Technical Documentation: College Scorecard Data by Field of Study

Contents

About 1

Data Methodology 1
Defining field of study 1
Number of credentials conferred 2
Post-completion earnings 3
Cumulative loan debt 4
Repayment rate 6
Privacy protection 7
Data files available 8

Data Sources and Limitations 9
Federally aided students 9
Earnings timepoints 10
Economic conditions 11
Aggregation of data 12
Monthly estimates 12
Nonemployment 13
Matching fields of study reported to different data systems 13

Appendix A – Excerpts from the November 2019 documentation report 15
Motivation for Calculating Earnings by Field of Study 15
Comparison of Earnings Data Against External Sources 16
Frequency of Six-digit CIP Codes Within Four-digit CIP Code-based Fields of Study 18
Data quality 19

Appendix B – Summary of documentation changes in this version 25

About

The primary purpose of the Scorecard is to provide data to help prospective postsecondary students make informed enrollment decisions. 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 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, debt, and repayment metrics 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 (i.e., number of completers per year, median earnings, median cumulative debt at the reference school, median cumulative debt at any school attended) 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 institutional identifiers, a four-digit Classification of Instructional Programs (CIP) code, and a credential level. Institutional identifiers 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**[2]**)**, 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 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.

[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.

- 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
- In addition to the above identifiers, each field of study is assigned an indicator that describes the availability of distance education programs within that field of study (**DISTANCE**). This indicator identifies fields of study where all credentials in the field of study may be fully completed via distance education and fields of study where at least one, but not all credentials in the field of study may be fully completed via distance education. Because there may be multiple programs available within a field of study, after identifying that a *field* contains distance education programs the user will need do further research to determine if their specific *program* of interest is available to be completed entirely via distance education.

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 NLSDS credential levels, some aggregation of IPEDS award levels was needed. Exhibit 1 indicates the recoding and aggregation scheme applied to IPEDS to calculate the number of awards provided in the field of study data.

Exhibit 1. Correspondence of NSLDS credential level to IPEDS award level[3]

NSLDS Credential Level	IPEDS Completions Data Award Levels Aggregated
1: Undergraduate Certificates or Diplomas	Certificates of less than 1 year, at least 1 but less than 2 years, or at least 2 but less than 4 years
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

Scorecard data describes the earnings of graduates one to seven years after completing a credential (**EARN_MDN_HI_1YR or EARN_MDN_1YR, EARN_MDN_HI_2YR, EARN_NE_MDN_3YR, EARN_MDN_*x*YR, x=4, 5, and 7**). The Department acknowledges that earnings in the early years after graduation may not be indicative of longer-term earnings, and so it plans to provide earnings data points for graduates several years after completion. Achieving this goal will take time.

The cohort of evaluated graduates for earnings metrics consists of those individuals who received federal financial aid, but excludes[4] those who:

- were subsequently enrolled in school during the measurement year,
- died prior to the end of the measurement year, or
- did not work during the measurement year[5]

Earnings data also include the number of completers who worked and were not enrolled in school during the measurement year (**EARN_COUNT_WNE_ *x*YR**) and the number of completers who worked and were not enrolled in school during the measurement year that earned more than a high school graduate (**EARN_GT_THRESHOLD_*x*YR**) were included in this release. These count metrics were calculated for *x*=1, 4, 5, and 7 years after completion of a credential for various cohorts. The total number of graduates who were not enrolled in school during the measurement year is provided for the three-year earnings cohort (**EARN_COUNT_NE_3YR**).

[3] See *Matching fields of study reported to different data systems* for more information.

[4] Earnings metrics that combine calendar year (CY) 2016/CY2017 earnings also excluded those who had loan deferments for military service during the measurement year. Exclusions for military deferments did not apply to metrics that combined CY2017/CY2018 earnings. The two years of calculations that combined CY2017/CY2018 and CY2018/CY2019 earnings also excluded those who received a higher-level credential than the credential level of the field of study measured. For example, if a student completed a bachelor's degree and then completed a master's degree, that student would be excluded from the bachelor's degree calculation. These rules for exclusions apply even when metric names do not change (e.g., the PELL/NOPELL and MALE/NOMALE disaggregations of earnings).

[5] The three-year earnings measurement combining calendar year (CY) 2018/CY2019 included those who did not work during the measurement year.

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

Earnings data within credential level and CIP Code by at the national level were calculated 4, 5, and 7 years after completion. The number of completers working and not enrolled, the median earnings, and the 25th and 75th percentiles of earnings were calculated and are included in the data file (EARN_COUNT_WNE_*x*YR_NAT, EARN_MDN_*x*YR_NAT, EARN_P25_*x*YR_NAT, EARN_P75_*x*YR_NAT). The goal of presenting these values is to allow for information about fields of study to be disseminated when examining these fields at the institution level might not be possible due to small counts.

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 Schedules SE (Self-Employment Tax) for each student measured. Earnings values are presented in inflation-adjusted dollars. The reference year for inflation adjustment is one year following the latest measurement year for the cohort (e.g., earnings measured in calendar years 2022 and 2023 were adjusted to 2024 dollars).

Cumulative loan debt

Scorecard provides loan debt metrics associated with individuals who were aided through Title IV federal loan programs who complete college in the same two consecutive award year timeframe. These metrics describe number of graduates and the mean and median loan amounts disbursed for two types of loans: 1) Parent PLUS loans (**DEBT_[*disaggregation*]_PP_[*method*]_N, DEBT_[*disaggregation*]_PP_[*method*]_MEAN, DEBT_[*disaggregation*]_PP_[*method*]_MDN**) and 2) Direct Subsidized and Unsubsidized Loans (also known as Stafford Loans) [6] combined with Graduate Direct PLUS loans (**DEBT_[*disaggregation*]_STGP_[*method*]_N, DEBT_[*disaggregation*]_STGP_[*method*]_MEAN, DEBT_[*disaggregation*]_STGP_[*method*]_MDN**). Perkins Loans are excluded from the cumulative loan debt calculations.

