

Understanding Labor Market Demand in Real Time in Argentina and Uruguay

Evelyn Vezza
Gonzalo Zunino
Luis Laguinde
Harry Moroz
Ignacio Apella
Marla Spivack



WORLD BANK GROUP

Social Protection and Labor Global Department

March 2025

Abstract

This paper explores how job vacancy data can enhance labor market information systems (LMISs) in Argentina and Uruguay where, as in many countries, data on in-demand skills is lacking. By analyzing job postings collected over four years in Argentina and Uruguay, this study assesses the potential of vacancy data to fill labor market data gaps. The findings reveal that vacancy data capture labor market dynamics across time and geography, showing a strong correlation with traditional labor market indicators such as employment and unemployment. However, the data are biased towards higher-skilled occupations. Despite these limitations, the large volume of postings allows for robust inferences and provides valuable insights into skills demand. The study presents three key applications of the data: 1) using postings as a leading indicator of labor market

health; 2) identifying in-demand skills; and 3) mapping similarities between occupations to improve the information available to job counselors to provide advice about job transitions. Finally, the paper contributes methodologically by developing both a manually created skills taxonomy and an experimental machine learning approach to classifying skills. The machine learning method, while less comprehensive, highlights in-demand skills and can complement the manual approach by keeping it up to date with minimal input. Overall, the paper demonstrates the potential of job vacancy data to improve LMISs and inform labor market policies in Argentina and Uruguay with immediate practical applications for labor market analysis, skills development, and workforce training.

This paper is a product of the Social Protection and Labor Global Department. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/prwp>. The authors may be contacted at hmoroz@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Understanding Labor Market Demand in Real Time in Argentina and Uruguay¹

Evelyn Vezza, Gonzalo Zunino, Luis Laguinge, Harry Moroz, Ignacio Apella, and Marla Spivack

JEL code: J21, J23, J24, J63, O15

Keywords: Argentina, employment, labor market information systems, machine learning skills, Uruguay

¹ This paper benefited from inputs from a team from Charles River Economics Labs at the University of Chicago that consisted of Utkarsh Dandanayak, Kiran Duggirala, Ishaan Goel, Giyoung Kwon, Katherine Papen, and Prakhar Saxena. The authors are grateful to the *Secretaría de Trabajo, Empleo y Seguridad Social* in Argentina and the *Ministerio de Trabajo y Seguridad Social* in Uruguay. The team received very useful comments, advice, and assistance from Sofia Belen De Benito, Maria Eugenia Bonilla-Chacin, Alejo Uriel Burgos, Carolina Crerar, Maddalena Honorati, Victoria Levin, Davor Miskulin, Mauro Pelucchi, Marcela Salvador, Aivin Vicquierra Solatorio, and William Wiseman.

Introduction

The nature of work is changing in Argentina and Uruguay. First, Argentina and Uruguay are experiencing *deindustrialization* as jobs move towards the services sector. Employment in industry declined from around one-third and one-quarter of jobs in 1991 in Argentina and Uruguay, respectively, to just over and just under one-fifth in 2019.² Services sector jobs now represent around three-quarters of employment in both countries. This implies a different growth model and, for labor markets, one that facilitates skills building in skill-intensive services generally and digital skills in particular (Nayyar, Hallward-Driemeier, and Davies 2021). Second, *technological developments* are having a substantial impact on the labor markets of both countries. Employment has already shifted strongly towards jobs that require the kinds of analytical and interpersonal tasks that complement new technologies and away from manual ones (Apella and Zunino 2017, 2022). Notably, technological progress has led to changes in the kinds of work done rather than increased unemployment. There is evidence that technological change is leading to labor market polarization, that is, increasing employment in low- and high-skilled jobs but hollowing out middle-skilled ones (Apella, Rofman, and Rovner 2020). Still, mobile robotics and artificial intelligence (AI) are capable of accomplishing some or even many analytical, interpersonal, and other nonroutine tasks, meaning that further labor market disruption may be on the horizon (Brynjolfsson, Mitchell, and Rock 2023; Brynjolfsson, Li, and Raymond 2023; Eloundou et al. 2023; Gmyrek, Winkler, and Garganta 2024). Finally, efforts to transition to a *low-carbon economy* will both create new skills demands in green growth sectors and reduce the demand for other skills in carbon-intensive industries (World Bank 2022).

These changes are occurring in challenging labor market contexts in both countries. Argentina's labor market has been weak since the late 2000s after the end of the commodity boom and continues to struggle as the government undertakes substantial economic reforms. Informality is high (50 percent of workers in 2023) and private sector job creation is stagnant (for example, 95 percent of job growth between 2012 and the start of the pandemic in 2019 was public).³ Young people and women have particularly poor labor market outcomes (World Bank 2023). Uruguay's labor market is stronger than Argentina's in several respects. The labor force participation rate is higher than Argentina's (64 percent versus 61 percent in 2023), driven by a smaller – though still sizable – gap between female and male participation rates (17 percentage points versus 20 percentage points in 2023).⁴ Uruguay's informality rate of 26 percent in 2023 is much lower than Argentina's and that of Latin America and the Caribbean (LAC) as a whole where the average rate is 50 percent.⁵ However, weaknesses have appeared in Uruguay's labor market since the end of a decade-long period of strong economic growth in 2014 (Torres and McKenzie 2020). The unemployment rate climbed several percentage points from 2014 to 2019 prior to the COVID-19 pandemic in a pattern similar to that of Argentina. Labor force participation has declined, driven by a decline in participation by men.

These overlapping challenges – long-term changes in the nature of work and short- and medium-term labor market weaknesses – are making it difficult for firms to find workers with the right skills and for workers to develop the skills that firms demand. Argentina and Uruguay both score poorly relative to comparator countries on a summary measure of how well human capital is deployed in the labor market (**Figure 1**). Large shares of firms in both countries identify an inadequately

² Data are from the ILO and available in the [World Development Indicators](#).

³ The informality rate is from ILOSTAT. The data on private sector job creation is from the *Encuesta Permanente de Hogares* (EPH) as compiled in the [Boletín de Estadísticas Laborales](#).

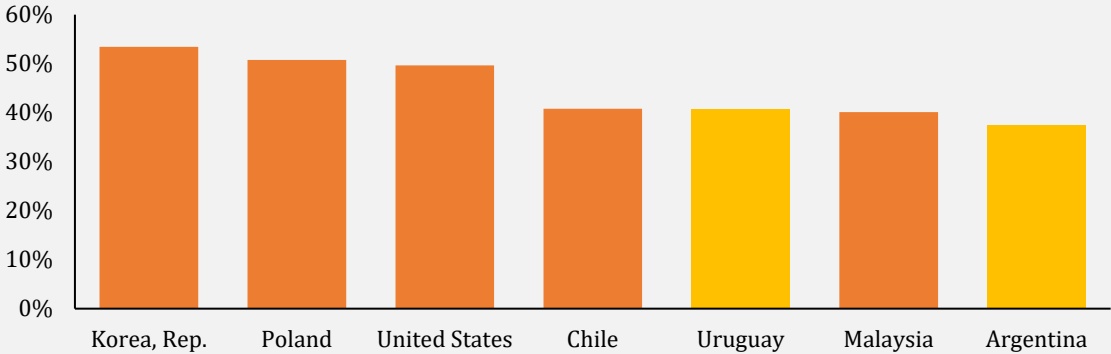
⁴ The data are from ILOSTAT.

