through both the William D. Ford Federal Direct Loan Program and the Federal Family Education Loan Program (FFELP). Since the FFELP was discontinued in 2010, Stafford Loans have been referred to as Direct Loans.

Exhibit 3. Percentage of data that are reportable following privacy suppression, by credential level

Credential level	Metric	Reportable fields of study (UNITID level)	Proportion of graduates in reportable fields of study (IPEDS)	Reportable fields of study (OPEID6 level) ¹	Proportion of graduates who received federal financial aid in reportable fields of study
All	Median debt	24.3%	68.7%	32.6%	83.4%
	Median earnings	21.4%	66.3%	28.2%	83.6%
	Both	19.4%	62.8%	25.1%	79.5%
Undergraduate Certificates	Median debt	15.0%	42.1%	24.5%	79.8%
	Median earnings	15.7%	45.5%	26.9%	87.9%
	Both	12.7%	37.6%	19.7%	77.7%
Associate's Degrees	Median debt	18.9%	65.2%	24.2%	74.4%
	Median earnings	17.5%	64.3%	22.2%	77.3%
	Both	14.8%	59.2%	17.9%	70.3%
Bachelor's Degrees	Median debt	39.2%	86.3%	44.0%	88.6%
	Median earnings	31.6%	80.1%	34.6%	84.2%
	Both	30.4%	79.2%	33.2%	83.7%
Post-Baccalaureate Certificates	Median debt	2.2%	n/a	3.5%	18.1%
	Median earnings	3.8%	n/a	6.0%	39.0%
	Both	1.4%	n/a	2.1%	14.5%
Master's Degrees	Median debt	25.1%	67.5%	30.3%	81.8%
	Median earnings	23.4%	66.0%	28.0%	82.1%
	Both	21.0%	62.6%	24.7%	78.7%
Doctoral Degrees	Median debt	7.5%	20.5%	10.0%	60.3%
	Median earnings	6.7%	20.0%	8.9%	60.1%
	Both	5.9%	17.0%	7.6%	56.3%
First Professional Degrees	Median debt	33.6%	71.9%	68.4%	97.7%
	Median earnings	32.4%	70.2%	66.3%	97.3%
	Both	30.9%	69.5%	63.2%	96.1%
Graduate/ Professional Certificates	Median debt	3.7%	22.0%	10.0%	49.8%
	Median earnings	3.6%	18.8%	8.7%	53.4%
	Both	2.8%	17.1%	6.6%	44.3%

¹ The denominator for this percentage calculation is the total number of fields of study with at least one Title IV completer in the cohort.

Note: n/a = not applicable

Source: National Student Loan Data System (as of July 2019) and the U.S. Department of Education, College Scorecard data.

Data Sources and Limitations

NSLDS is the Department's central database for monitoring Title IV federal student aid, with data on federal borrowers and grant recipients dating back to the 1960s. While primarily used for operational purposes, such as tracking federal grant disbursements and loan disbursements and repayment, NSLDS

provides administrative data from which the loan debt data element and cohorts for earnings can be constructed.

To gain insight into the labor market outcomes of individuals who graduate, Scorecard linked NSLDS records to administrative tax records maintained by the IRS within the Department of the Treasury. Specifically, post-completion earnings values are derived from the sum of wages and deferred compensation from all W-2 forms received for each individual, plus self-employment earnings from IRS Form 1040 Schedule SE.

IPEDS is a system of surveys conducted annually by the National Center for Education Statistics (NCES). The completion of all IPEDS surveys is mandatory for all institutions that participate in, or are applicants for participation in, any federal financial assistance program authorized by Title IV of the Higher Education Act of 1965, as amended. Scorecard derives the count of recognized postsecondary credentials conferred by field of study from the IPEDS completion survey results. In this survey, institutions report all credentials conferred by award level and six-digit CIP code. Scorecard aggregates these reported data up to the four-digit CIP code (by summing the counts of related six-digit CIP codes) and crosswalks IPEDS award levels to NSLDS credential levels.

Federally aided students

Since Scorecard earnings and cumulative loan debt data by field of study are derived from NSLDS data, these data elements describe only those students who received federal financial aid in the form of Title IV grants and loans. Because not all graduates received Title IV aid to attend school, data users should not assume that Scorecard data values are representative of all students who graduated from a particular field of study.

Scorecard provides cumulative debt calculations based on only federal loans for only Title IV borrowers. Further, Scorecard does not capture the amount borrowed from nonfederal sources such as loans administered by states, institutions of higher education, or other private entities. As an example of the prevalence of nonfederal borrowing, 15 percent of undergraduate fourth-year seniors between the ages of 18 and 24 had nonfederal loans in 2015-16.⁴ In addition, debt calculations are based only on the amount borrowed and do not include capitalized interest, even if that interest accrued prior to completion. Given these factors, Scorecard suggests users interpret these data with the understanding that estimates may often be lower than the amount owed upon completion.

Earnings are estimated for only those who receive Title IV federal financial aid. Exhibit 4 examines the proportion of Title IV federal financial aid recipients among undergraduates in 2015-16. This exhibit shows that, nationally, 54.5 percent of undergraduate students received federal financial aid and the proportion of aid recipients varied across different sectors of higher education. For example, among four-year institutions, 59.8, 64.4, and 77.7 percent of students received federal aid in public, private, and proprietary, institutions, respectively. Exhibit 4 also shows the number of undergraduates in each sector. The sectors with the highest number of undergraduates include public two-year institutions (7,639,500) and public four-year institutions (6,838,300).