Scorecard field of study debt metrics are provided for all graduates (**[*disaggregation*]=ALL**), disaggregated by Pell grant recipient status of the student associated with each loan[1,2] (**[*disaggregation*]=PELL/NOPELL**), and disaggregated by the sex of the student associated with each loan[1,2] (**[*disaggregation*]=MALE/NOTMALE**[7]). There are two methods of debt measurement included in these metrics. Debt statistics for loans originated only at the institution where the debt is reported (**[*method*]=EVAL**) and loans originated at any institution the student attended (**[*method*]=ANY**) are calculated for all levels of disaggregation.

For example, if a female Pell-recipient graduate borrowed $2,000 per semester for two years in Direct Loans at College A, then transferred to College B and borrowed $3,000 per semester for one year in Direct Loans and completed a credential, the $2,000 x 2 semesters x 2 years = $8,000 cumulative debt at College A would not be recorded in metrics for College A since the student did not complete a credential at that school. Since the graduate completed a credential at College B, these cumulative debts would be included in College B's metrics

[6] 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.

[7] Measurements by sex prior to the award year 2021-22, 2022-23 pooled cohort included records with no sex indicated along with females in the NOMALE metric. Beginning with the award year 2021-22, 2022-23 pooled cohort, the NOMALE metric only includes females. Records with no sex indicated are proportionally allocated to the MALE and NOMALE categories based on the institution's male and female distribution.

as follows: $3,000 x 2 semesters x 1 year = $6,000 would be included in calculations leading to DEBT_ALL_STGP_EVAL_MDN, DEBT_ALL_STGP_EVAL_MEAN, DEBT_PELL_STGP_EVAL_MDN, DEBT_PELL_STGP_EVAL_MEAN, DEBT_NOTMALE_STGP_EVAL_MDN, and DEBT_NOTMALE_STGP_EVAL_MEAN. The graduate's $8,000 + $6,000 = $14,000 total cumulative debt would be included in College B's DEBT_ALL_STGP_ANY_MDN, DEBT_ALL_STGP_ANY_MEAN, DEBT_PELL_STGP_ANY_MDN, DEBT_PELL_STGP_ANY_MEAN, DEBT_NOTMALE_STGP_ANY_MDN, and DEBT_NOTMALE_STGP_ANY_MEAN calculations.

The cumulative loan debt includes only loan disbursement amounts for the relevant loan program 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 loan debt for a particular field of study includes all debt dispersed for the relevant loan program at the same academic level as the credential level for the field of study, regardless of whether the graduate switched fields of study during their time at the institution. For example, if a graduate 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.

Individuals 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. Those who completed multiple awards within the same institution at the same four-digit CIP code and credential level were measured only once. Graduates who completed in multiple institutions were included at each institution where they were awarded a credential.

The estimated monthly payment is calculated for all graduates for both types of loans and both methods of debt measurement (**DEBT_ALL_PP_ANY_MDN10YRPAY, DEBT_ALL_PP_EVAL_MDN10YRPAY, DEBT_ALL_STGP_ANY_MDN10YRPAY, DEBT_ALL_STGP_EVAL_MDN10YRPAY**) and is based on the median cumulative loan debt amortized over 10 years using interest rates available at the time of calculation. While borrowers have many payment plan options beyond a 10-year fixed payment plan[8], 10 years is considered a standard payment plan and provides a higher estimate for the monthly payment. The higher estimated payment is conservative for budget planning purposes. For undergraduate programs, calculations are based on the interest rate at the time of calculation for undergraduate Direct Subsidized Loans and undergraduate

[8] For additional repayment options, see https://studentaid.gov/loan-simulator/.

Direct Unsubsidized Loans (6.52%). For graduate programs, calculations are based on the interest rate at the time of calculation for Direct PLUS Loans for graduate and professional students (9.07%). 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 higher. This scenario provides users with a higher estimate of a monthly payment based on a standard repayment plan. Parent PLUS Loan calculations are based on the interest rate at the time of calculation for Direct PLUS loans (9.07%).

Repayment rate

Scorecard provides 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. These metrics (**BBRR[*yr*]_FED_COMP_[*status*]**) aim to allow the user to examine the full picture of what happens to borrowers as they attempt to satisfy their loan obligations under the Direct Loan program, measured two to four years ([*yr*]=2, 3, or 4) after entering repayment. 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 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 [*yr*]*365 days have elapsed since that borrower entered repayment. For example, 2-year repayment rates are measured 730 days after entry into repayment.

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 cumulative debt metrics).

Repayment rates with borrowers evaluated in award years 2018-19 and 2019-20 should be interpreted with the understanding this 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.

Privacy protection

Data presented by field of study have been treated to reduce the risk of disclosure of confidential or identifying information. Debt metrics have been treated via cell suppression methods, and to further reduce disclosure risk, information about the specific suppression rules is not available to the public.

Earnings metrics have been treated using a perturbative approach based on differentially private methods.

- For the earnings counts including the count of non-enrolled workers in an earnings cohort and the count of non-enrolled workers earning more than a high school graduate, a differentially private algorithm was used to apply noise to the data. 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 fourth 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. Data users should exercise caution in calculating percentages, particularly in smaller cohorts where absolute differences of just one or two in either the numerator and/or denominator can more substantially perturb percentage estimates.

- For median income data values, noise was applied to each observation through a differential private algorithm that was independent of the algorithm used for counts. Following an elimination of cases based on small cell sizes, further suppression techniques were applied to ensure fitness-for-use of the data. Using the undisclosed IRS median income values, 'noisy' median incomes with relative errors above a certain threshold were then suppressed. For example, the following table presents the breakdown of relative error for the median calculations:

  Exhibit 2. Percentage distribution of calculated institution-level field of study medians, by years after completion and percentage difference range.