⁵ The informality rate for Uruguay is from ILOSTAT. The regional average is from ILO (2023).

educated workforce as a major constraint: 40 percent in Argentina and 37 percent in Uruguay versus a LAC average of 29 percent.⁶ Argentina has the third-highest rate of “qualification mismatch” among G-20 countries (OECD and ILO 2018; OECD 2021b). Uruguay’s continued success in nontraditional services like information and communications technology requires development of complementary technology and engineering skills, which may be undersupplied (World Bank 2015; Che 2021). Skills mismatch is cited as a possible explanation for the relatively high youth unemployment rate (Torres and McKenzie 2020). Recent labor market assessments in Argentina and Uruguay emphasize the importance of reskilling and upskilling programs to respond to rapidly evolving skills needs and of employment support services like labor market intermediation to help link (retrained) jobseekers, particularly disadvantaged ones, to good jobs (World Bank 2022b; Apella, Rofman, and Rovner 2020).

Figure 1: Deployment of human capital in Argentina and Uruguay, 2020

Percentage of productivity



Source: Pennings 2020. The year is 2020 for the Human Capital Index.

Note: The World Bank’s Human Capital Index (HCI) measures the productivity that a child born today could expect to have at age 18 based on their health and education as a percentage of the productivity they could have enjoyed with complete education and full health. The utilization-adjusted HCI (UHCI) incorporates information about employment rates to assess how inefficiencies disrupt deployment of human capital in labor markets (Pennings 2020).

Strong labor market information systems are the backbone of the demand-driven training and employment services systems that can help countries respond to changing skills demands. Establishing skills development pathways that are responsive to labor market needs requires inputs in the form of high-quality, reliable, and up-to-date labor market information and a labor market information system (LMIS) capable of channeling this information to relevant users with different needs. LMISs are increasingly recognized as a cornerstone of labor market policies, particularly as labor market disruptions associated with the changing nature of work require demand-driven upskilling and reskilling (World Bank 2019). When fully realized, these systems support the development and deployment of human capital through labor market intelligence, labor intermediation, career and skills guidance, and links to active labor market policies. While private providers are essential suppliers of these services, public provision is often needed to support disadvantaged groups that are not served by the private sector.

Data extracted from job vacancy postings can help LMISs meet the demand for high-quality, reliable, and up-to-date labor market information. Job postings made by employers in newspapers or, more commonly now, online have a range of information about job demands including location, skill and task requirements, and salary. These postings are made, and can be collected, in real time offering an opportunity for immediate insight into demand.

⁶ World Bank Enterprise Surveys. The year is 2017 for both countries.

This paper investigates whether and how data from job vacancy postings can be used to inform labor market policy in Argentina and Uruguay. The paper first provides a brief introduction to the role of labor market information in general and job postings in particular in improving labor market outcomes. After reviewing the landscape of existing labor market information in Argentina and Uruguay, the paper investigates the quality of online job postings data collected in Argentina and Uruguay between 2020 and 2023. These quality checks reveal the data's strengths and weaknesses.⁷ The paper then explores three potential use cases for the job vacancy postings: 1) using job vacancy postings as a (leading) indicator of labor market health; 2) identifying skills demand at the aggregate and occupation level to inform education and training courses; and 3) identifying similarities between jobs to inform jobseekers, public labor market intermediaries, training institutions, and other actors about potential job transitions.

The paper advances both knowledge about and the use of an important new data source to inform training and employment policy. The paper makes several contributions. First, the online job postings data fills gaps in knowledge about skills demand in Argentina and Uruguay and improves on past approaches to analyzing job transitions, which have been curtailed in scope because of data limitations. Second, the data is expansive, covering multiple sources of online job postings including job search websites, employers, and recruiters. Though important contributions, recent analytical efforts have been limited to keyword analysis and have been focused on a single job search website (Di Ionno and Mandel 2016; Bennett et al. 2022). Additionally, we have access to data from two additional countries – Chile and the United States – that are used for benchmarking in several cases. Third, the paper offers practical guidance for policymakers on how to use the data taking into account the analysis of its strengths and weaknesses.⁸ Finally, the paper makes two methodological contributions. First, we categorize skills both manually and using an unstructured machine learning approach. We provide guidance on how to apply both approaches that can be useful to others working with skills and job postings data. Second, we explore alternative approaches to measuring occupational similarity, again providing guidance on the benefits and drawbacks of the different approaches.

Section 1: The role of labor market information in improving education and employment outcomes

Inadequate information about the labor market can hinder productivity with consequences for growth and equity. Inadequate or inaccurate information about the labor market can lead workers to underinvest in education and training, raise search costs for firms, and reduce the quality of matches between firms and workers. Without a steady supply of good labor market information, educational institutions face difficulties responding to changing skills needs and public sector institutions face difficulties making informed choices about skills investments (World Bank 2021). This can undermine the accumulation and deployment of human capital that underpin economic growth (World Bank 2020; Pennings 2020). Lack of information may pose greater challenges to young, informal, and less-skilled workers because they often rely on personal networks that lack accurate information about good jobs (Carranza and McKenzie 2024).

⁷ Previous research has established that job vacancy postings are suitable for studying skills dynamics including in developing countries. See World Bank (2022c) for Indonesia; Cunningham et al. (2022) for Malaysia; Nomura et al. (2017) for India; Brancatelli, Marguerie, and Brodmann (2020) for Kosovo; and Del Carpio et al. (2017) and Muller and Safir (2019) for Ukraine.

⁸ The paper has been undertaken in close coordination with the *Secretaría de Trabajo, Empleo y Seguridad Social* in Argentina and the *Ministerio de Trabajo y Seguridad Social* in Uruguay.

Labor market information systems (LMISs) can help alleviate information problems. LMISs collect, analyze, store, and disseminate information about labor markets. This labor market intelligence function then feeds into LMISs’ three other core functions: job matching; career and skills guidance; and referral services that connect jobseekers and other labor market stakeholders to other government programs (Testaverde and Posadas 2021). These systems range from basic ones, which focus primarily on the labor market intelligence function, to advanced ones, which utilize the labor market information collected to inform service provision (**Table 1**). More advanced LMISs collect data from a range of sources including surveys, administrative data, and private companies, among others. Strong IT infrastructure, user-friendly interfaces, and a client orientation are key attributes for reliable and efficient systems that seek to be relevant to a range of different users.

Table 1: Hierarchy of labor market information systems

Stage	Description
Basic	<ul style="list-style-type: none"> • Generate basic labor market statistics primarily from survey data • Do not provide services based on labor market data collected • Collaborate with a limited set of public actors
Intermediate	<ul style="list-style-type: none"> • Incorporate additional data sources (e.g., administrative data) • Provide basic services to jobseekers and firms • Collaborate with a range of public actors, including education and training institutions
Advanced	<ul style="list-style-type: none"> • Provide tools to collect, produce, evaluate, and disseminate labor market data from a wide range of sources (including real-time sources) • Offer a full range of services (labor market intelligence, job matching, career and skills guidance, and referral services) targeted to different users • Expand collaboration to include private sector actors

Source: Sorensen and Mas 2016; Testaverde and Posadas 2021.

