Revised Summary Report

U.S. Department of Labor Research Roundtable on Assessing Workforce Skill and Competency Change Over Time

December 9, 2022

Submitted to:

Monica Mean
U.S. Department of Labor
Chief Evaluation Office
200 Constitution Ave, NW
Washington, DC 20210

Submitted by:

Marilia Mochel
Manhattan Strategy Group
4340 East-West Hwy Suite 1100
Bethesda, MD 20814

Table of Contents

RESEARCH ROUNDTABLE ON ASSESSING WORKFORCE SKILL AND COMPETENCY CHANGE OVER TIME

Introduction

The U.S. Department of Labor (DOL) Chief Evaluation Office (CEO), in collaboration with the Employment and Training Administration (ETA), sponsored a roundtable of subject matter experts[1] to learn how researchers assess workforce skill and competency change over time, to investigate new data sources to measure this change, and to collect recommendations for possible public investments to support new data sources. The roundtable was hosted virtually on October 21, 2022, from 3:00 to 4:30 Eastern Time. CEO and ETA disseminated invitations for a selected guest list composed mostly of federal staff and members of the Workforce Information Advisory Council, which includes state agency leaders. Forty-six participants attended.

This report summarizes the presentations delivered by the experts at the roundtable. First, we provide a background and motivation for the event's organization. We then introduce the panelists who participated in the event, describing their relevant credentials. The remainder of the report describes the key elements of the panelists' presentations, followed by their recommendations for the federal government relating to its role in data collection on skills and competency change.

Background

The U.S. labor market is changing. In upcoming years, new jobs will be added, existing jobs will be lost, and the nature of work in the remaining jobs will change due to factors such as (1) the increase in remote work and other enduring impacts of the COVID-19 pandemic, (2) the role of gig work, and (3) future adoption of automation and the impacts of artificial intelligence (AI) on work.

Obtaining timely intelligence on these labor market impacts could help inform individuals, education and training providers, and the workforce system by identifying which workers might need reskilling for different pathways—flagging which tasks, skills, and competencies might become obsolete and what new skills workers will need to acquire for various career paths.

It is valuable for ETA to learn about feasible and optimal ways to assess skill and competency change over time, as various technological, social, and environmental factors impact jobs. It is no longer human curatable to capture skill and competency levels solely through surveys and focus groups; real-time data from online and administrative databases also need to be leveraged. The four overarching goals for the research roundtable were as follows:

1. Learn from researchers in the field various perspectives on the feasibility of analyzing skill and competency change over time.
2. Explore needed data elements to assess skill and competency changes for occupations.
3. Explore various methods or approaches, such as AI and machine learning (ML), to assess skills and competency demand and trends over time, as certain occupations increase or decrease in size or as task composition of occupations changes.

[1] Manhattan Strategy Group (MSG), a contractor to the CEO, supported the planning and staging of this roundtable funded by contract # AG-3198-S-16-0048.

4. Identify options to support such analytical capabilities either through government or public-private partnerships to build capacity for capturing and assessing such information.

The Roundtable Panelists

CEO and ETA collaborated to select panelists from a list of experts MSG identified for their research experience in examining changes in occupational skills in the U.S. labor market. MSG recruited the panelists to participate in the roundtable. The panel included four panelists whose expertise we summarize below.

Dr. Daniel Rock is an Assistant Professor of Operations, Information, and Decisions at the Wharton School of the University of Pennsylvania. His research focuses on the economic effects of digital technologies, with a particular emphasis on the economics of AI. He has recently worked on studies addressing the types of occupations that are most exposed to ML, measuring the value of AI skillsets to employer firms, and adjusting productivity measurement to include investments in intangible assets.

Dr. Ben Zweig is the CEO of Revelio Labs, a market intelligence company that derives competitive insights based on unstructured human capital data. Revelio Labs leverages the latest advances in AI research methods to create structured and accurate representations of raw labor data contained in millions of resumes, online profiles, and job postings. Dr. Zweig is also an Adjunct Professor at New York University's Stern School of Business, where he teaches a class about the future of work.