⁴ See <https://nces.ed.gov/DataLab/TablesLibrary/TableDetails/12639?keyword=nonfederal&rst=true>

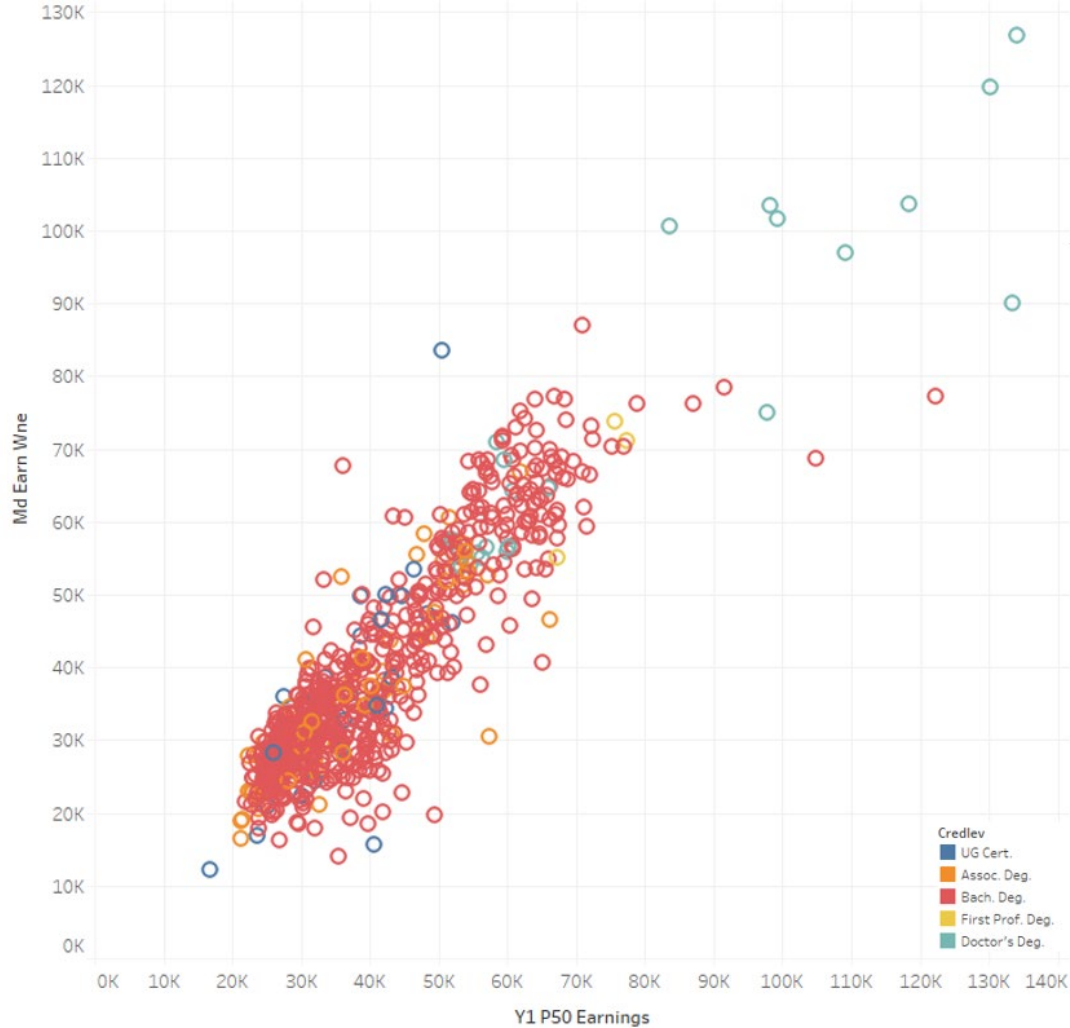
Exhibit 4. Percentage of undergraduates receiving federal aid and number of undergraduates by institutional sector 2015-16

Sector	Percent of Undergraduates receiving federal financial aid	Number of undergraduates
Total	54.5	19,532,300
Public 4-year	59.8	6,838,300
Public 2-year	40.0	7,639,500
Public less-than-2-year	45.2	66,100
Private 4-year	64.4	2,952,800
Private 2-year	75.2	94,500
Private less-than-2-year	62.3 ¹	4,000
Proprietary 4-year	77.7	1,121,700
Proprietary 2-year	76.0	449,500
Proprietary less-than-2-year	78.3	365,900

¹ Interpret data with caution. Estimate is unstable because the standard error represents more than 30 percent of the estimate. Note: Total federal aid excludes Veterans'/DOD. Number of undergraduates is based on weighted sample sizes. Computation by NCES PowerStats; see https://nces.ed.gov/datalab/index.aspx?ps_x=bkbbmaaa6. Source: U.S. Department of Education, National Center for Education Statistics, 2015-16 National Postsecondary Student Aid Study (NPSAS:16).

Scorecard earnings data by field of study were compared with publicly available data from the Census Bureau's Postsecondary Employment Outcomes (PSEO) program,⁵ which provides post-completion earnings estimates by field of study for institutions in the University of Texas system, public institutions in Colorado, the University of Michigan–Ann Arbor, and the University of Wisconsin–Madison. Since PSEO does not restrict cohorts to only students who receive federal financial aid, these data provide a comparison of whether or not there is general concordance between the calculations made with and without completers who did not receive federal financial aid. However, there are other differences in methodology that may explain differences in outcome measurements. For example, PSEO excludes individuals who were not working the majority of the year (Scorecard excludes those who did not work for the entire year). In addition, PSEO relies on unemployment insurance records (Scorecard uses tax records) and does not include self-employment or earnings from some public sector entities. Given the differences in methodology, differences in earnings estimates may represent an upper bound to the impact of including or excluding completers who do not receive federal financial aid. When examining the scatter plot in Exhibit 5, the positive correlation between the Scorecard and PSEO data suggests that the two estimates from the two different sources are generally in concordance.