Effective LMISs are important because better labor market information can contribute to better labor market outcomes. Evaluations of the impact of labor market information interventions show modest impacts (McKenzie 2017). But better design can lead to more impactful programs. Such programs tend to encourage job search in different locations and help update a job searcher’s beliefs about the labor market (Carranza and McKenzie 2024). Providing students with better information about the returns to schooling can increase investments in human capital and adjust them to areas more aligned with labor market demand, as has been shown in the Dominican Republic, Mexico, and the United States (Avitabile and de Hoyos 2018; Jensen 2010; Wiswall and Zafar 2015). Providing workers with access to better information can adjust labor market expectations, encourage search in new occupations and locations, and even improve labor market outcomes directly in some instances, as has been found in India, Peru, the Philippines, and the United Kingdom (Beam, Mckenzie, and Yang 2016; Beam 2016;; Belot, Kircher, and Muller 2019; Belot Muller, and Kircher 2022; Dammert, Galdo, and Galdo 2015; Jensen 2012). More effective training programs also tend to be those that target sectors with growing demand, requiring inputs from LMISs about what jobs and skills are needed (Katz et al. 2022).

LMISs are increasingly incorporating digital tools to enhance the effectiveness and efficiency of delivery. These tools include both improvements in data collection and in how services are provided to jobseekers and other clients. There is evidence that these tools improve outcomes. In Peru, an impact evaluation of an automated job recommendation system showed a significant increase in usage of the labor intermediation system from 3 percent of unemployed people to 14 percent. A randomized controlled trial found positive impacts on employment from using the system’s digital

job matching tool (Dammert, Galdo, and Galdo 2015). A similar tool deployed in the United Kingdom that provided low cost, automated job search advice to jobseekers found that the tool increased the number of interviews by 44 percent, driven mostly by jobseekers who were searching more narrowly prior to the intervention (Belot, Kircher, and Muller 2019; Belot, Muller, and Kircher 2022).

LMISs are also increasingly turning to job postings, particularly online job postings, as a source of labor market insights and translating these insights into policy. Job postings data is now commonly used for labor market analytics.⁹ Frequent uses of these data include labor market monitoring and analysis, including over time and at the local level (Forsythe et al. 2020; Shen and Zhu 2023; Evans et al. 2023); assessing demand for skills, including newly emerging skills and including predictive analysis (OECD 2021, 2022; Cunningham et al. 2022; Sato et al. 2023; Borgonovi et al. 2023); and improving skills matching (Apella and Zunino 2022b; Samek, Squicciarini, and Cammeraat 2021).¹⁰ **Table 2** provides examples of how governments across the development spectrum are utilizing job postings data.

Table 2: Examples of uses of job postings data

Country	Example
Australia	• The Internet Vacancy Index identifies changes in demand in real time
European Union	• The European Union’s Skills-OVATE tool provides detailed jobs and skills data based on online job ads from 28 European countries
Indonesia	• The Online Vacancy Outlook provides detailed skills profiles of occupations
Malaysia	• Online job postings are used as an indicator of shortage in the Critical Occupations List, which identifies occupations in high demand
Malawi	• Online job postings have been used to identify in-demand jobs
Myanmar	• Online job postings have been used to identify in-demand jobs
Netherlands	• Public employment services use online job postings and CVs for job matching
New Zealand	• The Ministry of Business, Innovation, and Employment packages information from online job postings for an online education and career exploration tool
Singapore	• SkillsFuture Singapore incorporates online job postings into assessments of priority skills to inform its lifelong learning initiatives
United Kingdom	• The Migration Advisory Committee uses online postings as an indicator of shortage for its Shortage Occupation List
United States	• The occupational database O*NET uses online job postings to identify quickly evolving changes in skills and job titles

Source: See World Bank (2022c) for Indonesia; UNESCO (2019) for Malawi; CSC (2019) for Malaysia; UNESCO (2019b) for Myanmar; CEDEFOP, European Commission, ETF, ILO, OECD, and UNESCO (2021) for the Netherlands and New Zealand; SkillsFuture Singapore (2022) for Singapore; MAC (2017) for the United Kingdom; and World Bank (2019b) for the United States.

Job postings data complements survey and administrative data. The relative advantages of job postings data include granularity across multiple dimensions (jobs, skills, geographies, and employers); near real-time collection; and detailed information on skills and qualifications that are

⁹ For recent published work using job postings, see Atalay et al. (2020); Azar et al. (2020); Brown and Matsa (2020); Conzelmann et al. 2023; Deming and Kahn (2018); Deming and Noray (2020); Forsythe et al. (2020); Hansen et al. 2023; Hershbein and Kahn (2018); Kuhn and Shen (2013, 2015); Marinescu (2017); Modestino, Shoag, and Ballance (2016); and Napierala and Kvetan (2022), among others. For World Bank publications, see Nomura et al. (2017) for India; World Bank (2022c) for Indonesia; Brancatelli, Marguerie, and Brodmann (2020) for Kosovo; Cunningham et al. (2022) for Malaysia; and Del Carpio et al. (2017) and Muller and Safir (2019) for Ukraine. For an overview of recent literature, see Fabo and Kureková (2022).

¹⁰ Beyond postings themselves, metadata on how jobseekers utilize portals can provide insights into job search strategies and preferences (Faberman and Kudlyak 2019).

in demand (**Table 3**). Survey data, in contrast, tends to be collected only periodically at high levels of aggregation. For example, survey data cannot provide any insight into labor markets at the level of job titles while job postings data can be analyzed from the level of skills, job titles, and occupations. Importantly, the cost of job postings data is typically a fraction of that of survey data. Collecting postings data can be done digitally and undertaken by specialized firms benefiting from economies of scale while survey operations generally require governments to hire substantial human resources and deploy them regularly (ILO 2020). However, survey data are typically representative while job postings may have less coverage of certain occupations and industries, an issue that may be exacerbated in contexts of high labor market informality (Cammeraat and Squicciarini 2021). Survey data are also able to provide information about subpopulations (for example, by gender or education), which permits identification of potentially disadvantaged groups. Job postings and survey data also offer insights into different aspects of labor markets. Job postings data provide insight into labor market demand, while survey data shows labor market supply as well as the outcomes of labor market matches (employment) and non-matches (unemployment). Administrative data may offer some of the advantages of job postings data, but their utility is highly dependent on the program to which they are attached and the quality of collection efforts. In sum, job postings data can make up for some of the weaknesses of survey and administrative data and survey data can make up for some of the weaknesses of job postings data.

