College Scorecard Data Errata
Graduate Students Inadvertently Included in Undergraduate Earnings
October 15, 2026

The Department of Education (ED) became aware of issues pertaining to previously published data on institutional earnings outcomes in the College Scorecard. This document outlines the issues and provides an analysis indicating that they had little impact on the results.

ED has published undergraduate, institution-level, earnings outcomes data for many years. Table 1 describes the set of data elements and years in which these data were published.

Table 1. Institution-level, Post-completion, Undergraduate, Earnings Data Elements

Data Elements	Measurement Point	Affected Data Files
• Counts of individuals working and non-working. • Counts of working individuals earning greater than a threshold.	1-year post-completion	merged_2018_19_PP.csv, merged_2019_20_PP.csv, merged_2020_21_PP.csv
• Counts of individuals working and non-working. • Counts of working individuals earning greater than a threshold.	3-year post-completion	merged_2017_18_PP.csv, merged_2018_19_PP.csv,
• Counts of individuals working and non-working. • Median earning of working individuals	4-year post-completion	merged_2019_20_PP.csv, merged_2022_23_PP.csv
• Counts of working individuals earning greater than a threshold.	4-year post-completion	merged_2019_20_PP.csv
• Counts of individuals working and non-working. • Median earning of working individuals. • Counts of working individuals earning greater than a threshold.	5-year post-completion	merged_2020_21_PP.csv

In March 2026, the Department released 4-year, post-completion, institution-level calculations based on earnings in calendar year 2022 and 2023 for undergraduate students who completed in 2017-18 and 2018-19 respectively. These data were included in the back-end College Scorecard data files, and they were also presented in other formats, including the:

- College Scorecard consumer site.
- FAFSA "lower earning" summary page on the FAFSA application.

For these calculations, a code error inadvertently included some graduate-level students. The code error only impacts 4-year measurement calculations in the 2022-23 data file. Impacts of the code error include:

- At institutions that offer both graduate and undergraduate programs, some graduate-level completers were inadvertently included in undergraduate cohorts.

- Some graduate-only institutions have non-NULL values for undergraduate earnings metrics.

Institutions that do not offer graduate programs were unaffected by the code error. Further, all program-level calculations were unaffected.

Among institutions that offer both undergraduate and graduate programs, the median cohort included approximately 9 percent graduate-level completers. Table 2 summarizes the distribution of graduate completers (as a percentage of the overall cohort) in each intended undergraduate cohort.

Table 2. Distribution of percent of graduate students inadvertently placed into earnings cohorts that offer both undergraduate and graduate programs.	
Institution Count	1,553
Mean	13.3%
Std. dev.	16.1%
Percentiles	
1%	0.0%
5%	0.0%
10%	0.4%
25%	3.9%
50%	8.9%
75%	16.3%
90%	28.0%
95%	44.3%
99%	95.8%

While the other cohort calculations for institution-level, undergraduate earnings are not impacted by the code error, these cohorts may also include graduate-level students because of limitations in institution-level enrollment records used in the calculations. Graduate-level students are identified and excluded if they received title-IV benefits at the graduate level, but if they previously received title-IV benefits at the undergraduate-level they may be inadvertently included in the undergraduate-level calculations at the institution where they were a non-title-IV benefit recipient as graduate-level student.

