Roundtable on College Scorecard
CAIR Conference

November 17, 2016
8:00am – 8:45am
Crystal Ballroom

Facilitators: Ryan Chan, ryan.chan@ucop.edu; Chris Furgiuele, chris.furgiuele@ucop.edu

Agenda

1. Quick poll of participants' level of knowledge and experience with College Scorecard

2. Overview of Scorecard
    a. Brief history of development
    b. Participating agencies
    c. Audiences and products to reach them
    d. Scorecard.gov website vs. data for researchers

3. UCOP experience using Scorecard data
    a. Quick walkthrough of published analytical products
    b. Usage for media, communications, and executive talking points
    c. Validity check for internal quarterly wage data
    d. Contemplation of release of campus level earnings

4. Practical tips on how to use the Scorecard data
    a. Getting started – which datasets to use
    b. Optimal tools
    c. Comparative analysis basics
    d. Key variables of interest
    e. Weighting variables

5. Thoughts from the group
    a. Experiences, successes, challenges
    b. Questions

Background Materials

Overview

The College Scorecard was created to increase accountability and transparency in higher education, and to put more information into the hands of those choosing a college and those working to improve college quality. In doing so, the Scorecard provides data for benchmarking in three key college outcomes: graduation rates, indebtedness and default, and earnings.

Key takeaway: "While there is variation in the amount of debt and fraction of students borrowing by sector, on average, students at private for-profit two-year and four-year institutions have high rates of borrowing and their graduates often have large amounts of debt. While debt per se may not be problematic where students are able to repay their loans, it should be paired with other data, such as completion rates and post-school earnings, to provide a more comprehensive picture of student outcomes."

Public portal: https://collegescorecard.ed.gov, focuses on key metrics and general institutional information using the most recent data. Underlying dataset: https://collegescorecard.ed.gov/data/ Part of the data.gov ecosystem—the same API calls can be used to query and build applications/websites.

Data:

Sourced from three primary datasets:

- IPEDS Data
- National Student Loan Data System (NSLDS)
- Treasury data (NSLDS combined with tax records and census data)

A few variables come from other sources (Office of Postsecondary Education for HSI, AANAPII, PBI, and tribal institution flags; Department of Education for HBCU; and Federal Student Aid data center for default rates).

Key advantages of the scorecard dataset:

- Comparative NSLDS and Treasury data not available anywhere else (e.g. debt at graduation)
- Multiple datasets matched up for ease of use – a one stop shop for higher education data.
- Name-brand recognition

Major limitations:

- Most metrics are only for federal aid recipients
- No earnings disaggregation by major/program (though promised in future versions)
- Earnings data includes both graduates and non-completers, but does not distinguish between them.
- Scorecard data pulled from IPEDS may not be as current as what's available in IPEDS survey datasets

How to get started:

- Options: the "full" dataset, the "most recent data", the "scorecard" data, "earnings" (just the Treasury data), or "NSLDS" (just the loan data).
- We recommend using the "full" dataset or the "most recent data" dataset. The others are just filtered views.
    - If you use the "full" dataset, each year is a separate file, and certain elements are not available for the most recent year. For example, NSLDS variables are mostly unavailable in the 2014-15 file; variables that "pool" years are not available in the 2012-13 or 2013-14 file. On the other hand, the full dataset allows the creation of historical trends and also allows "snapshots" where the comparable "cohorts", at least in Scorecard's opinion, are lined up.
    - Using the "most recent dataset" avoids any issues with trying to find which year has the most recent value for a variable, but mixes years together.
    - Crosswalks are included for institutions that merge/dissolve across years.
- We recommend using a statistical program (e.g., SAS, Stata, SPSS) or a BI program (Tableau, Microsoft Power BI, QlikView) to analyze the data. Excel is possible but difficult/unwieldy. There are about 2,000 columns and nearly 8,000 institutions. Note that many institutions have "NULL" or "PrivacySuppressed" values for columns, which may cause numeric values to be recognized as character formats.
- Make sure to review the "cohort map" within the data dictionary to understand which metrics to which cohorts. The earnings, debt, and graduation rates all apply to different student cohorts

Basic analysis

- Create a comparison set of institutions.
    - UC uses "AAU Publics" and "AAU Privates" as a comparison group.
    - Use institutions that have comparable U.S. News Rankings (we don't endorse the rankings, but this is an easy way to find similar institutions)
    - IPEDS auto-generates comparison institutions if you don't pick them yourself. You can find them for any institution by looking for the "Data Feedback Reports".
    - Beware:
- Pick variables to use
    - Check to see that they are not missing/suppressed for the bulk of your comparison group
- Creating aggregate statistics
    - When creating "averages of averages", make sure to weight by the proper "N" (cohort size, enrollment, etc) indicated in the cohort map within the data dictionary
    - One possible approach for medians: sort them and assign the number of observations to each median, then find the observation that falls in the middle. In the example below, the 23rd observation would be in the middle, which corresponds to 5000.

Reported Median	N
2000	10
3500	6
5000	25
6300	4
	Total=45

- o Warning: some institutions with branches are multiple campuses are reported separately in IPEDS, but their statistics are combined and duplicated for NSLDS and Treasury data—they have separate IPEDS ids, but the NSLDS/Treasury data are aggregated at the OPEID6 level. This can result in overweighting of their responses by counting the same students multiple times.

Variables of interest.

Just a suggestion, there are many more.

Label	Name	Weight
Three-year cohort default rate	CDR3	CDR3_DENOM
Three-year repayment rate for completers	COMPL_RPY_3YR_RT	COMPL_RPY_3YR_N
Three-year repayment rate for non-completers	NONCOM_RPY_3YR_RT	NONCOM_RPY_3YR_N
Percentage first-generation students	PAR_ED_PCT_1STGEN	
Average family income of dependent students in real 2015 dollars.	DEP_INC_AVG	DEP_INC_N
Average family income of independent students in real 2015 dollars.	IND_INC_AVG	IND_INC_N
Number of students not working and not enrolled 10 years after entry *(calculate unemployment rate)*	COUNT_NWNE_P10	**N/A**
Number of students working and not enrolled 10 years after entry *(calculate unemployment rate)*	COUNT_WNE_P10	**N/A**
Mean earnings of students working and not enrolled 10 years after entry	MN_EARN_WNE_P10	COUNT_WNE_P10
Median earnings of students working and not enrolled 10 years after entry	MD_EARN_WNE_P10	COUNT_WNE_P10
The median original amount of the loan principal upon entering repayment	DEBT_MDN	DEBT_N
The median debt for students who have completed	GRAD_DEBT_MDN	GRAD_DEBT_N
The median debt for students who have not completed	WDRAW_DEBT_MDN	WDRAW_DEBT_N

Data Release History:

First release: September 2015, Updated March 2016. Data update: September 2016.

https://collegescorecard.ed.gov/data/changelog/

More Reading and Resources

StackExchange forum (the scorecard team usually replies to queries)
http://opendata.stackexchange.com/questions/tagged/collegescorecard

Where the College Scorecard Has Gained Traction So Far — and Where It Hasn't. Chronicle of Higher Education, September 28, 2016. http://www.chronicle.com/article/Where-the-College-Scorecard/237919

'How Much Will I Make After Graduating?' College Scorecard Offers Only Clues. Chronicle of Higher Education, September 15, 2015. http://www.chronicle.com/article/How-Much-Will-I-Make-After/233111