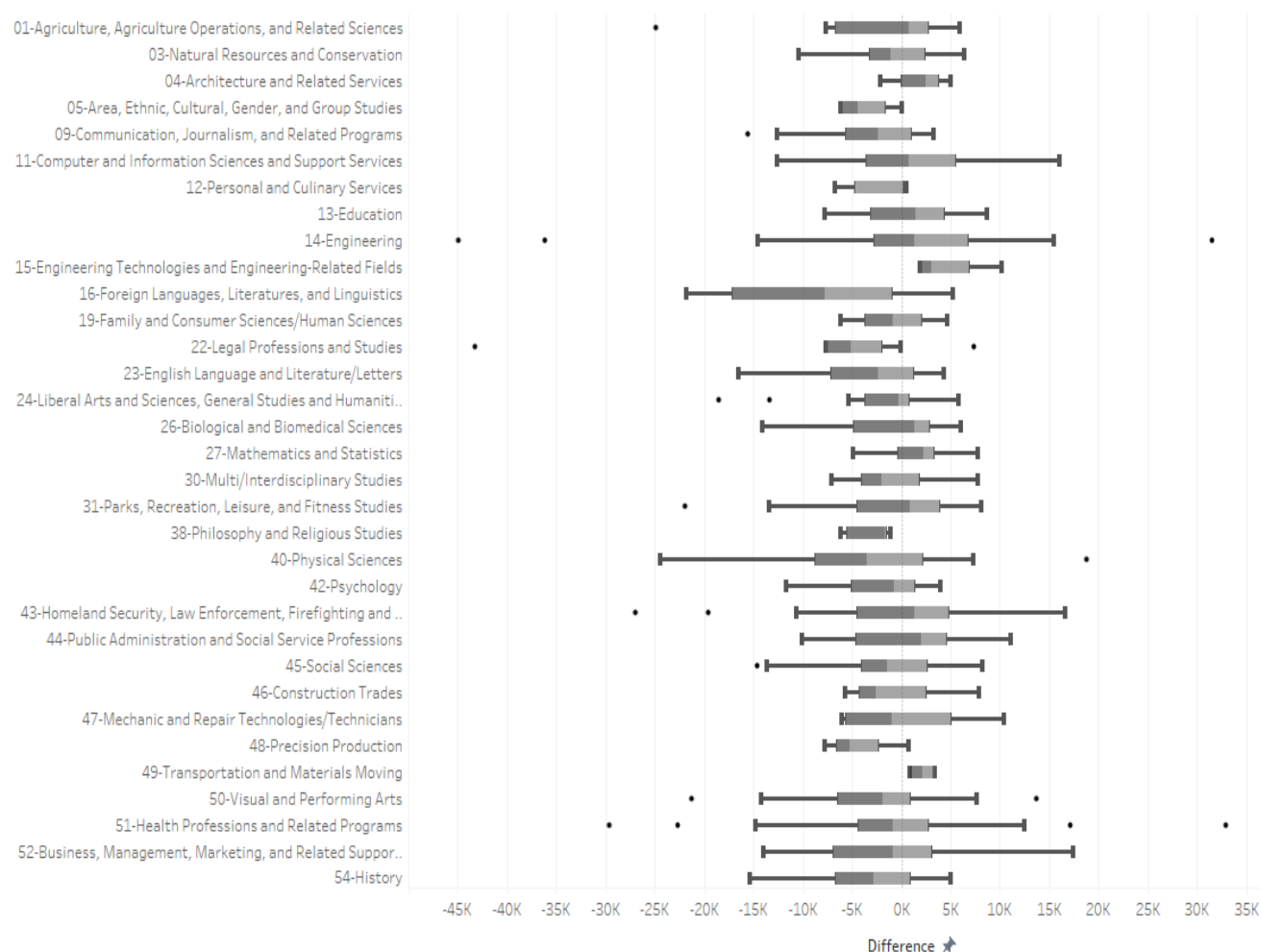
⁵ See https://lehd.ces.census.gov/data/pseo_beta.html

Exhibit 5. Scatterplot of Scorecard median earnings and median earnings from Census PSEO earnings data

Correlation coefficient= 0.9

Source: U.S. Census Bureau. (2019). Post-Secondary Employment Outcomes Data (Beta) (2001-2016) [computer file]. Washington, DC: U.S. Census Bureau, Longitudinal-Employer Household Dynamics Program [distributor], accessed on 9/30/19 at https://lehd.ces.census.gov/data/pseo_beta.html R2019Q2 [version] and U.S. Department of Education, College Scorecard data.

Despite the positive correlation, there is variation by broader categories of field of study. In Exhibit 6, the distribution of differences between Scorecard and PSEO data values are shown by two-digit CIP codes. Overall, considering all field of study differences regardless of the two-digit CIP code grouping, the median difference was -\$640, with an interquartile range from -\$4,977 to \$2,733.

Exhibit 6. Differences in median earnings (Scorecard values minus PSEO values) by 2-digit CIP codes

Source: U.S. Census Bureau. (2019). Post-Secondary Employment Outcomes Data (Beta) (2001-2016) [computer file].

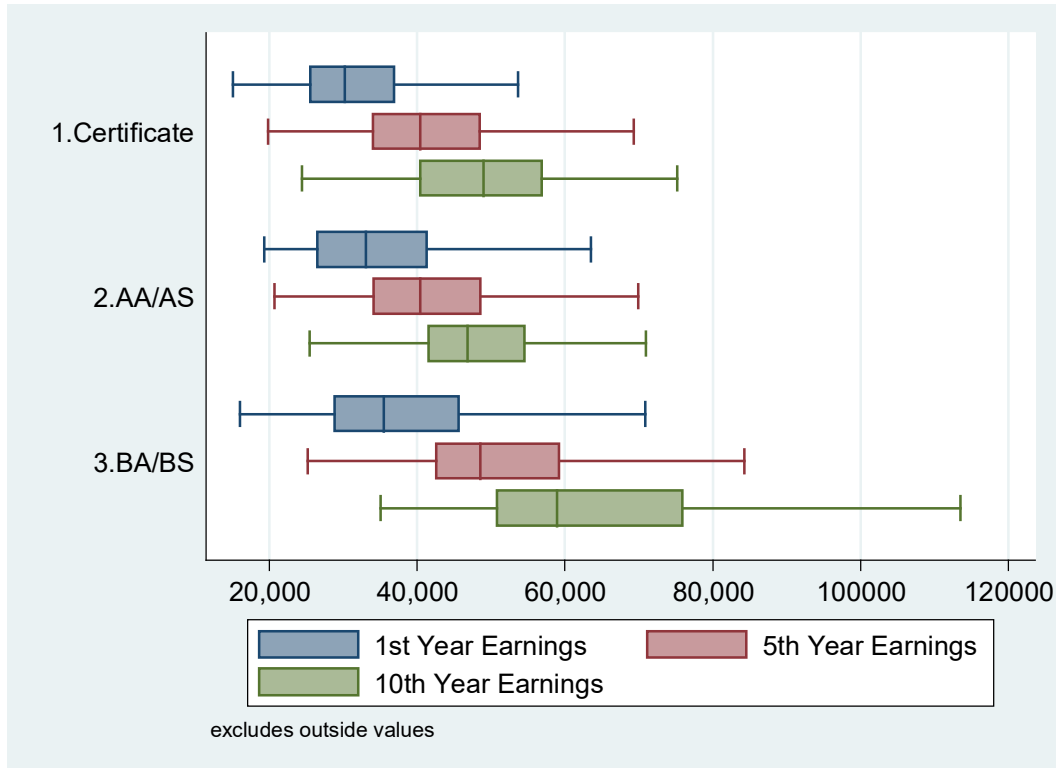
Washington, DC: U.S. Census Bureau, Longitudinal-Employer Household Dynamics Program [distributor], accessed on 9/30/19 at https://lehd.ces.census.gov/data/pseo_beta.html R2019Q2 [version] and U.S. Department of Education, College Scorecard data.