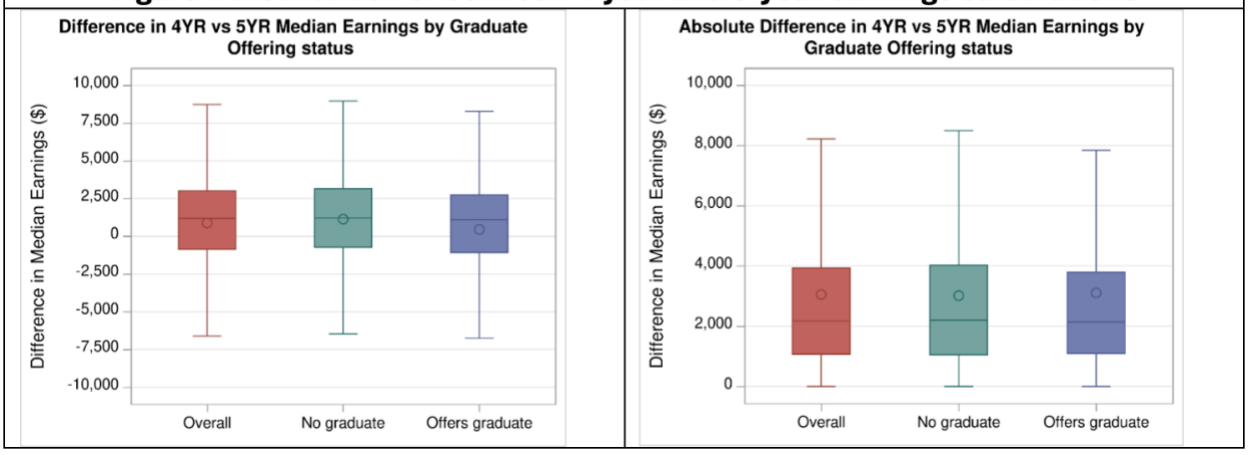
To resolve limitations with institution-level enrollment records, all College Scorecard new calculations in the 2022-23 data file (except for the aforementioned 4-year measurement calculations) use program-level completion records to identify students as graduate-level or undergraduate-level.

Analyzing Impacts of the Code Error

To measure the impact of the code error and the limitations to institutional records, the Department compared previously published 4-year earnings calculations in the 2022-23 data file to newly published 5-year calculations that leverage new improved methodology. Both calculations measure earnings in calendar years 2022 and 2023. However, the 4-year calculation is based on the 2017-18 and 2018-19 award year cohort and the 5-year calculation is based on the 2016-17 and 2018-19 award year cohort.

Changes from 4-year to 5-year median earnings are distributed very similarly across institutions offering graduate programs — which could be affected by the inclusion error — and institutions not offering graduate programs, which could not. The median change from 4-year to 5-year was 1,103 among affected institutions and 1,212 among unaffected institutions, a difference of 109, or approximately 9% of the observed change and only 2.5% of the standard deviation across all institutions (4,433). The magnitude of the difference (2.5%) falling below the conventional 10% threshold indicates there is no meaningful difference in central tendencies between the two groups. The IQRs of affected and unaffected institutions differ by only 82, indicating comparable variability between the two groups. Additionally, absolute differences are less than 8% of the standard deviation at every percentile which indicates that the error did not cause an unusual shift anywhere in the distribution.