Table 3: The strengths and weaknesses of job postings and survey data

Characteristic	Job postings	Surveys
Coverage	Granular, but not representative	Representative, but not granular
Frequency	Near real-time	Periodic
Focus	Demand	Supply, matches of supply and demand
Subpopulation analysis	No	Yes
Skills	Yes	No, except for specialized surveys*

Section 2: Labor market information systems in Argentina and Uruguay

The governments of Argentina and Uruguay are undertaking efforts to improve their labor market information systems. Argentina’s *Fomentar* program supports an ecosystem of demand-driven training and employment services. An [occupational observatory](#) has been developed that collects information on tasks undertaken at work to help inform these services. The employment portal (*Portal Empleo*) has been created as a single place for jobseekers to search for jobs, access workshops to improve job search and readiness skills, and enroll in training. In Uruguay, the *Instituto Nacional de Empleo y Formación Profesional* (INEFOP) was created in 2008 to monitor skills demand and provide training. While INEFOP has the mandate and is well funded to provide training, its focus is mostly on technical skills that are in high demand in the short term, and not on longer-term strategic reskilling. Following the example of the O*NET occupational database in the United States, Uruguay piloted a skills monitoring tool with World Bank technical assistance to help detect skills needs across occupations. The pilot covered 22 occupations.¹¹ This pilot informed subsequent analytical efforts including to understand similarities between jobs, which can inform the efforts of employment services providers working with jobseekers (Velardez 2021; Parrilla 2022).

However, these efforts are incipient in both countries, and both lack detailed data on in-demand skills. In Argentina and Uruguay, the improvements in labor market information systems are in the

¹¹ The *Ministerio de Trabajo y Seguridad Social* began designing questionnaires to collect occupational information in 2017. The pilot was undertaken in late 2019 and early 2020.

early stages. Recent labor market assessments emphasize the importance of developing demand-responsive training and employment programs to respond to rapidly evolving skills needs (World Bank 2022, 2022b; Apella, Rofman, and Rovner 2020). One of the key components missing from the LMISs of both countries is a source of up-to-date, detailed information about skills demand and supply (**Table 4**). Apart from the early-stage occupational observatories developed in Argentina and Uruguay, there is no source of detailed information on employer demand for skills or worker supply of skills. The *Encuesta Permanente de Hogares* and the *Encuesta de Indicadores Laborales* in Argentina and the *Encuesta Continua de Hogares* in Uruguay provide general insight into labor market demand, but only do so at an aggregated occupational level, limiting their utility for jobseekers, training institutions, and other labor market stakeholders.¹² The occupational observatories do provide detailed insight into skills, activities done at work, and other job characteristics, but given their resource intensity can only be done periodically, for a subset of occupations each year, and at a relatively high cost meaning that quickly changing skills needs may not be captured.¹³ In the absence of detailed skills data, researchers have turned to data from other countries, particularly O*NET in the United States, to help governments understand skills dynamics in Argentina and Uruguay. However, the assumptions underlying this research – that occupations have the same skills across countries and within occupations over time – are unreliable.¹⁴

Jobs postings data are available in both countries but have not yet been used widely for labor market analysis. In Argentina, Di Ionno and Mandel (2016) uses job postings from the job search site Indeed.com to identify jobs associated with the “App Economy.” The authors use keywords and phrases to identify these jobs. The Universidad Torcuato di Tella in Argentina compiled a now defunct index of labor market demand using job advertisements in the newspapers *Clarín* and *La Nación* (Albertini, Poirier, and Trupkin 2019). The *Secretaría de Trabajo, Empleo y Seguridad Social* has also investigated the possibility of collecting and analyzing online job postings. In Uruguay, Bennett et al. (2022) analyzes job postings and applications from the online job search site *Buscojobs*. The authors restrict their scope to an evaluation of whether the postings and vacancies are suitable for studying skills dynamics, though follow-up work is ongoing (Escudero, Liepmann, and Vergara 2023). These analyses demonstrate that raw job postings data is available and can be used for labor market analysis. However, the data have not yet been processed and made available in accessible, policy-relevant ways.

¹² The occupational classification schemes used in these surveys – the *Clasificación Nacional de Ocupaciones* (CNO) in Argentina and the *Clasificación Internacional Uniforme de Ocupaciones* (equivalent to the International Standard Classification of Occupations (ISCO) 2008) in Uruguay – provide information about the tasks and qualifications required for occupations. However, these schemes provide general information about common tasks and little information about specific skills. Additionally, they are updated infrequently. Argentina’s CNO faces a particular challenge because it is only comparable to ISCO at a high level of aggregation, limiting the ability to benchmark Argentina’s performance against comparators. See INDEC (2018); La Buonora (2017); and Molina, Bernasconi, and de la Fuente (2020).

¹³ Even the best practice O*NET occupational observatory in the United States only updates its occupational information periodically.

¹⁴ For work showing cross-country differences in skills content within occupations, see Caunedo, Keller, and Shin (2021); Dicarolo et al. (2016); Lewandowski et al. (2022); and Lo Bello, Sanchez-Puerta, and Winkler (2019). For work showing differences in skills content over time within occupations, see Atalay et al. (2020); Autor (2022); and Lin (2011).

Table 4: Sources of labor market information in Argentina and Uruguay

a. Argentina					
Source	Example	Coverage	Frequency	Strengths	Weaknesses
Household surveys	<i>Encuesta Permanente de Hogares</i>	Large urban areas	Monthly	- Data on formal and informal jobs - Job titles can be classified in ISCO	- Coverage gap (workers living in small urban and rural areas not included) - Limited data about skills - No data on labor demand - ISCO classification only possible at 2 digits
Firm surveys	<i>Encuesta de Indicadores Laborales</i>	Firms with 10+ formal employees in 12 urban centers	Monthly	- Data on hirings and dismissals - Data on unmet labor demand	- Coverage gap (12 provinces, small firms, public employers not included) - No data on informal and self-employment - Limited data on skills - Classification in ISCO not possible
Administrative data	<i>Sistema Integrado Previsional Argentino</i>	All formal workers	Real time	- Data on hirings and dismissals - Employment and firm dynamics - Job titles can be classified in ISCO	- No data on informal employment - Limited data on skills
Occupational profiles	<i>Observatorio Ocupacional Argentina</i>	Subset of occupations	Periodic	- Detailed data on skills and other job characteristics - Job titles can be classified in ISCO	- Limited coverage of occupations
b. Uruguay					
Source	Example	Coverage	Frequency	Strengths	Weaknesses
Household surveys	<i>Encuesta Continua de Hogares</i>	Whole country	Monthly	- Data on formal and informal jobs - Job titles can be classified in ISCO	- Limited data on skills - No data on labor demand
Firm surveys	NA	NA	NA	NA	- No firm surveys
Administrative data	<i>Banco de Previsión Social</i>	Formal workers	Real time	- Employment can be classified in ISIC - Panel structure available	- No data on informal employment - Limited data on skills - Classification in ISCO not possible - Not freely accessible
Occupational profiles	<i>O*NET Uruguay</i>	Subset of occupations	Periodic	- Comparable to O*NET USA - Detailed data on skills and other job characteristics - Job titles can be classified in ISCO	- Limited coverage of occupations - Not freely accessible

Note: ISCO = International Standard Classification of Occupations; ISIC = International Standard Industrial Classification; NA = Not applicable.