First-year earnings

Scorecard's long-term goal is to measure the annual earnings of former students at multiple points in time up to 10 years post-completion. However, NSLDS did not begin collecting program-level data until the 2014-15 award year, limiting historical cohorts that can be derived from NSLDS by field of study. Further, Scorecard combines students into two-year cohorts to maximize cohort cell sizes in order to minimize privacy suppressions. Therefore, the first complete two-year cohort that can be derived from NSLDS by field of study is the 2014-15 and 2015-16 award years. With 2017 calendar year (and prior) earnings available, Scorecard provides first-year earnings for those who graduated in 2014-15 and 2015-16, measured in calendar years 2016 and 2017, respectively. As more data become available, Scorecard will calculate salary after completion measured multiple years post-completion. Data users should use caution in using these first-year earnings data because they may not be predictive of longer-term earnings outcomes. To illustrate an example of how earnings outcomes may change over time, Exhibit 7

examines earnings 1, 5, and 10 years after completion using PSEO data. Exhibit 7 shows that for fields of study in this set of institutions, the distribution of typical (median) earnings is similar among undergraduate certificate, associate's degree, and bachelor's degree programs in the first year post-completion. While the distributions increase over time for all three credential types, the distribution of typical bachelor's degree earnings increases at a faster rate than the distribution of typical sub-baccalaureate earnings over the next 10 years.

Exhibit 7. Distribution of median earnings 1, 5, and 10 years post-completion for fields of study in PSEO participating institutions by undergraduate credential level



Source: U.S. Census Bureau. (2019). Post-Secondary Employment Outcomes Data (Beta) (2001-2016) [computer file]. Washington, DC: U.S. Census Bureau, Longitudinal-Employer Household Dynamics Program [distributor], accessed on 9/30/19 at https://lehd.ces.census.gov/data/pseo_beta.html. R2019Q2 [version]

Data users should interpret first-year earnings with the understanding that these data values may not be indicative of longer-term earnings outcomes and factors such as credential type may substantially influence the change in earnings over time.

Other examples of how first-year earnings can differ substantially from longer-term earnings are fields of study where graduates frequently go on into residency programs (e.g. medical residency) in the first years after graduation. In cases like these, data users are encouraged to find more information about individual fields of study to provide a fuller context in interpreting earnings data points.

Economic conditions

Data users should also understand that earnings of past completers may not accurately predict benefits generated by schooling because of trends in the overall national and global economy. For example, earnings values calculated during a time of economic prosperity may not reflect how students would fair

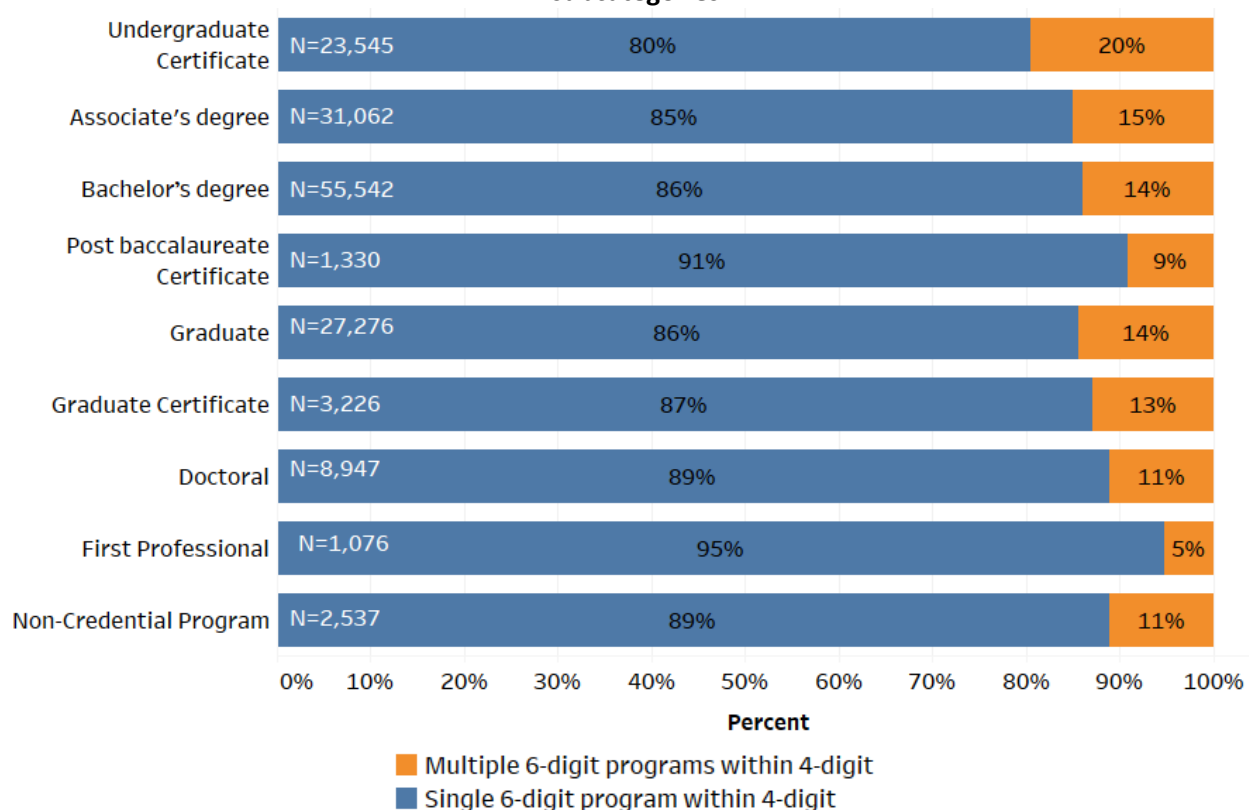
if they graduated during a recession. In addition, specific industries may have their own peaks and valleys based on occupation-specific trends in the labor market where certain credentials may become obsolete while other credentials may skyrocket in demand. Data users are encouraged to examine Scorecard earnings data in conjunction with other online resources such as the [O*NET OnLine](#) to explore the outlook for specific occupations.