Figure 1 Distribution of between 4-year and 5-year earnings calculations

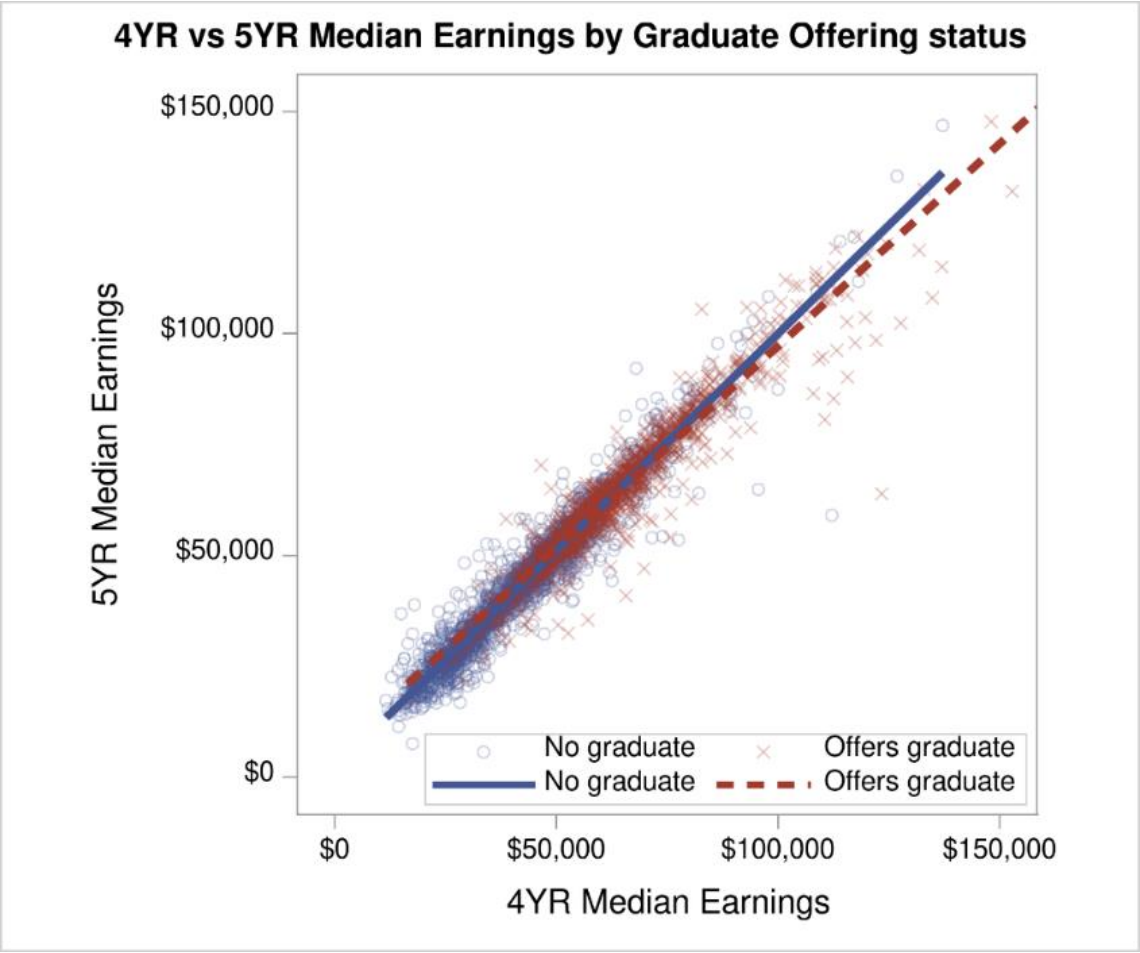


Looking more closely at the medians of the two groups, the small difference (-109) is in the direction that would be expected if higher-earning graduate students were included in the 4-year measures. Assuming the year-to-year increase is similar across groups, the inclusion error made the change from inflated 4-year earnings to corrected 5-year earnings in the affected group slightly smaller than the change from uninflated 4-year to corrected 5-year earnings in the unaffected group.

Finally, the visual similarity of the regression lines indicates that the difference in 4-year and 5-year earnings is largely unchanged between the two groups. In conclusion, there is

no evidence in the distributions or regression lines (see Figure 2) suggesting that the inclusion error made a meaningful difference.

Figure 2



Appendix I:

Descriptive Statistics for Differences between 4-year and 5-year measurement

Signed differences (5YR – 4YR)

	n	mean	std	p5	p10	p25	p50	p75	p90	p95	IQR
All	3,929	881	4,433	-5,587	-3,505	-853	1,187	3,009	5,075	6,785	3,862
No grad	2,440	1,137	4,142	-5,108	-3,363	-735	1,212	3,162	5,402	7,318	3,897
Offers grad	1,489	463	4,845	-6,745	-3,795	-1,076	1,103	2,739	4,619	5,912	3,815
Difference		-674	703	-1,638	-432	-341	-109	-423	-783	-1,406	-82
% SD				-37%	-10%	-8%	-2%	-10%	-18%	-32%	

Absolute differences

	n	mean	std	p5	p10	p25	p50	p75	p90	p95	IQR
All	3,929	3,052	3,333	209	433	1,071	2,174	3,933	6,235	8,530	2,862
No grad	2,440	3,018	3,055	201	423	1,052	2,203	4,029	6,283	8,437	2,977
Offers grad	1,489	3,109	3,744	229	452	1,092	2,139	3,795	6,128	8,679	2,703
Difference		91	688	28	29	41	-64	-234	-155	242	-274
% SD				1%	1%	1%	-2%	-7%	-5%	7%	