Difference compared to unperturbed value	Years after completion			
	1 year (measured in CY2022 and CY2023)	4 years (measured in CY2022 and CY2023)	5 years (measured in CY2022 and CY2023)	7 years (measured in CY2022 and CY2023)
Less-than 1 percent	64%	68%	62%	60%
1 to 4 percent	31%	28%	33%	34%
5 to 9 percent				
Percent	5%	4%	5%	6%
Median difference	$ 2,795	$ 3,436	$ 3,736	$ 4,068

Exhibit 3. Percentage distribution of calculated national-level field of study 4-year post-completion median earnings, by percentage difference range.

Difference compared to unperturbed value	Years after completion			
	1 year (measured in CY2022 and CY2023)	4 years (measured in CY2022 and CY2023)	5 years (measured in CY2022 and CY2023)	7 years (measured in CY2022 and CY2023)
Less-than 1 percent	81%	84%	82%	80%
1 to 4 percent	16%	14%	15%	17%
5 to 9 percent				
Percent	3%	2%	2%	3%
Median difference	$3,356	$3,313	$4,094	$4,924

- Reported metrics may not satisfy expected relationships in a small number of cases due to the perturbation process. For example:
    - perturbed counts in the disaggregates (by Pell Grant recipient status or sex) may not sum to the reported corresponding total count.
    - the count of workers earning more than a high school graduate may be higher than the count of all workers in the cohort.
    - median earnings values for subgroups may both be greater-than or both be less-than the overall median value.

Any debt or earnings data points suppressed for privacy are indicated by the "PS" data code.

Borrower-based repayment rate data were suppressed for privacy, and unsuppressed values have been rounded and, in some cases, coarsened into ranges. Because of the rounding and coarsening in to ranges, less data needed to be suppressed, allowing more information to be reported in the data files.

Data files available

The field of study data are organized into data files labeled according to the award years included in the pooled debt cohort. Additionally, a data file is available that contains the most recent data for each metric, regardless of which of the "yearly" data files those data are from. For example, some earnings metrics have been discontinued because of changes to the metric definitions over time. Not all "yearly" files will have data for these metrics, but the "most recent" data file will have data for all earnings metrics, pulling that information from the last time it was calculated.

Data are matched to IPEDS institutions slightly differently in the "yearly" files compared to the "most recent" file. Matching in the "yearly" files is based on the OPEID/UNITID relationships during that time period. The "most recent" file is matched according to the latest OPEID/UNITID relationships available via the Scorecard institutional data. As a result, when comparing data in the "most recent" data file to the appropriate "yearly" files there may be some unexpected differences due to changes in the OPEID/UNITID relationships.

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 loan debt data elements 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 an annual survey consisting of interrelated components, conducted by the National Center for Education Statistics (NCES). The completion of all applicable IPEDS survey components is mandatory for 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 Completions survey component. In this component, 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 undergraduates between the ages of 18 and 24 who were in their 4th year or above had nonfederal loans in 2015-16.[9] 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 3 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 4-year institutions, 59.8, 64.4, and 77.7 percent of students received federal aid in public, private, and proprietary institutions, respectively. Exhibit 3 also shows the number of undergraduates in each sector. The sectors with the highest number of undergraduates include public 2-year institutions (7,639,500) and public 4-year institutions (6,838,300).

[9] See https://nces.ed.gov/Datalab/TablesLibrary/TableDetails/12639?keyword=nonfederal&rst=true

Exhibit 3. Number and percentage of undergraduates receiving federal aid by Institutional Sector 2015-16

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

! 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=bkkbmaaa6 (requires registration/log in to access).
Source: U.S. Department of Education, National Center for Education Statistics, 2015-16 National Postsecondary Student Aid Study (NPSAS:16).

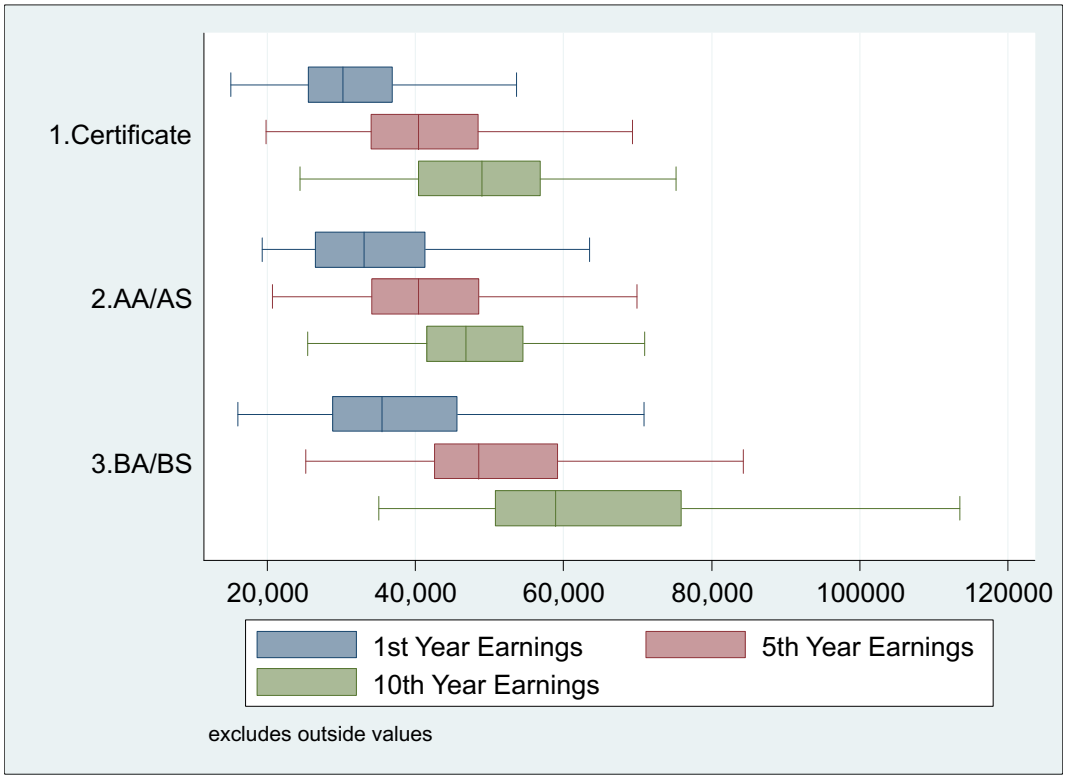
For more information on calculations for federally aided students compared to calculations for all students, see *Comparison of Earnings Data Against External Sources* in Appendix A.

