

Employment Data Taskforce Report Out

Overview

The Office of Cradle-to-Career Data convened the Employment Data Taskforce for the purpose of identifying ways to consistently calculate employment and earnings data points from the unemployment insurance source data provided by the Employment Development Department.

The taskforce was made up of representatives from California Community Colleges Chancellor's Office (CCCCO), University of California (UC), Employment Development Department (EDD), Georgetown University Center for Education and Work, Columbia University Community College Research Center, California State University-Northridge, University of California-Berkeley, and University of California-Davis.

Over the course of three public meetings held during summer 2022, the group completed the following tasks:

- Conducted analyses on UI data to address questions about data variations and potential missing values
- Developed considerations for processing UI data
- Developed considerations for how to construct employment and earnings metrics
- Identified policy implications of how data are processed and metrics would be calculated

The report out will be shared with the Governing Board and taken into consideration by the Office of Cradle-to-Career (C2C) staff as they move forward with calculating employment and earnings metrics for the analytical data set that will be used for the dashboards and the query builder tool.

Findings from UI data set analysis

When analyzing data pulled for each of the last 26 years, with between 12.3 million - 21.1 million Unemployment Insurance (UI) wage records reviewed in each of the years, EDD found:

- The data set appears to be complete from reporting employers, in that there were no null records found in any fiscal quarter.
- A very small number of social security numbers (SSNs), which are used as the unique identifier for each person, appear to be used by multiple individuals in a fiscal quarter. 85-87% of SSNs are present for only one employer, 12-14% of SSNs had between two and nine employers, 0.2-0.3% of SSNs are present for ten or more employers (roughly 32,700-44,000 records), and 0.04-0.06% of SSNs are present for twenty or more employers (roughly 7,300-10,400 records).
- The number of cases where it appears that employers submitted duplicate records for the same individual diminished over time, with most years having duplications in 0.01-0.03% of records, with the highest figure being from 1995 (0.2% or 29,470 out of 15 million records).
- Most individuals have records in all four fiscal quarters of a year (about 70%). Generally, the distribution of individuals with records in one, two, or three quarters is about 10% each.
- Most individuals have an earnings record in each fiscal quarter of the year. 5-10% of individuals have a record in one fiscal quarter, but no record in the next fiscal quarter, and 1-3% have records in two consecutive fiscal quarters, but no record in the next fiscal quarter. People with a record in fiscal quarters 3 and 4 were slightly more likely to be missing records in subsequent quarters.
- There are significant earnings differentials between the lowest 95% of worker salaries, the top 1%, and the top 0.1%, which may distort averages. This gap has increased over time.
- Analyses conducted by CCCCCO and CSU showed that when data are disaggregated, there are significant differences in match rates, including by segment, race, gender, financial aid status, academic discipline, completion status, and over time.
- CCCCCO indicated that between 9-12% of students in career education disciplines report being self-employed, and thus would not be found in the UI data set.

Key points of discussion

Principles for Calculating Earnings

- Earnings metrics should seek to realistically reflect the true earnings of each individual, as opposed to what someone working full-time should expect to earn based on factors such as education attainment level or field of study.
- Minimum thresholds for earnings should not be set for data to be displayed, such as removing records that are below the federal minimum wage.

Employment

- C2C should not calculate an employment rate due to limitations in the underlying data set, and instead should indicate the number of students who were found in/matched to the state wage file. In data displays, this information could be displayed as a percentage of the relevant cohort.

Matching Individuals in the C2C Data Set with UI Wage Data

- All individuals for whom C2C has an SSN should be matched to the UI wage file, although some individuals may be excluded from specific cohorts or metrics.
- Each year, the match conducted by EDD should include all available years of data for new SSNs and the most recent year of data for all previously matched SSNs.
- Quarters in which there is no match should be counted as zero earnings in the match file, although some individuals with zero earnings may be excluded from specific cohorts or metrics.
- When returning the file to C2C, EDD should eliminate the 0.04-0.06% of records for individuals with 20 or more employers.
- When returning the file to C2C, EDD should eliminate the 0.01-0.03% of records for individuals that have exact duplicates.

Matching Individuals in the C2C Data Set with National Student Clearinghouse Data

- Secure information from the National Student Clearinghouse to:
 - identify people who still enrolled in postsecondary, to determine which individuals should be classified as exiters
 - determine whether students graduated from colleges in other states and their highest degree awarded

Cohorts

- In the query builder, users should be able to create cohorts based on three dimensions:
 - The number of fiscal quarters in which earnings were found (users may select 1, 2, 3, or 4 quarters within a fiscal year)
 - The highest level of postsecondary educational attainment (users may select no award, certificate, associate's degree, bachelor's degree, master's degree, doctor's degree, post-baccalaureate and post-master's certificate)
 - Whether the individual had exited all postsecondary education in the time period in which earnings are being calculated, meaning that they are no longer enrolled but may have left without earning a certificate or degree

Timeframes

- In the query builder, users should be able to identify which and how many years of earnings information to display, such as every year from 2015-2022, or only the second year after the individual exited postsecondary education