Section 3: The quality of online job postings data in Argentina and Uruguay

This section investigates the quality of online job postings data collected by Lightcast in Argentina and Uruguay between 2020 and 2023. These quality checks reveal the data's strengths and weaknesses. The section also discusses methodological choices made in cleaning the data.

3.1 Overview of the data

Data on online job postings were obtained from Lightcast, a leading provider of labor market information. Lightcast collects online job postings data for the United States, many European Union countries, and several countries in Asia and began doing so in 2020 for several countries in Latin America. Lightcast collects data in many developing country labor markets including Indonesia and Malaysia and in several countries in Africa. Data was obtained for Argentina and Uruguay for this work. Data includes the raw text of the job postings as well as fields extracted by Lightcast from this text. These fields include posting date, posting duration, job location, employer name, job title and occupation name categorized into the International Standard Classification of Occupations (ISCO) at the 4-digit level, sector categorized into the Nomenclature of Economic Activities (NACE) at the 4-digit level, skill keywords and categories, and salary offered.

Lightcast collects data on job postings from around 1,800 sources for Argentina and around 550 for Uruguay. Most sources are employers followed by job boards and recruiters. Education and government sources are also included. Data is available beginning in May 2020 for Argentina and August 2020 for Uruguay (**Table 5**). Data is available for both countries through December 31, 2023. Thus, data is available for around 3.5 years for Argentina and Uruguay. A total of 3,090,055 postings is available for Argentina and 144,272 for Uruguay (**Table 6**). In both countries, the number of postings has increased over time, demonstrating in part Lightcast's improved ability to capture postings.¹⁵

Table 5: Online job postings data collection period

	Argentina	Uruguay
First date	21-May-20	1-Aug-20
Last date	31-Dec-23	31-Dec-23
Total period (months)	43	41

Source: Lightcast.

Table 6: Online job postings by year

Year	Argentina		Uruguay	
	#	%	#	%
2020	158,967	5.1%	8,606	6.0%
2021	680,930	22.0%	44,362	30.8%
2022	884,053	28.6%	41,159	28.5%
2023	1,366,105	44.2%	50,113	34.7%
Total	3,090,055	100%	144,240	100%

Source: Lightcast.

¹⁵ Half of online job postings were posted to job search sites for 60 days in Argentina and Uruguay. Longer vacancy duration does not seem to be correlated with the type of jobs advertised, as the distribution of occupations does not vary with posting duration.

3.2 Quality checks

The assessment of the quality of the online job vacancy data follows a two-step approach. First, the quality checks evaluate the quality of the data itself focusing on missing values and the classification of values. Second, the quality checks investigate the representativeness of the data looking at the distribution of the online job postings over time, across geographies, across occupations, and across employers. This process follows approaches used in Vermuelen and Gutierrez (2024) and Cammeraat and Squicciarini (2021) to assess the quality and utility of (Lightcast’s) online job postings data.

3.2.1 Data quality

Location, employer, occupation, and skills identifiers are available for most vacancy postings in Argentina and Uruguay. Around 90 percent of vacancies across both countries has information on location, employer, occupation, and skills (**Table 7**). Information on sector and salary, in contrast, is limited, as is the case for online job vacancy data collected by Lightcast in English-speaking countries (Tsvetkova et al. 2024). In the case of sector information, the substantial volume of missing values is in part due to the lack of information about the sectors of individual employers, which can assist with sector classification.¹⁶ Salary information is commonly omitted by employers when posting job vacancies.

Table 7: Missing values in online job postings data, 2020-23

Attribute	Argentina		Uruguay	
	#	%	#	%
Location	262,887	9	5,163	4
Employer name	124,142	4	3,296	2
Occupation	93,372	3	2,734	2
Sector	1,943,604	63	92,729	64
Skills	263,462	9	10,957	8
Salary	3,065,601	99	141,619	98

Source: Lightcast.

Beyond missing data, problems with classification of location are apparent in Argentina, though these problems decreased over time as Lightcast’s algorithm improved. The location of vacancies is misclassified in two cases in Argentina. First, job vacancies are at times incorrectly allocated between the City of Buenos Aires (*Ciudad Autónoma de Buenos Aires* (CABA)) and the Province of Buenos Aires (*Provincia de Buenos Aires* (PBA)). For example, if a job advertisement specifies a municipality within PBA the name of the province is used as the location. Second, job vacancy locations are not identified for CABA if the location specified is a *comuna*¹⁷ within CABA. These two factors lead to a distribution of job vacancies between CABA and PBA that does not reflect their relative population. This occurs until the end of 2021 when Lightcast updated its algorithm resulting in an increase in the share of vacancies in CABA.

¹⁶ In the case of Argentina, attempts to link employer name to the *Administración Federal de Ingresos Públicos* (AFIP) business registry dataset did not produce a useful match. This is because the names of companies in job advertisements often differ substantially from their legal names as registered with the tax authorities. However, additional efforts to clean employer names may yield better results and are an area for further exploration.

¹⁷ A *comuna* is an administrative unit in the City of Buenos Aires.

3.2.2 Data representativeness

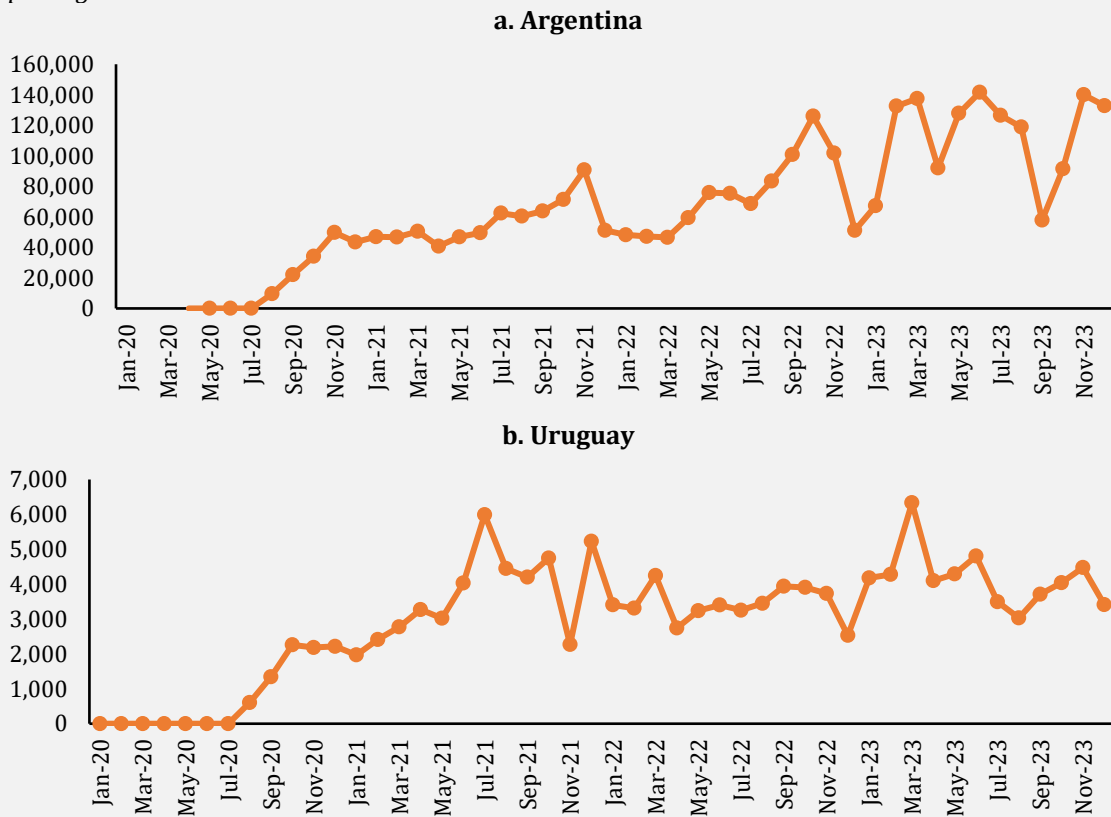
Representativeness is a common concern with job postings data. Job postings are often thought to be biased towards higher-skilled, formal, urban jobs while less-skilled, informal, nonurban ones are more likely to be filled through informal means like word of mouth. The following sections examine the representativeness of the postings data.