Earnings timepoints

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 2-year cohorts to increase cohort cell sizes and decrease privacy suppressions. Due to these limitations, Scorecard can only provide earnings up to four years after graduation at this time. As more data become available, Scorecard will calculate salary after completion measured multiple years post-completion. Data users should use caution in using earnings data that are measured close to graduation because they may not be predictive of longer-term earnings outcomes. To illustrate an example of how earnings outcomes may change over time, Exhibit 4 examines earnings 1, 5, and 10 years after completion using Census Bureau's Postsecondary Employment Outcomes (PSEO) program[10] data. Exhibit 4 shows that for fields of study in this set of institutions, the distribution of median earnings is similar among undergraduate certificate, associate's degree, and bachelor's degree programs in the first year after completing a credential. While both the median earnings and the spread increase over time for all three credential types, the change over 10 years is pronounced when measuring earnings for bachelor's degree recipients compared to the change when measuring earnings for recipients of sub-baccalaureate credentials.

[10] See https://lehd.ces.census.gov/data/pseo_beta.html

Exhibit 4. Distribution of median earnings 1, 5, and 10 years after 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]

In summary, users should interpret earnings soon after graduation 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.

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[11] to explore the outlook for specific occupations.

[11] See https://www.onetonline.org/.

Aggregation of data

CIP codes

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 01.0105, which places it in the category of "Agriculture Business and Management" (four-digit CIP code of 01.01), which is a subdivision of "Agriculture, Agriculture Operations, and Related Sciences" (two-digit CIP code of 01). Note that in the Scorecard data files, the decimal point has been removed from CIP codes.

Even within the most granular CIP code classification, 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 because combining subcategories increases cell sizes, leading to fewer privacy suppressions. The trade-off is loss of granularity in describing individual program offerings by institutions. To the extent that outcomes vary across these more granular subcategories, this variation will not be observed using the current methodology. See Appendix A for related analyses conducted while developing the analysis plans for Scorecard field of study data.

OPEIDs

The Department recognizes separate branch "locations" for all Title IV-eligible institutions that have a Program Participation Agreement[12] and assigns each location an eight-digit OPEID. Similar to CIP code construction, the OPEID that identifies individual institutions, and their branch locations, is hierarchical. Digits two through six of the OPEIDs of related institutions are the same, which is tied to the "main campus." The first digit and 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, many institutions still 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 a six-digit OPEID institutional identifier created by adding a '0' to the base digits (in positions 2 through 6) of related institutions' eight-digit OPEID numbers. 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 payments 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. Monthly payment estimates are based on a standard 10-year fixed payment plan, which provides a higher estimate for the monthly payment. The higher estimated payment is conservative for budget planning

[12] The Program Participation Agreement is the formal document establishing an institution's eligibility to participate in Title IV programs.

purposes. Data users are encouraged to use the loan simulator to better understand the range of monthly payment figures available to borrowers.

In July 2019, 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 threshold 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 at most two years after completion. In contrast, GE earnings are 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 include information on the number 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 using this to calculate 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.

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 1). 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 of Divinity). As a result, there is some bias in the matching of the 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 5 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 students completed that were associated with federal loans, which would explain some of the increase. However, the marked increase, especially at the sub-baccalaureate level, suggests there may be additional mismatching between the two data sources.

Exhibit 5. Number of fields of study in the award year 2022-23 and award year 2023-24 pooled cohort NSLDS data, and number after matching to IPEDS Completions component data

		Public			Private			Proprietary			
		4+ years	2 to 3 years	Less than 2 year	4+ years	2 to 3 years	Less than 2 year	4+ years	2 to 3 years	Less than 2 year	Foreign
Undergraduate Certificate or Diploma	NSLDS Only	6,131	10,191	944	1,187	280	128	1,003	1,557	1,832	13
	NSLDS+IPEDS	10,117	15,905	1,660	1,943	317	149	1,113	1,678	1,948	13
Associate's Degree	NSLDS Only	7,454	14,758	50	2,194	149	3	1,162	655	16	6
	NSLDS+IPEDS	9,867	19,344	116	2,700	184	3	1,242	722	16	6
Bachelor's Degree	NSLDS Only	27,091	10	0	30,371	1	0	1,183	7	0	1,276
	NSLDS+IPEDS	28,145	15	0	32,998	1	0	1,264	7	0	1,276
Post-baccalaureate Certificate	NSLDS Only	1,023	5	0	505	2	2	27	1	0	9
	NSLDS+IPEDS	1,023	5	0	505	2	2	27	1	0	9
Master's Degree	NSLDS Only	14,067	0	0	10,725	0	0	711	0	1	2,077
	NSLDS+IPEDS	17,138	0	0	13,170	0	0	826	0	1	2,077
Doctoral Degree	NSLDS Only	4,661	0	0	2,255	0	0	129	0	0	355
	NSLDS+IPEDS	6,849	0	0	3,656	0	0	149	0	0	355
First Professional Degree	NSLDS Only	503	0	0	441	0	0	24	0	0	65
	NSLDS+IPEDS	863	0	0	1,033	0	0	67	0	0	65
Graduate/Professional Certificate	NSLDS Only	2,189	0	0	1,471	0	0	204	0	0	38
	NSLDS+IPEDS	5,835	1	0	3,518	1	4	311	1	0	38

Source: National Student Loan Data System and Integrated Postsecondary Education Data System as of October 2026.