Dr. Ina Ganguli is an Associate Professor of Economics at the University of Massachusetts Amherst and Associate Director of the UMass Computational Social Science Institute. Her research areas are labor economics, the economics of science and innovation, international development, and economic history. She works on topics related to immigration, gender issues, and how individuals acquire and use their skills, particularly in science and innovation. Much of her work is focused on the behavior of high-skilled "knowledge" workers: scientists and engineers.

Dr. Robert Seamans is an Associate Professor at New York University's Stern School of Business. His research focuses on how firms use technology in their strategic interactions with each other, as well as the economic consequences of AI, robotics, and other advanced technologies. Dr. Seamans and collaborators developed the AI Occupational Exposure measure, which uses data on AI progress and O*NET data to assess the potential impact of AI on occupations. Dr. Seamans served as the Senior Economist for technology and innovation on President Obama's Council of Economic Advisers.

Summary of Panel Presentations

The roundtable organizers instructed panelists to be prepared to speak for up to 15 minutes. Their remarks should engage the following areas: their use of occupational/skills datasets in research, new and potential data sources and methods to assess changes in skills over time, strengths and weaknesses of new approaches, and recommendations for the role of the federal government and public-private partnerships. We discuss below the main themes raised by the panelists in their presentations.

Theme #1: ML use of digital data sources to understand labor market dynamics is growing

Dr. Rock was the first panelist to speak. His remarks emphasized the huge opportunity posed by using ML on digital data sources to measure movements in the labor market. He noted that while economists have worked with patents in this manner, using ML to assess job postings is more recent. He equated online job postings to digital breadcrumbs available for the government and private sectors to use to learn about trends in the labor market.

He shared ways in which researchers have been using these data to better understand labor market trends, showcasing his own research. One such example, which has received interest from researchers and government alike, refers to ways to measure demand for AI skills in the labor market. Dr. Rock showed how his research used private sector data to track changes over time in AI demand. Using digital data sources enabled him to explore various facets of demand for AI skills within jobs and also at a deeper level, showing how demand is concentrated in specific firms.

He noted that ML is highly adaptable and allows researchers to explore changes in the labor market quickly. He described analysis illustrating how digital data sources can be used to show how occupations are standardizing over time, moving from more varied activity profiles to more similar ones. He showed how the labor market itself is expanding with job groups spreading out. He expanded on the potential of ML to make sense of vast troves of digital data (job postings) by describing how researchers can use ML to track the expansion of the use of remote work. He showed how such techniques enable researchers to assess how specific aspects of jobs are changing over time and in close to real-time.

"You can automate classifying certain jobs if you have a job description. Maybe what we want to do, for example, is index work based on how remotable it is. So, we clip out some of the job posting […] and add in 'this is a full-time remote position and employees can be based anywhere in the United States'. We put that in the posting at random somewhere. Then we classify both the original posting and this one and look at how much our predictions changed once we have this remote work text in there and we can see, oh, okay, well, the ones that we shift toward, they're like first-line supervisors of office and administrative support workers, human resource managers, and so on. They're more likely to be amenable to remote work."

Theme #2: Digital data sources and ML can be used to create a common taxonomy of occupations

Dr. Zweig emphasized how advanced techniques enable analysis of the occupational landscape in a way that is faster and less costly. He noted that survey-based data sources can become prohibitively expensive to measure as the economy becomes more complex. ML, and specifically sophisticated natural language processing techniques, provide an alternative.

Dr. Zweig started by describing jobs as bundles of activities or tasks. Therefore, he noted that to "train" a machine to classify jobs or occupations requires available data on these activities or tasks. In his work, the data comes from online profiles, including LinkedIn resumes, and job postings. Resumes include an individual's position, job title, start/end dates for positions, and bullets describing what an individual has done at the job, which constitutes a list of activities involved in that job. Similarly, job postings also list responsibilities, qualifications, skills, and a little about the company.

Analysis of job postings enables researchers to combine jobs that are similar given the description of responsibilities. ML analysis provides representation for every job and allows us to combine jobs that are similar. The potential for use of ML includes the ability to cluster jobs into occupations, to find similarities across groups, and to develop skill groups or families. Ultimately, ML enables us to create a taxonomy of activities and develop a common language where occupations, skills, and activities can be cross-referenced. Indeed, other elements can also be analyzed, including geography or seniority.