3.2.2.1 Representativeness across time

The number of online job postings collected each month generally increased over time (**Figure 2**). The number of postings collected in December 2023 was substantially higher than in December 2021 in Argentina and Uruguay. Several factors likely explain this increase over time. First, Lightcast began data collection during the COVID-19 pandemic. The growth in the volume of online job postings over time may reflect the recovery of the labor market as lockdown measures were eased. Second, Lightcast’s data collection efforts improved over time, leading to collection of more postings. In addition to the increase in postings over time, seasonal trends are observable with dips in postings in the summer months of the southern hemisphere (January to March).

Figure 2: Online job postings in Argentina and Uruguay, 2020-23

of postings



Note: Postings are available from May 2020 to December 2023 for Argentina and from August 2020 to December 2023 for Uruguay.

Source: Lightcast.

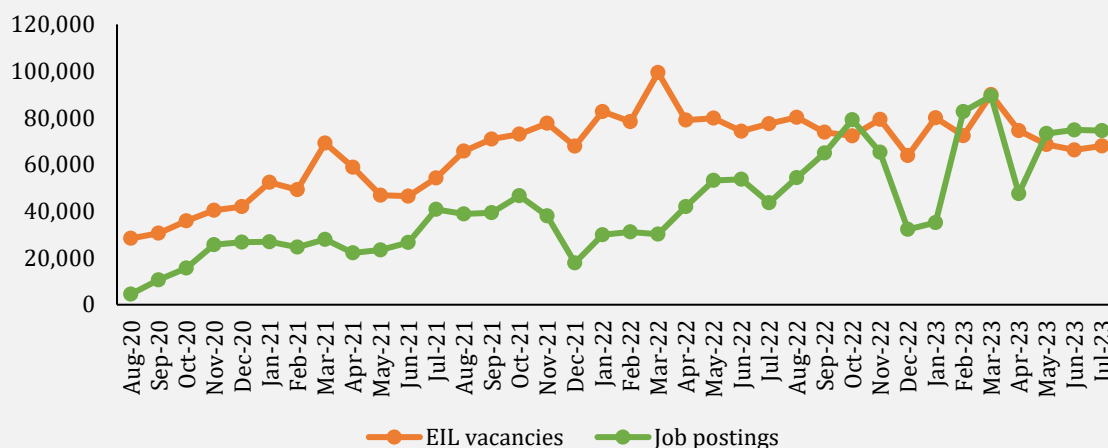
The representativeness of the online job postings data over time can be tested by comparing the postings data to information on vacancies from other sources. In Argentina, the *Encuesta de Indicadores Laborales* (uh) collects information on vacancies and so can be used to evaluate the time

trends in vacancies collected by Lightcast. Similar data is not available for Uruguay. The EIL collects monthly vacancies from private employers with more than 10 employees located in 12 large urban centers in the northern and central regions of Argentina. 14 provinces are excluded from the sample. The EIL does not collect detailed information on occupations or report a breakdown by geographical area, so only time trends in total vacancies can be assessed. The Lightcast data have been processed to make the two data sources as comparable as possible by including only the vacancy data for the same provinces as the EIL, excluding postings from staffing firms and the public sector, and taking into account only those online job postings that have been active for two months (to compare flow rather than stocks). Even after controlling for these characteristics, the vacancies collected by Lightcast still differ from the EIL dataset in two ways. First, EIL vacancies refer to vacancies in the previous month and new hires in the current month of the survey, whereas Lightcast data represent active online vacancies. Second, it is not possible to distinguish the proportion of EIL vacancies that are advertised online versus offline.

Time trends in online job postings collected by Lightcast are consistent with those observed in the EIL in Argentina. The two series, beginning in August 2020 and ending in July 2023, which is latest month for which EIL data are available, have a correlation of 0.57 (Figure 3). This is in the same range as correlations that have been found in other settings.¹⁸

Figure 3: Comparing job postings and vacancies in the Argentina EIL, 2020-23

of postings/vacancies



Note: The period is August 2020 to July 2023. Postings from provinces not covered by the EIL, from staffing firms and the public sector, and that are active for more than two months are excluded. EIL = *Encuesta de Indicadores Laborales*.
Source: Lightcast; EIL.

In Argentina, the evolution of online job postings during the pandemic showed a strong negative relationship with the COVID Stringency Index. During the pandemic, Argentina had a long period of highly restrictive measures compared to Uruguay. The Stringency Index¹⁹ ranged from 90 out of 100 in April 2020 to around 75 out of 100 in mid-2021 in Argentina, while the maximum values in

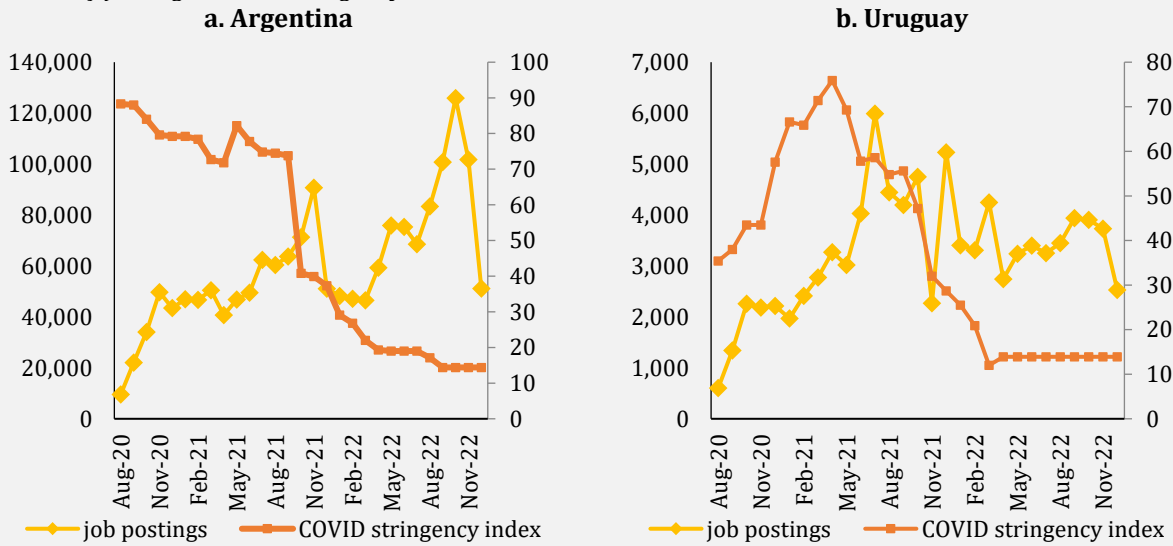
¹⁸ For example, Carnevale, Jayasundera, and Repnikov (2014) find a correlation coefficient of 0.75 between job postings and survey-measured vacancies in the United States. See Cammeraat and Squicciarini (2021).