Aggregation of data

CIP codes are a taxonomy where academic disciplines are classified by a six-digit code with the first two digits representing the broadest category for a discipline, the first four digits representing a more granular category, and all six digits representing the most granular description of an academic discipline. For example, “Agriculture/Farm Supplies Retailing and Wholesaling” has a six-digit CIP code of 010105, which places it in the category of “Agriculture Business and Management” (four-digit CIP code of 0101), which is a subdivision of “Agriculture, Agriculture Operations, and Related Sciences” (two-digit CIP code of 01).

Even within the most granular category of discipline, institutions often provide different types of program offerings with different course catalog program names. Scorecard uses the first four digits of the CIP code in its calculations. In doing so, Scorecard can provide information in categories that may be easier for prospective students to understand. In addition, Scorecard can provide more information that is not privacy-suppressed because combining subcategories increases cell sizes. The trade-off is loss of granularity in describing individual program offerings by institutions. To the extent that outcomes vary across more granular subcategories, this variation will not be observed using the current methodology.

Few institutions reported students completing in multiple six-digit CIP code subcategories within a four-digit CIP code category. Among the fields of study with at least one completer reported in NSLDS, approximately 15 percent reported students completing in more than one six-digit CIP code subcategory. In addition, nearly 70 percent of credentials conferred were in four-digit CIP code categories with only one six-digit CIP code subcategory with at least one completer. Exhibit 8 shows the percentage of four-digit CIP code offerings with completers in one or multiple six-digit CIP categories by credential level.

Exhibit 8. Percentage of fields of study with students completing in one or multiple six-digit CIP subcategories

Source: National Student Loan Data System as of March 2019.

The Department recognizes separate branch “locations” for all Title IV-eligible institutions that have a Program Participation Agreement, the formal document establishing their eligibility to participate in Title IV programs and assigns each location an eight-digit OPEID. Similar to CIP code construction, the OPEID that identifies individual institutions and their branch locations is hierarchal. Related institutions use the same first six digits of an OPEID, which is tied to the “main campus.” The last two digits identify specific branch locations in a larger family of campuses. While NSLDS is increasingly moving toward reporting student enrollments and aid-receipt at the eight-digit OPEID level, this is not yet universal, and many institutions report information on where students enroll or receive aid only at the six-digit OPEID level (since the main campus is often where aid is managed for all branches of an institution). Given the lack of consistency with reporting to NSLDS at the more granular eight-digit-OPEID level, Scorecard constructs cohorts and calculates data values by field of study using the six-digit OPEID institutional identifier. To the extent that outcomes vary across branch locations, this variation will not be observed using the current methodology.

Monthly estimates

The Scorecard consumer tool provides both an estimate of monthly income and monthly debt payment for each field of study. The monthly income is based on the median earnings divided by 12. While this provides some contextual information for prospective students, data users should not interpret this as a monthly earnings rate because these data do not adjust for individuals who only worked part of the year. In addition, data users should be aware that monthly debt is based on only one of many payment plans available to borrowers. The monthly payment estimate is based on a standard 10-year fixed

payment plan. Data users are encouraged to use the [repayment calculator](#) to better understand the range of monthly payment figures available to borrowers. The standard 10-year payment fixed plan represents a relatively high monthly estimate in comparison to other the payment options.

Recently, the Department rescinded the Gainful Employment (GE) regulations, which compared an annual debt-to-earnings ratio to a threshold level. Data users should exercise caution in making comparisons between Scorecard monthly earnings and debt estimates with the thresholds in the GE regulation because the data-generating process for both debt and earnings is substantially different. For example, the median loan debt that was used in GE was calculated using all federal financial aid recipients, whereas in Scorecard, this is calculated only for federal borrowers. In addition, the earnings calculated in Scorecard represent the earnings one year after completion. In contrast, GE earnings is measured further out in time depending on the type of program. These differences represent just a few of the many differences between Scorecard and GE.

Nonemployment

Earnings data allow users to derive 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. In particular, data users should note that those who did not work may have chosen to be out of the labor market for a variety of reasons.

Data quality

In 2014-15, Federal Student Aid (FSA) began requiring institutions to report program-level enrollment data to NSLDS. While data reporting seems to have stabilized over time, enrollment reporting in the earlier years was likely incomplete. In order to address this, Scorecard published preliminary cumulative loan debt data in May 2019, and the Department asked institutions to make corrections to their own historical NSLDS program-level enrollment data by mid-July 2019. In order to examine the overall data quality of program-level reporting to NSLDS after institutions were able to make historical updates and corrections, the Department compared NSLDS program-level data with other internal FSA reporting and available samples from NCES statistical collections.

To assess consistency of reporting by institutions, the Department examined the process where institutions initially report program-level CIP codes at the time of aid disbursement and then confirm or update those CIP codes for students via NSLDS program-level enrollment reporting; for the years in question, those CIP codes are almost always either confirmed or corrected during that secondary FSA process step. For example, for the 2014-15 award year, 97 percent of CIP codes were confirmed or corrected.

To assess the completeness of reporting by institutions, institution-level completion records in NSLDS can serve as a comparison. Institution-level completion records serve as a trigger for a student entering repayment due to a graduation and should have a corresponding program-level graduation rate. Exhibit 9 provides aggregate ratios of program-level completion records to institution-level completion records if they occur within 180 days for an individual student at an individual institution (that is, for every institutional completion record, how often is there a corresponding program-level completion record for that same student). The exhibit shows that the 2014-15 reporting was less complete than in subsequent years.