Please see Appendix A for additional, historical data quality information.

Appendix A – Excerpts from the November 2019 documentation report

This appendix contains information and analyses that were presented in the November 2019 documentation report that are still relevant for the currently published data.

Motivation for Calculating Earnings by Field of Study

During the 2015 Scorecard redesign, Scorecard 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 A1 describes the differences in earnings estimates by credential level.

Exhibit A1. 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 (November 2019).

Scorecard's 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 after graduation.

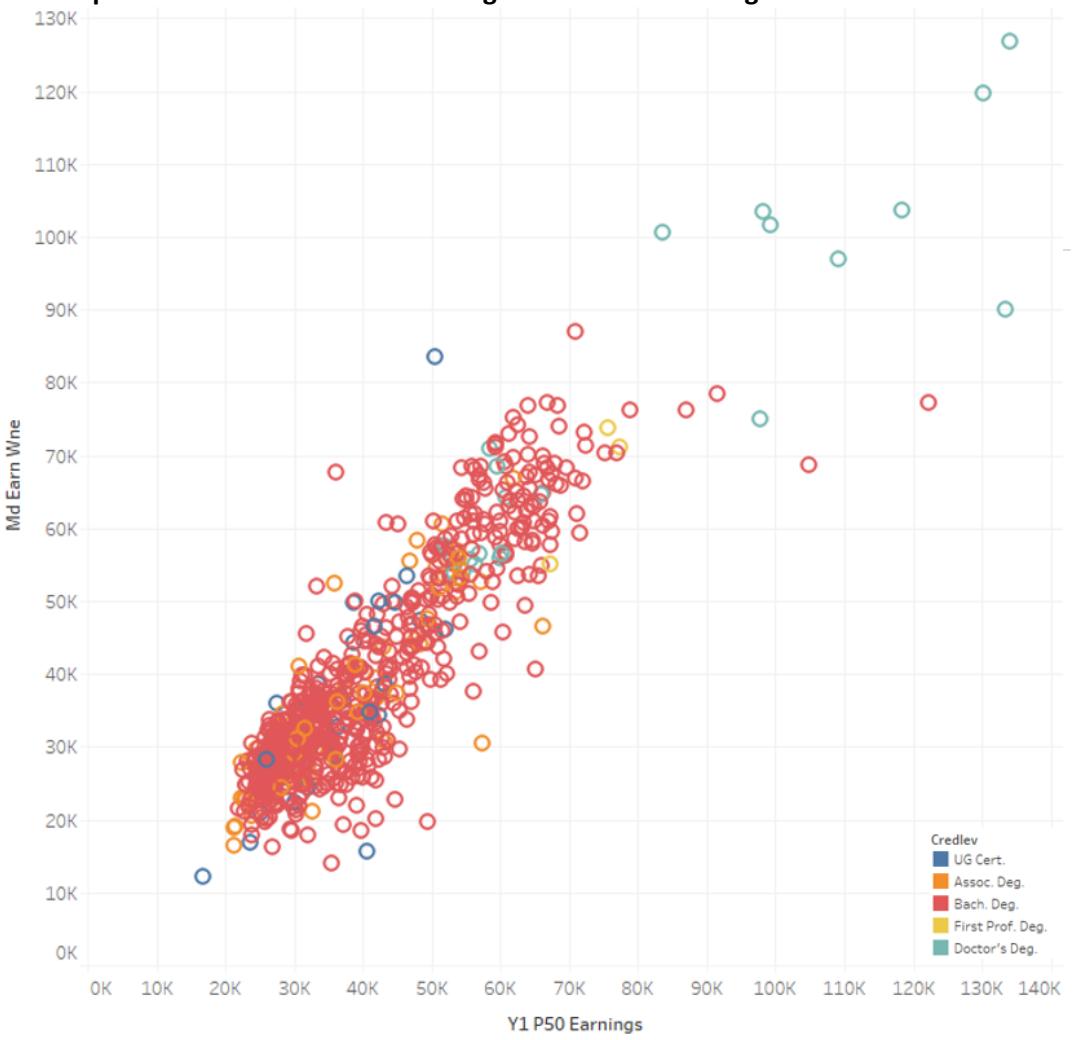
In order to provide more comprehensive data to prospective students, Scorecard will continue to develop data metrics prescribed by the Executive Order. For now, Scorecard is releasing information on post-completion earnings and cumulative debt by field of study.

Comparison of Earnings Data Against External Sources

The 2019 Scorecard earnings data by field of study were compared with publicly available data from the Census Bureau's Postsecondary Employment Outcomes (PSEO) program,[13] 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 some 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 A2, the positive correlation between the Scorecard and PSEO data suggests that the two estimates from the two different sources are generally in concordance.