¹⁹ The index summarizes the mean score of nine response measures that reflect the severity of government policies during the pandemic, including school closures; workplace closures; cancellation of public events; restrictions on public gatherings; public transport closures; stay-at-home requirements; public information campaigns; restrictions on internal movement; and international travel controls. The higher score (100) indicates a stricter response (Hale, et al 2021).

Uruguay were only above 70 for March and April 2021 (**Figure 4**). In Argentina, the correlation between the index and the online job postings was -0.66 for the period between August 2020 and December 2022. Conversely, in Uruguay the correlation coefficient of -0.07 is much weaker, perhaps reflecting a more limited labor market impact of the less restrictive COVID-19 measures.

Figure 4: Job postings and COVID Stringency Index in Argentina and Uruguay, 2020-22

Number of postings, COVID Stringency Index



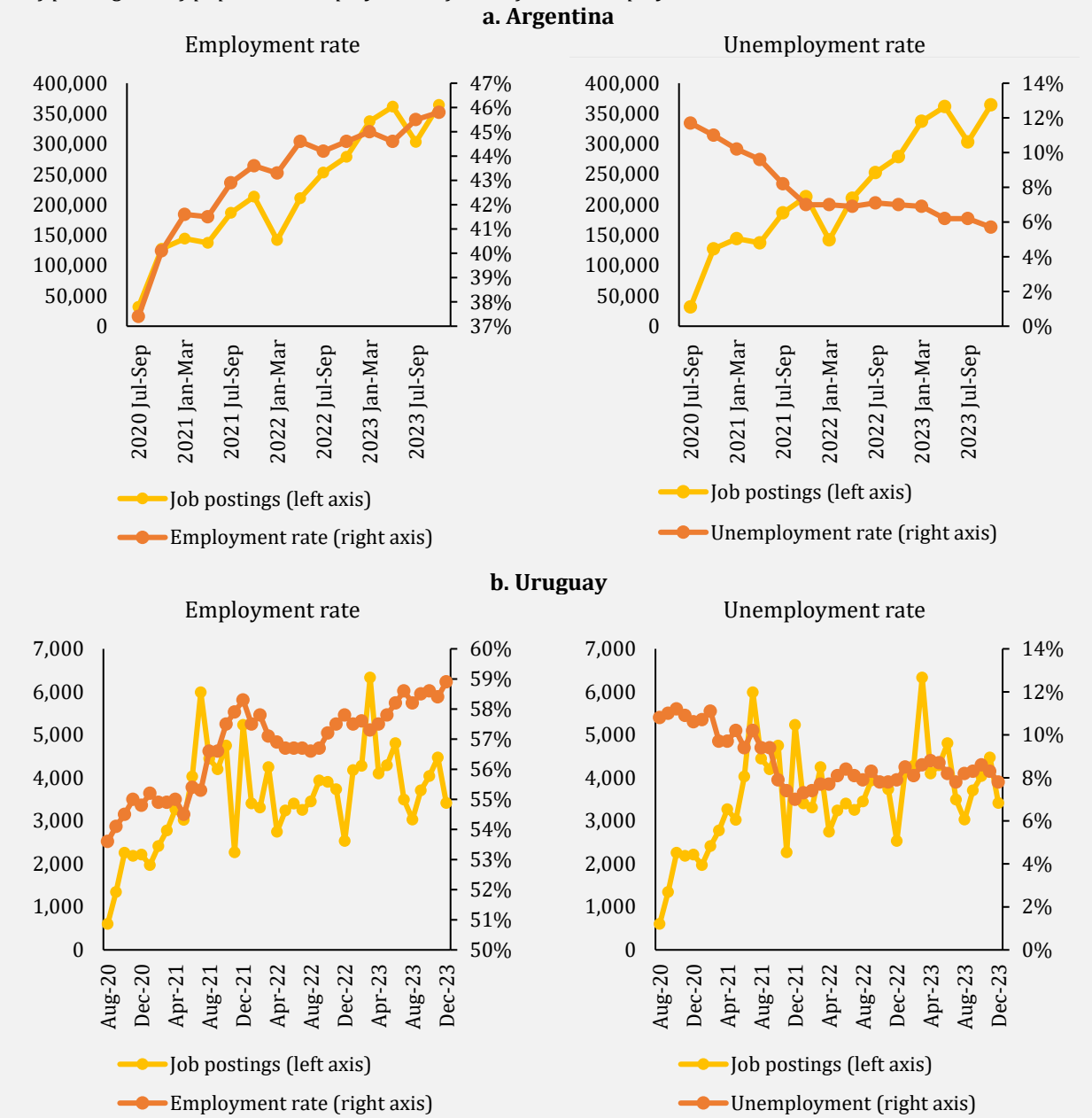
Note: The period corresponds to August 2020 to December 2022 for Argentina and Uruguay.

Source: Oxford COVID Stringency Index; Lightcast.

In Argentina and Uruguay, job postings are closely correlated with labor market indicators over time. Representativeness across time can also be tested by comparing how closely trends in vacancies follow trends in employment and unemployment rates, with positive correlations expected for the former and negative correlations expected for the latter (**Figure 5**). The correlations are high in both countries, though they are stronger in Argentina (0.88 and -0.86) than in Uruguay (0.56 and -0.50).

Figure 5: Job postings, employment rate, and unemployment rate in Argentina and Uruguay, 2020-23

of postings, % of population employed, % of labor force unemployed



Note: The period corresponds to the third quarter of 2020 to the fourth quarter of 2023 for Argentina and August 2020 to December 2023 for Uruguay.

Source: Encuesta Permanente de Hogares (Argentina); Encuesta Continua de Hogares (Uruguay); Lightcast.