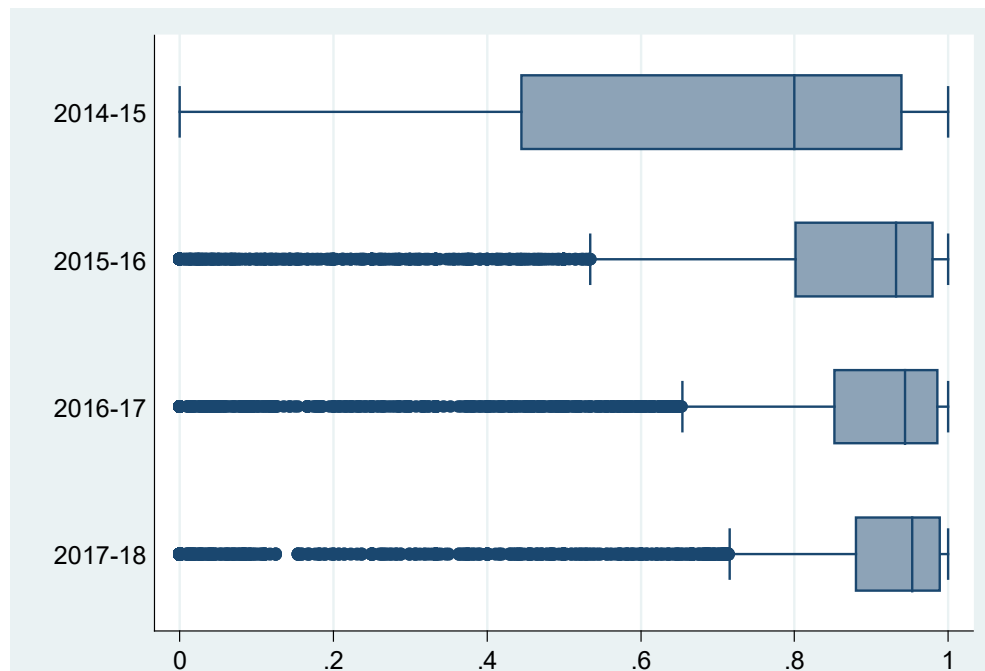
Exhibit 9. Percentage of institution (IHE)-level completion records with corresponding program-level completion records as of mid-July 2019

Award year	Percent
2014-15	73
2015-16	90
2016-17	91
2017-18	92

Source: NSLDS as of July 2019

Further analysis suggests that these ratios are not substantially different across institutional characteristics including institutional control (i.e., public, private, proprietary) and institution size (i.e., number of completers).⁶ In examining the subgroup of students who received grants, but no loans, the results were also similar (85 percent, 87 percent, and 96 percent in 2014-15, 2015-16, and 2016-17, respectively).

Further examination of NSLDS data (Exhibit 10) shows that in the initial year of reporting (2014-15), 80 percent of institutional completion records were also reported at the program level for the typical (median) institution and 25 percent of institutions reported less than 45 percent of institutional completion records at the program level. In the next three years of reporting (2015-16, 2016-17, and 2017-18), the typical institution reported 93, 95, and 95 percent of institutional completion records at the program level, respectively.

Exhibit 10. Distribution of ratios for IHE-level completion records vs program-level completion records

Source: NSLDS as of July 2019

To further explore NSLDS program-level reporting, the Department examined a sample of undergraduate students who began postsecondary education in 2011-12. The 2012/17 Beginning Postsecondary Students Longitudinal Study (BPS:12/17) contains a nationally representative sample of

⁶ Analysis not shown in this document.

first-time beginning postsecondary students who first enrolled in postsecondary education in 2011-12. The sample includes a diverse set of beginning postsecondary students representing different demographic characteristics (e.g., age, gender, etc.), students who study on a full-time or part-time basis, and students who enrolled in multiple types of institutions and degree/certificate programs. BPS:12/17 includes a wealth of information on its sample of students including degree and certificate attainment for up to six academic years (2011-12 to 2016-17) after beginning postsecondary education. As part of their data collection, BPS:12/17 obtained administrative data on degree and certificate attainment from the National Student Clearinghouse (NSC). NSC provides educational reporting, data exchange, data verification, and research services to a majority of institutions of higher education in the United States and has a substantial database with student enrollment data, including graduation records.

Using the BPS sample of students, the Department compared graduation records from NSC with the NSLDS program-level data from which Scorecard field of study cohorts are derived. Exhibit 11 shows the number and percentage of cases that match out of all Title IV–aided BPS sample members. Matches indicate cases where completion records were observed within one month of each other with the same two-digit CIP code and credential level in both data sources. Unmatched events from either source could be due to a difference in records (e.g., different CIP code) or a lack of records in either data source. For example, an unmatched event could represent an NSLDS-reported completion record with no corresponding NSC event because the student was at a school that does not report graduation or CIPs to NSC. Exhibit 11 shows that 65 percent of cases matched in the combined 2014-15, 2015-16, and 2016-17 academic years and the number of unmatched cases due to a record in NSC but not NSLDS was similar to the number of unmatched cases due to a record in NSLDS but not NSC. Exhibit 11 also shows that bachelor’s degree cases had a higher match rate than sub-baccalaureate cases. However, because the sample of students is based on a 2011-12 beginning postsecondary sample and completion records are evaluated much later (beginning in 2014-15), the credential levels with shorter program lengths (i.e., undergraduate certificates and associate degrees) have small sample sizes, and their match rates should be interpreted with caution.

Exhibit 11. BPS:12/17 program-level completion match status of Title IV–aided students, by academic year, and credential level

			Matched		Unmatched		
	Academic Year	Total Count	Count	Percent Matched	Count	Percent of unmatched in NSC, but not in NSLDS	Percent of unmatched in NSLDS, but not in NSC
Overall	2014-15	3,140	2,110	67	1,040	52	48
	2015-16	2,290	1,450	63	840	46	54
	2016-17	1,120	700	63	420	56	44
	All years	6,550	4,260	65	2,290	51	49
UG Certificate	2014-15	170	30	15	150	66	34
	2015-16	150	30	20	230	67	33
	2016-17	100	20	23	80	69	31
	All years	420	80	19	340	67	33
Associate's Degree	2014-15	410	170	43	230	64	36
	2015-16	400	200	51	200	54	46
	2016-17	270	140	52	130	56	44
	All years	1,080	510	47	560	59	41
Bachelor's Degree	2014-15	2,570	1,910	74	660	45	55
	2015-16	1,740	1,220	70	530	39	61
	2016-17	750	540	72	210	51	49
	All years	5,060	3,670	73	1,390	44	56

Notes: Unweighted sample sizes rounded to the nearest 10. Detail may not sum to totals due to rounding. Percent matched columns are calculated with unrounded numbers so they may not match exactly.