Earnings

- Earnings should be calculated by adding together earnings in the number of fiscal quarters selected for the cohort (users may select 1, 2, 3, or 4 quarters within a fiscal year).
- Earnings should be calculated as median earnings, with 25th percentile, 50th percentile, and 75th percentile provided.
- When showing results for exiters, use event time, meaning starting with the first full fiscal quarter after the student left postsecondary education.
- Earnings should be adjusted for inflation using the California Consumer Price Index for All Urban Consumers (CPI-U), using the most recent year of calculation as the base year

Earnings Gains

- Earnings gains should only be calculated for people with both pre- and post-records sufficient for calculating earnings. If the user indicates that the cohort includes individuals with earnings values in two fiscal quarters, then

earnings gains should only be calculated for individuals with earnings values greater than zero in two fiscal quarters both before and after participating in postsecondary education.

- Calculate pre-wages using the timeframe of the four fiscal quarters prior to entry to the postsecondary institution. Users may select cohorts that have earnings in one, two, three, or four of these quarters.
- People who were enrolled in either K-12 or postsecondary in the time period for pre-earnings should be excluded from the calculation.
- Allow users to select the timeframe after exit for the post-wage (for example, the four quarters immediately after exit or for each year of the ten years after exit).

Living Wage

- Living wage should be calculated in three ways, using the MIT Living Wage Calculation for California:
 - Living wage attainment for a single adult: Earnings will be compared to the variable for Living Wages for 1 adult with 0 children for the county in which the institution is located.
 - Living wage attainment for a family of four with two working adults: Earnings will be compared to the MIT Living Wage Calculation for California variable for Living Wages for Two Adults (Both Working) with 2 Children for the county in which the institution is located.
 - Living wage attainment for a family of four with one working adult: Earnings will be compared to the MIT Living Wage Calculation for California variable for Living Wages for Two Adults (1 Working) with 2 Children for the county in which the institution is located.
- Earnings should be calculated by adding together earnings in the number of fiscal quarters selected for the cohort (users may select 1, 2, 3, or 4 quarters within a fiscal year)
- Living wage attainment should be calculated for each individual in each year that they are matched to the UI wage file.
- An individual's geographic location should be the county in which the last post-secondary institution they attended is located.
- Documentation should address the fact that these figures do not include expenses such as loans, and prepare users for how low the figures are.

Testing To Ensure Data Quality

- Data testing should not be so exhaustive that it prevents information from being calculated. C2C should be aware that matches will never be exact, so comparison values should be used to look for whether results are reasonable.
- Possible points of comparison include:
 - Examine data to see if there are patterns one would expect during recessions.
 - Compare median earnings to values shown in the US Department of Education's College Scorecard.
 - Use the American Communities Survey to measure the share of people in California in a particular age/education range who have positive earnings in a year to evaluate match rates (note that ACS has multiple categories for "some college, no degree" including options that take less than the year and more than the year).
 - Compare earnings information to data available from the U.S. Census Bureau Post-Secondary Employment Outcomes.
 - Compare results by region.
 - Compare values to match rates and earnings figures calculated by each of the segments, taking into account methodological differences. This includes:
 - comparing salary figures
 - comparing match rates based on self-employed students
 - comparing results for specific populations and fields of study, such as female Biology bachelor's degree earners, taking into account industry of employment

Policy and implementation considerations

- It may be helpful to provide explicit contextual information for people accessing the query builder to understand variations in data available through the C2C and information posted by other state agencies. The prevalence of missing records will be more pronounced for specific education segments and student populations. For example, there may be a lower match rate for community college students, people who major in some disciplines, or people who live in regions close to borders with other states. As a result, values may be more variable or be suppressed when users disaggregate information.

- Any changes to formulas should be carefully considered once the data have been made public and clearly documented, to build confidence in the information being provided.

Items for future consideration

- The metrics will not fully clarify who is employed in the California labor market and what they are earning because it is impossible to distinguish between someone who has lost a job and is seeking to work, someone who has stepped out of the workforce, someone who is earning money in California that is not reported to EDD, or someone who has left the state. C2C should conduct additional research, once the data set is constructed, to identify individuals who are no longer in the California labor market.
 - It would be beneficial to augment the UI data set with other data sources, such as Franchise Tax Board records, to gain a better understanding of employment status and earnings for individuals not in the UI wage file.
 - C2C should explore using unemployment insurance claims to better understand whether an individual was unemployed in a specific quarter (although this measure will not include people such as those who exhausted UI benefits, did not file for benefits, or left the state).
- The taskforce raised a number of important considerations about the best way to calculate earnings for the dashboards, where only one cohort can be used, such as the need for further analysis to determine how to handle factors such as departure from the California labor market, exit status, highest educational attainment, and the number of quarters of earnings.
- Given that the query builder will allow the public to construct earnings metrics that will vary from information posted by other state agencies, and that members of the public may not understand how to annualize earnings from incomplete data sets (such as cases where only one or two fiscal quarters of earnings data are available), C2C should consider what information should be freely available to the public.
- C2C should identify additional measures of economic mobility, such as leveraging information on parental earnings from financial aid data sets as a point of comparison for students' earnings.