[13] See https://lehd.ces.census.gov/data/pseo_beta.html

Exhibit A2. Scatterplot of Scorecard median earnings and median earnings from Census PSEO earnings data

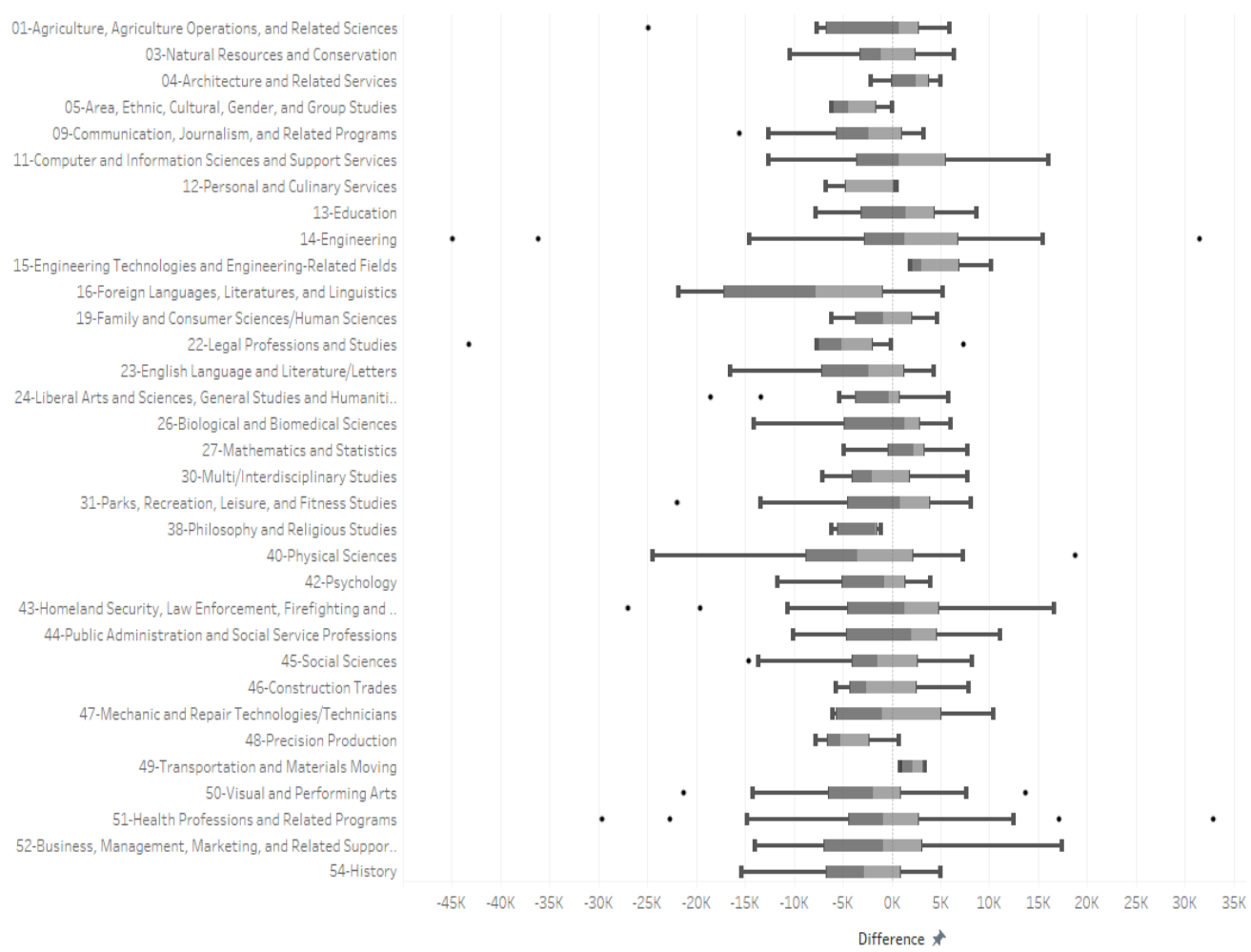


Correlation coefficient= 0.9

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 (November 2019).

Despite the positive correlation, there is variation by broader categories of field of study. In Exhibit A3, the distribution of differences between Scorecard and PSEO data values are shown by 2-digit CIP codes. Overall, considering all field of study differences regardless of the 2-digit CIP code grouping, the median difference was -$640, with an interquartile range from -$4,977 to $2,733.

Exhibit A3. Differences in median earnings (Scorecard values minus PSEO values) by 2-digit CIP codes