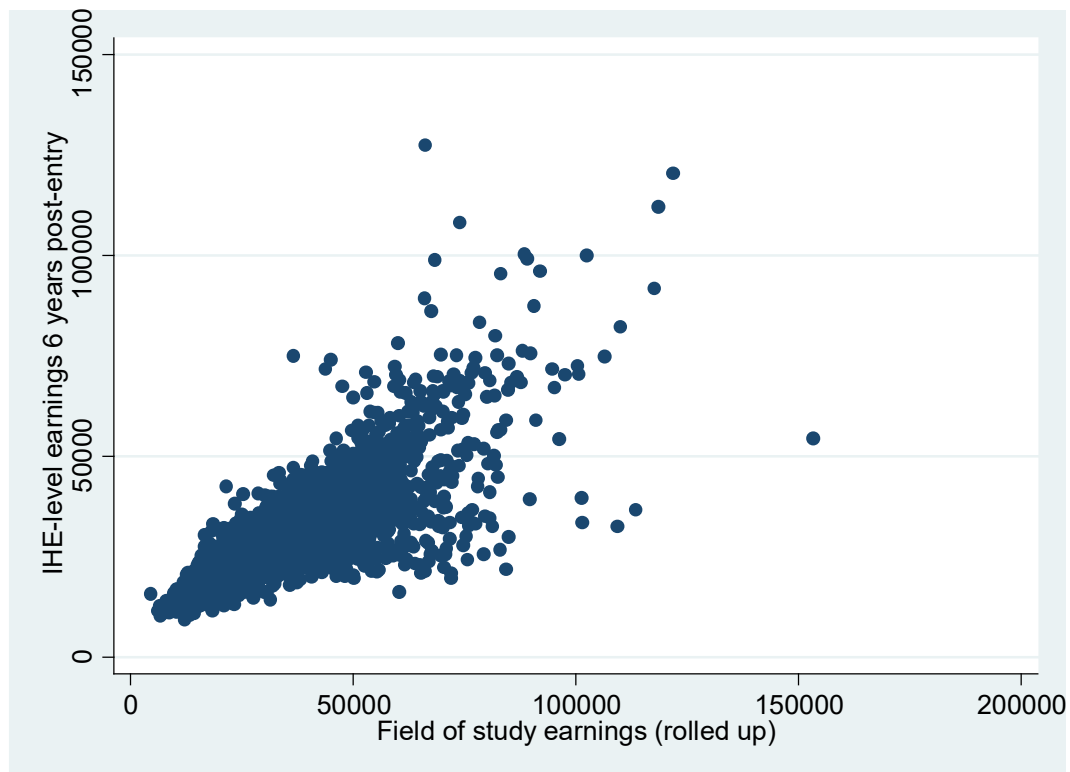
Source: U.S. Department of Education, National Center for Education Statistics, 2012/17 Beginning Postsecondary Students Longitudinal Study (BPS:12/17) unpublished data. (This table was prepared in August 2019.)

To assess the overall external validity of data generated from program-level reporting in NSLDS, earnings calculations by field of study were compared to other earnings calculations derived from a different cohort-generating process. In previous years, IRS provided Scorecard earnings data at the institution level. Previously published institution-level earnings data were based on an entry cohort and included any student who entered into the institution regardless of subsequent completion, transfer, or withdrawal. These institution-level earnings were measured at 6, 8, and 10 years post-entry. Given the differences in cohort construction and measurement between the field of study-level and institution-level earnings data, comparisons across these two data points should show differences. However, to

examine the overall concordance between these two data points, Exhibit 12 shows the relationship between field of study–level data aggregated to the institution level (weighted average of field of study–level values) and the previously published institution-level median earnings calculations measured six years post-entry. Exhibit 12 shows that the earnings measurements from the two sources are highly correlated.

Exhibit 12. Scatterplot of Scorecard field of study–level and IHE-level earnings values

Correlation Coefficient: 0.79, n= 3555



GE calculations published in 2017 provide program-level earnings for all programs at proprietary institutions and nondegree programs at public and private nonprofit institutions. As mentioned previously, the data-generating process for GE earnings data is substantially different from the IRS field of study–level Scorecard process. For example, the unit of analysis for GE calculations is based on a six-digit CIP code, whereas the unit of analysis for Scorecard is based on a four-digit CIP code. Other key differences include the fact that GE measures earnings at a further point in time post-completion and the fact that GE cohorts include those who were not working (Scorecard cohorts include only those who worked). Exhibit 13 shows the relationship between GE median earnings calculations aggregated to the four-digit CIP code (weighted average) and the IRS field of study–level scorecard data. The scatterplot indicates a high degree of correlation between the two data sets.

Exhibit 13. Scatterplot of Gainful Employment comparison values and IRS earnings
Correlation Coefficient: 0.92, n= 5032