Dr. Zweig envisions that the development of these data sources could lead to developing a common language to provide on-demand applications for any end users. It would facilitate the work of firms, as they would not need to "reinvent the wheel" in creating their own internal taxonomies. It would make labor market data more ubiquitous, facilitating the allocation of labor in the same way financial data enabled better allocation of capital. Above all, the development would not need any human curation or judgment calls, automatically constructing taxonomies that adapt to a changing economy.

He emphasized that online profiles are the richest data source, the most foundational to understand the full dynamics of the workforce. They enable one to get headcounts by occupation, skill, geography, and seniority in addition to job activities. What is more, with this individual-level data, you can get an approximation of demographics (ethnicity and gender) via first name, last name, and geography. In his view, the analysis can get to the very local level.

Theme #3: ML has limitations due to sampling bias of digital data sources

Dr. Rock recognized a key weakness of digital data sources. He noted they have limited coverage on the lower end of the wage distribution and blue-collar work. Using his analysis of IT occupations, he showed how LinkedIn coverage rates vary by occupation. As an example, he showed his research documenting how there is one posting or resume on LinkedIn for every software engineer job (as seen on Burning Glass data on job postings) while at the same time there are not as many LinkedIn profiles in construction. Although adoption is getting better, he noted that this requires researchers to calibrate digital data. He recommended combining digital data sources with survey data to provide a more accurate picture—it allows the researcher to adjust the dimension of the digital breadcrumbs to the prevalence identified in the surveys, inflating or deflating as needed.

Dr. Zweig agreed that there is sampling bias in these digital datasets. This means researchers must adjust for the likelihood of individuals/job postings being represented in the sample. This can be accomplished by developing sampling weights that rely on other data sources. Additionally, he noted another challenge of such data sources: profiles typically encounter lags for when people report their job transitions. This may also bias the analysis, and the data need to be adjusted for it. At this time, he noted, this is a hard problem to solve.

*Theme #4: O*NET can be used to examine skills change, but creating a panel is an arduous process*

Dr. Ganguli presented how her research used O*NET data to create a panel to assess occupational changes. She, along with her collaborators, looked at the O*NET incumbent worker profiles from 2003 to 2008 and again from 2013 to 2018 to create measures for 2 years, 2005 and 2015, with 371 (of more than 800 in O*NET) occupations in both periods of time. The research team analyzed changes in these occupations in the two periods to show that within-occupation changes dominate aggregate changes in this period.

She also presented research underway that uses O*NET to measure work from home. They linked 2004 and recent O*NET measures to individual-level data in the 2004 and 2020–2021 Current Population Survey (CPS), which include individual demographic information and questions about whether someone worked from home. The research shows that O*NET 2004 data on work from home predicts change in work from home from 2004–2020 as seen in the survey data.

As she demonstrated how O*NET can be used to examine change over time, Dr. Ganguli noted that handling O*NET data is an arduous process, especially when the goal is to create a panel. Collaboration with DOL and O*NET experts is often necessary to build the panel. This limits widespread use in this manner.

Theme #5: There is now a variety of data sources and new metrics for technology's impacts on jobs

Dr. Seamans described the variety of data sources he uses and their applications in his research. In his first example, he explored how AI applications would affect different occupations at the industry and geography levels. With his co-authors, Dr. Seamans developed an AI Occupational Exposure index using categories of development in AI produced by the Electronic Frontier Foundation. They mapped these advances to the 800 occupations from O*NET to create the index and examine where the AI Occupational Exposure measure would have an effect. This measure can be applied to geographical areas to assess their exposure based on survey prevalence of the occupations.

For another project, Dr. Seamans was interested in AI skills demand across occupations. For this effort, he used Burning Glass data on job postings to measure whether AI skills were required for a job, aggregated at the occupation level. He also used Burning Glass data to see if the opportunity zones initiative had generated job growth in the geographical areas. He compared job-posting levels in opportunity zones with other areas to see that the job growth had not been different. Finally, he also shared his more recent project focused on AI prevalence that uses data to assess job descriptions at the firm level.