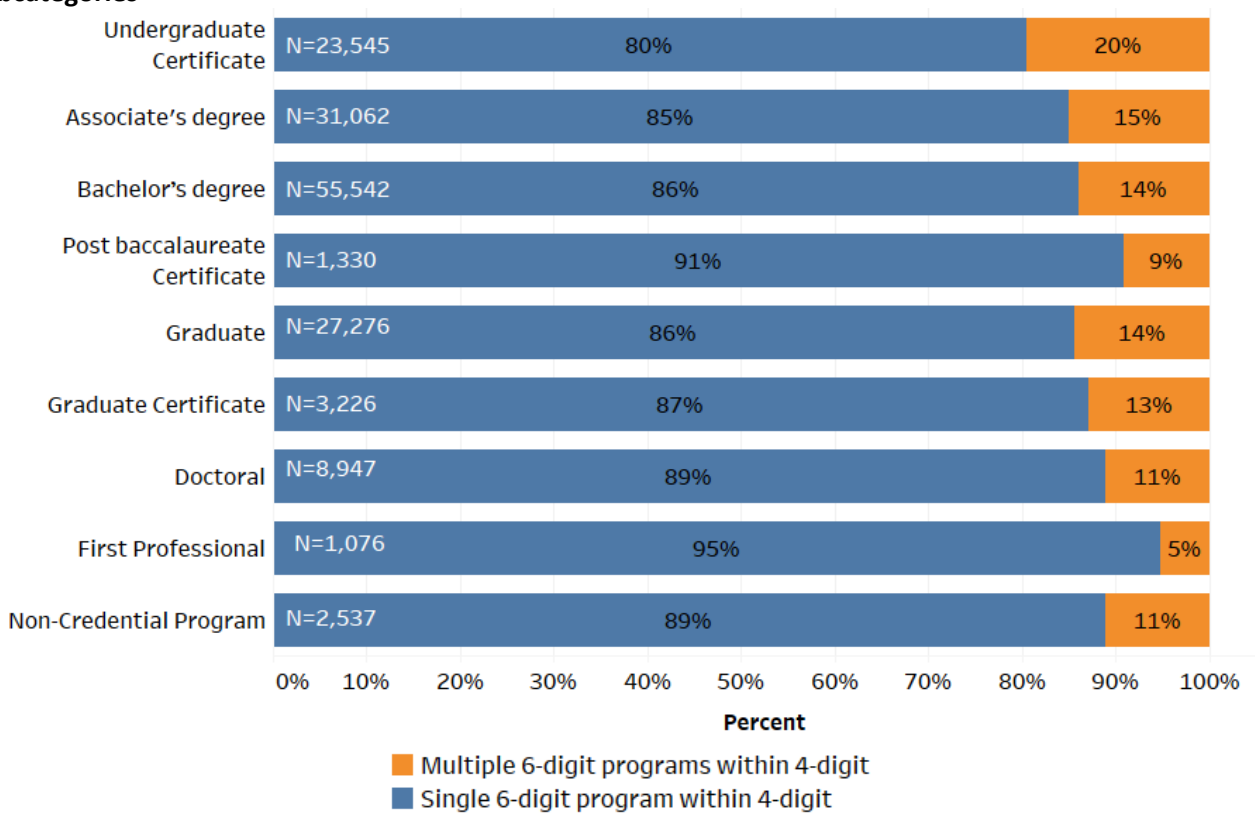


Source: U.S. Census Bureau. (2019). Post-Secondary Employment Outcomes Data (Beta) (2001-2016) [computer file]. Washington, DC: U.S. Census Bureau, Longitudinal-Employer Household Dynamics Program [distributor], accessed on 9/30/19 at https://lehd.ces.census.gov/data/pseo_beta.html R2019Q2 [version]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

Frequency of Six-digit CIP Codes Within Four-digit CIP Code-based Fields of Study

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 A4 shows the percentage of four-digit CIP code offerings with completers in one or multiple six-digit CIP categories by credential level.

Exhibit A4. 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.

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% 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 A5 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 A5. Percentage of institution-level completion records with corresponding program-level completion records as of mid-July 2019

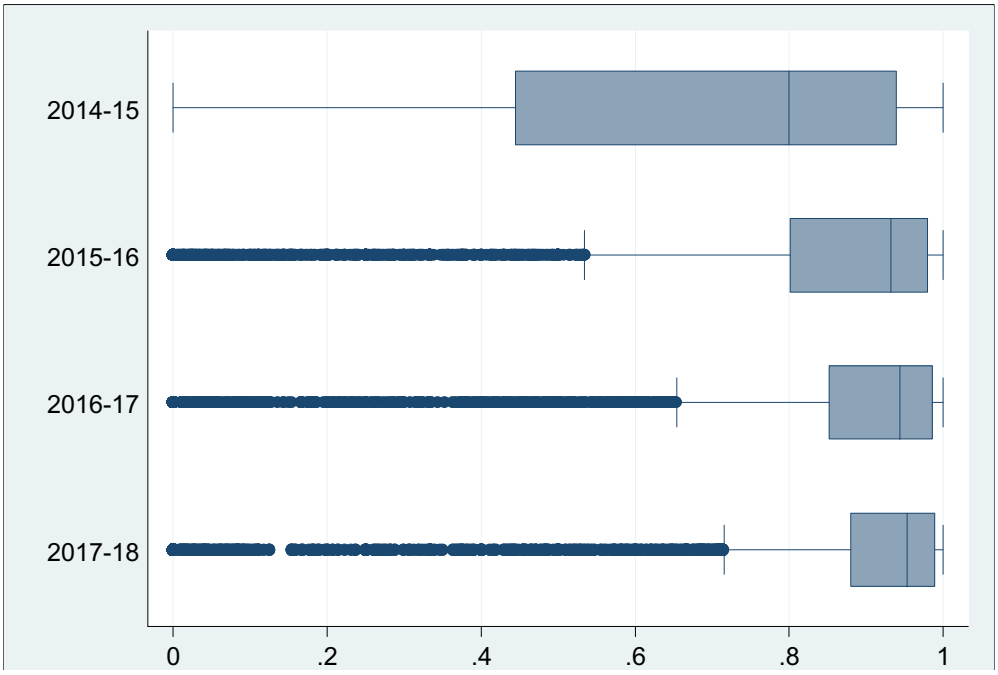
Award year	Percent
2014-15	73
2015-16	90
2016-17	91
2017-18	92

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).[14] 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 A6) 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 A6. Distribution of ratios for institution-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

[14] Analysis not shown in this document.

Longitudinal Study (BPS:12/17) contains a nationally representative sample of 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, sex, 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 A7 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 1 month of each other with the same 2-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 A7 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 A7 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.