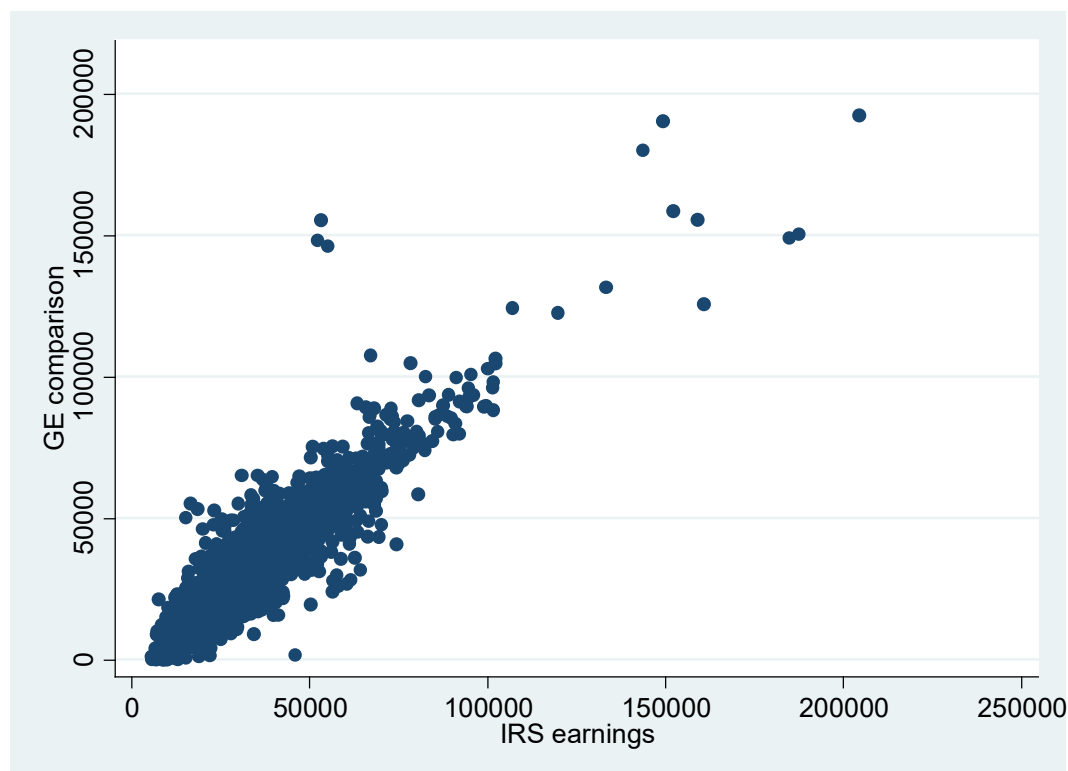


Exhibit 14 examines the correlation coefficients at specific credential levels and shows a high level of agreement within each credential level category, except for first professional degrees. This could be explained by the low number of observations that could be compared in both data sets (15) or by the fact that the GE data-generating process is different for medical residency programs (measurements are taken further out after graduation), which may fall under first professional degrees.

Exhibit 14. Correlation coefficients for Gainful Employment earnings (rolled up) compared with IRS Scorecard earnings, by credential level

Credential level	Correlation Coefficient	Observations
1: Undergraduate Certificate or Diploma	0.88	3,428
2: Associate's Degree	0.88	818
3: Bachelor's Degree	0.85	425
4: Postbaccalaureate Certificate	0.85	7
5: Master's Degree	0.94	204
6: Doctoral Degree	0.82	40
7: First Professional Degree	0.43	15
8: Graduate/Professional Certificate	0.94	95

Source: NSLDS and U.S. Department of Education, College Scorecard Data

Matching fields of study reported to different data systems

NSLDS-identified completers should also appear in the IPEDS Completions component for the corresponding cohort (award year) period. However, the NSLDS and IPEDS classifications of students may vary due to definitional differences between the two systems, and data users should exercise caution in comparing NSLDS-derived counts with reported IPEDS counts.

To align the two data sources, IPEDS Completions data values were aggregated across award levels to correspond to the NSLDS credential levels (as noted in Exhibit 3). This mapping, however, is subject to bias including, but not limited to, the following situations:

- The IPEDS Postbaccalaureate Certificate category may contain some awards that NSLDS classifies as Post-Baccalaureate Certificates and others that NSLDS classifies as Graduate/Professional Certificates. Examining the data, the reported fields of study in IPEDS more closely aligned with the Graduate/Professional Certificate classification.
- The NSLDS Post-Baccalaureate Certificates category may contain awards that IPEDS classifies as undergraduate certificates (e.g., teacher licensure programs). Examining the data, there was no support for mapping IPEDS undergraduate certificates to the NSLDS Post-Baccalaureate Certificate category.
- IPEDS has not collected a First-Professional Degree category since the 2009-10 collection. These degrees are primarily reported in the Doctor's Degree – Professional Practice or Doctor's Degree – Other categories, but ultimately classification of formerly first-professional degrees is up to the institution, and the two doctoral degree categories mentioned may contain awards that were not previously regarded as first-professional. Additionally, some non-doctoral-level programs are considered first-professional (e.g., Master's of Divinity). As a result, there is some bias in the matching of these two IPEDS categories to the NSLDS First Professional Degree category.

While CIP code definitions do not vary across systems, there is some evidence that institutions may report graduates under different CIP codes. Exhibit 15 shows the increase in the number of fields of study before and after adding fields of study that exist only in IPEDS. These counts include non–Title IV fields of study and those where no borrowers completed, which would explain some of the increase. However, the marked increase, especially at the undergraduate certificate level, suggests there is additional mismatching between the two data sources.

Exhibit 15. Number of fields of study in the NSLDS (PL) data only, and number of fields of study after matching to IPEDS Completions (C) component data

		Public			Private			Proprietary			Foreign
		4+ years	2 to 3 year	Less-than-2-year	4+ years	2 to 3 year	Less-than-2-year	4+ years	2 to 3 year	Less-than-2-year	
Undergraduate Certificate	PL-only	3,042	13,715	1,134	1,006	230	127	718	1,724	1,814	35
	C+PL	5,853	22,182	1,711	2,265	293	158	974	2,001	2,000	35
Associate's degree	PL-only	4,850	21,430	29	2,037	273	8	1,227	1,142	56	10
	C+PL	5,993	26,515	31	3,230	324	8	1,448	1,274	59	10
Bachelor's degree	PL-only	23,832	68	0	29,493	9	1	1,194	41	2	902
	C+PL	25,983	69	0	34,714	11	1	1,390	48	3	902
Post baccalaureate Certificate	PL-only	725	14	1	531	2	1	34	2	2	18
	C+PL	3,963	17	1	2,795	2	1	174	2	2	18
Graduate	PL-only	14,077	15	0	10,170	2	0	512	6	5	2,489
	C+PL	16,517	15	0	13,026	3	0	604	6	5	2,489
Graduate Certificate	PL-only	1,746	7	3	1,309	0	0	100	4	7	50
	C+PL	2,606	7	3	2,231	0	0	127	4	7	50
Doctoral	PL-only	5,459	1	1	2,790	2	0	132	0	0	562
	C+PL	6,730	1	1	3,563	2	0	162	0	0	562
First Professional	PL-only	506	1	0	457	1	0	26	0	2	83
	C+PL	847	1	0	914	1	0	51	0	2	83

Source: National Student Loan Data System and Integrated Postsecondary Education Data System as of March 2019.