Panelist Concluding Observations

At the end of their presentations, panelists were encouraged to provide recommendations to and about the federal government and its role in data collection related to skills and competency changes. We describe their recommendations below.

Automate the ability index to build an observatory of skills

Dr. Rock recommended the development of an ability index using automation and digital sources. This could be done via ML analysis of job postings, with their detailed skills descriptions, in combination with high-quality sampling from government sources, to build an observatory of skills and labor that has not existed in the past. Such data would be useful to learn about the big picture, allowing us to identify where the displaced workers are and where people may need help. His work shows that the concern is not that work is going away. In fact, his data show that every year the sample gets more diverse in terms of the types of work people are doing and that more recombination of activities is happening. Instead, there are areas that are vulnerable to the impact of automation on distributional changes and not to replacement due to automation.

Encourage flexibility in how we categorize jobs

Along these lines, Dr. Zweig echoed that jobs do not get automated; *components* of jobs do. For this reason, he favors providing flexibility in defining occupations, allowing definitions to be adaptable. He encourages flexibility in categorizing jobs and putting information out for the general public.

*Facilitate the creation of panels using O*NET to provide an "off-the-shelf" dataset*

Dr. Ganguli discussed how time consuming it is to crosswalk O*NET occupations to create a panel. Her team has an expert on O*NET that helped, but there were many challenges. Therefore, she senses a lot of desire for more off-the-shelf measures that researchers can easily download. ETA's Pam Frugoli noted that a team from the University of Michigan and Florida State University is unveiling a crosswalk tool to make this process more accessible.[2]

Encourage partnerships between statistical agencies and external researchers

Dr. Seamans encouraged collaboration and partnerships between federal statistical agencies and external researchers, in other parts of the government, in think tanks or academic institutions. He considers these partnerships fruitful and hopes the agencies will continue and expand these opportunities.

Create new modules describing work activities that cover coordination elements of an individual's job

Dr. Rock had a recommendation related to survey-based data collection produced by government agencies. He noted that it would be useful to have survey modules that address how people work together at the job. He argued that to understand how work is changing, it is important to know how workers coordinate with other different types of workers. This is a gap he identifies in current data collection.

Consider frequency of measurement of occupations using online data sources

Dr. Rock suggested that the frequency of data collection on occupations could be adjusted to facilitate longitudinal analyses. He noted that online data sources could allow more frequent measurements for different roles with more variation cross-sectionally. This would help us understand what is happening in labor markets.

*Use O*NET to capture adjacency of occupations*

While noting this was not a recommendation per se, Dr. Seamans suggested that there is value in using O*NET in ways that are policy relevant. For instance, he suggested that O*NET (or another government dataset) could identify adjacent occupations or sets of skills.[3] This could help individuals find other good occupations when there is a retraction in their own occupation in a geographical area by showing other occupations that can use individuals with these skills. Dr. Rock indicated that he has used digital data sources to build such adjacency measures for a large employer. The employer found the measures useful in trying to target adjacent roles when trying to fill roles or increase the supply of adjacent workers.

Enumerating job activities comes before skills to enable a move away from hiring based on job titles.

As noted, Dr. Zweig described jobs as bundles of activities. As such, he indicated that moving towards skills-based hiring misses a step. Employers are seeking to hire people to get the work done and to ascertain what skills are needed, they need first to enumerate the activities the job

[2] For more information, see https://claudepeppercenter.fsu.edu/onet/ and https://hrs.isr.umich.edu/news/data-announcements/health-and-retirement-study-linkage-occupational-information-network-data.

[3] See "Developing Related Occupations for the O*Net Program" for some of the work underway in this realm: https://www.onetcenter.org/reports/Related_2022.html

entails. Once that is produced, one can add a layer of translating activities into skills. This last also enables one to identify the people who have the skills.

Next Steps

As U.S. labor market trends continue to change and evolve, it becomes pertinent to accurately measure in-demand knowledge, skills, and competencies continuously. This roundtable presented best practices thus far and provided some key considerations on how various data sources can be improved for the purposes of assessing the demand for skills and competencies more quickly and frequently. Through the annual learning agenda process, the U.S. Department of Labor continues to gather additional information and explore options to ensure its systems and data can best support this important work.