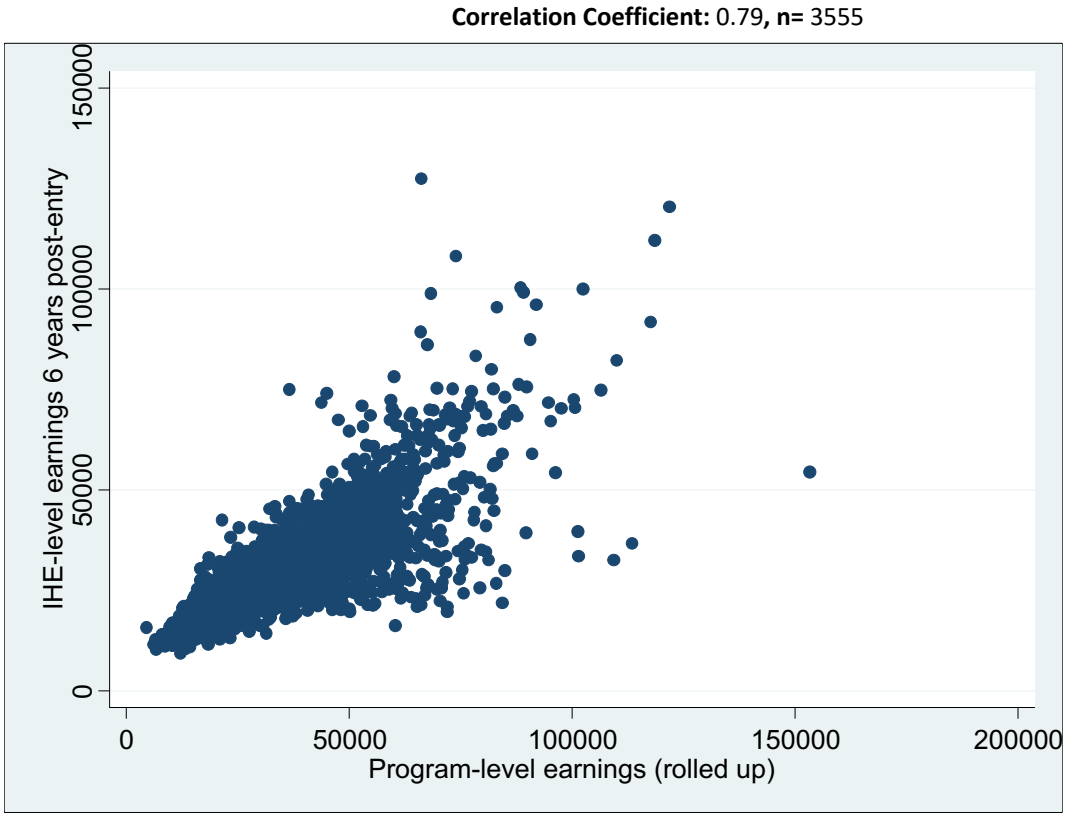
Exhibit A7. BPS:12/17 program-level completion match status of Title IV–aided students, by academic year, and credential level

			Matched		Unmatched		
	Academic Year	Total Count	Count	Percent Matched	Count	Percent of non-matched in NSC, but not in NSLDS	Percent of non-matched 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 agreement between these two data points, Exhibit A8 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 6 years post-entry. Exhibit A8 shows that the earnings measurements from the two sources are correlated.

Exhibit A8. Scatterplot of Scorecard field of study–level and IHE-level earnings values



Gainful Employment (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 A9 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 A9. Scatterplot of gainful employment comparison values and IRS earnings

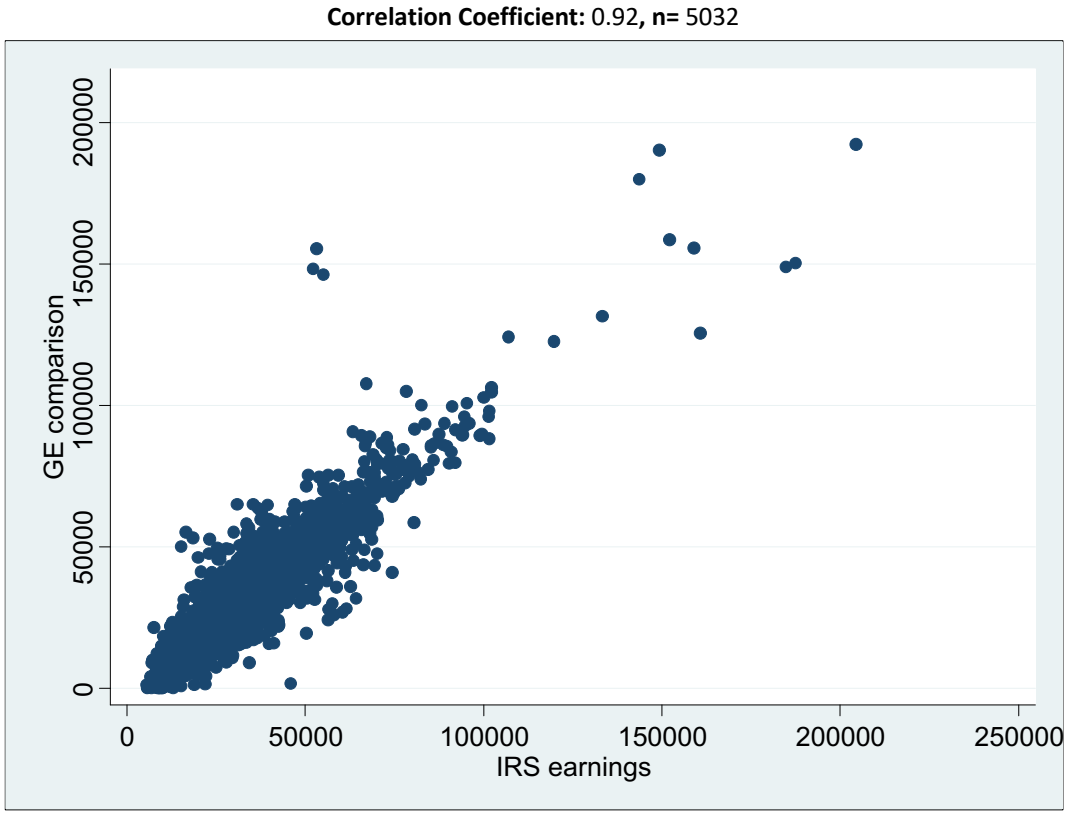


Exhibit A10 examines the correlation coefficients at specific credential levels and shows a high level of agreement within each credential level category, except for first professional degrees. This could be explained by the low number of observations that could be compared in both data sets (15) or 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 A10. 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 (November 2019)

Appendix B – Summary of documentation changes in this version

- Updates were made to page 4 to reflect the new national-level earnings measurements available.
- The table on page 7 was updated to display statistics associated with the most recent perturbation process for the 4-year timepoint.
- A table was added to page 8 to show statistics associated with the perturbation process for the 4-year national-level median.