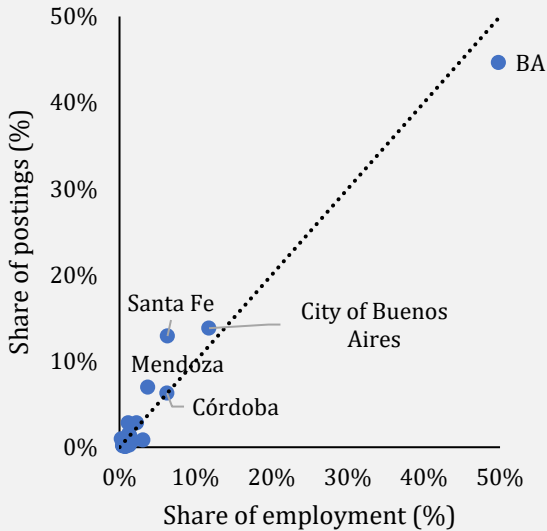
3.2.2.2 Representativeness across geographies

Job postings generally follow the distribution of employment in Argentina and Uruguay. In Argentina, five provinces – the Province of Buenos Aires, the City of Buenos Aires, Santa Fe, Córdoba, and Mendoza – are responsible for 82 percent of job postings. These provinces also have the largest share of jobs in Argentina, though there are some differences in the distribution of postings and jobs (**Figure 6**). In the last quarter of 2023, the share of job postings was double that of jobs in the

provinces of Santa Fe (13 percent of postings versus 6 percent of jobs) and Mendoza (7 percent of postings versus 4 percent of jobs), while the share of postings is only slightly higher in the City of Buenos Aires (14 percent of postings versus 12 percent of jobs). Jobs are more concentrated than postings in the Province of Buenos Aires (45 percent of postings versus 50 percent of jobs). In the case of Uruguay, around 60 percent of job postings are located in Montevideo, which accounts for approximately 41 percent of employment. This difference could be explained by a greater dynamism of job inflows and outflows in the capital, though this hypothesis cannot be tested due to the absence of a labor vacancy survey in Uruguay.

Figure 6: Job postings and employment by geographical areas in Argentina, 2023

% of postings, % of employment



*Note: The period corresponds to the fourth quarter of 2023.
Source: Encuesta Permanente de Hogares (Argentina); Lightcast.*

3.2.2.3 Representativeness across occupations

High-skilled occupations are overrepresented in the job postings data. There is a pattern of underrepresentation of low-skilled occupations and overrepresentation of high-skilled occupations across the three countries (**Table 8**). In Argentina and Uruguay, elementary occupations account for less than 2 percent of job vacancies but 14 and 18 percent of jobs, respectively. In contrast, managerial, professional, and technical occupations account for around 70 percent of vacancies but only around 25 percent of employment. The results are consistent with recent assessments for English-speaking countries, which find that the occupational representativeness of Lightcast data is low compared to official vacancy data (Tsvetkova et al. 2024).

Table 8: Percentage distribution of job postings and employment by occupational group, 2020-23

Occupation	Argentina		Uruguay	
	Postings	Employed	Postings	Employed
Managers	14	5	11	3
Professionals	41	13	34	14
Technicians	18	9	20	8
Clerical support	5	11	6	11
Service & sales	11	25	13	22
Skilled agricultural	0	0	0	4
Craft & related trades	4	13	5	13
Plant & machine	3	9	4	7
Elementary occupations	1	14	2	18
Not identified	3	2	4	0
Total	100	100	100	100

Note: The period for vacancies data is May 2020 to December 2023 for Argentina and August 2020 to December 2023 in Uruguay. The period for employment data corresponds to the second half of 2022 for Argentina and 2022 for Uruguay.
Source: Encuesta Permanente de Hogares (Argentina); Encuesta Continua de Hogares (Uruguay); Lightcast.

3.2.2.4 Representativeness across employers

A small number of companies is responsible for many postings (**Table 9**). Of the job postings with a classified employer²⁰, the top 10 companies account for 11 percent in Argentina and 22 percent of postings in Uruguay. Most of these top companies are staffing agencies. They account for 5 percent of the total number of postings in Argentina and 5 percent in Uruguay.

Table 9: Top employers by share of job postings, 2020-23

a. Argentina		b. Uruguay	
Employer name	%	Employer name	%
Unclassified	32.4	Unclassified	33.5
ManpowerGroup*	2.1	ManpowerGroup*	5.2
Adecco*	1.5	Lockheed Martin	4.6
Randstad*	1.5	BairesDev	3.7
Adn Recursos Humanos	1.1	Randstad	2.7
Marriott International	1.1	And Advice	2.3
Grupo Gestión	0.8	Adecco	1.3
BairesDev	0.8	Sabre	0.7
Consultores De Empresas	0.8	Sophilabs	0.7
Accenture	0.8	Grupo Humano	0.6
Bayton	0.6	Work Office S.A.S	0.6

Note: The period for vacancies data is May 2020 to December 2023 for Argentina and August 2020 to December 2023 for Uruguay. The * indicates a staffing agency.
Source: Lightcast.

3.3 Quality assessment conclusions

The job postings data are a quality source of information about the labor market, though biases are present and must be accounted for when analyzing the data. Comparison of the job postings data

²⁰ In Argentina, Lightcast is able to classify 44,755 of the 182,408 unique company names in the dataset. In Uruguay, Lightcast is able to classify 2,628 of the 7,428 unique company names.

with vacancy and other labor market data show that the postings broadly capture labor market dynamics, including over time and across geographies. However, postings are not representative of occupations. They are biased towards higher-skilled jobs and away from lower-skilled ones. This likely reflects the method to collect the vacancies: the data capture formal recruitment processes, and so are likely less useful to investigate informal or less-skilled jobs that are more likely to be filled through word-of-mouth or other informal recruitment channels. Still, the size of the data means that inferences can at times be made even where the data is biased. For instance, the 4 percent of job postings represented by craft and related trades occupations in Argentina is equivalent to nearly 125,000 job postings, providing a large sample on which to base analysis of skills in jobs within this occupation category. The section also revealed a few areas where additional data cleaning could increase the utility of the data, including cleaning employer names to facilitate mapping job postings to an economic sector.

Section 4: Putting job postings data to work in Argentina and Uruguay

This section discusses practical applications of the job postings data to labor market analysis. Three use cases are discussed: 1) using the postings to identify skills demand; 2) using the postings to identify similar occupations; and 3) using the postings to describe labor market health.

4.1 Use case #1: Using job vacancy postings to identify skills demand

This section uses the job postings data to assess the current demand for skills in Argentina and Uruguay. The first subsection describes the methodology for classifying skills into conceptual skill groups. Descriptive statistics based on the resulting taxonomy are then presented to illustrate the prevalence of skill types in job postings and occupations. Results are also presented for Chile and, at the occupation level, the United States as comparators. Skills were classified using data from online job postings in Chile along with Argentina and Uruguay to permit inclusion of Chile as a comparator. The classifications were then applied to the United States.

4.1.1 Methodology

The skills data in the online job postings dataset is structured into individual skills that are identified by Lightcast. The postings data contain a list of skills required for each vacancy that is drawn from the raw text of the job advertisement. There are 19,344 unique skills in the Argentina postings and 12,101 in the Uruguay postings. Lightcast classifies each of these skills into 32 categories (for example, Administration and Analysis) and more than 500 subcategories (for example, Office and Productivity Software and Statistics), though not all skills are categorized.²¹ Lightcast also identifies software skills. Lightcast creates the categories based on skills that appear together frequently but are not created with specific end users in mind.

Two approaches to skills classification are undertaken. Argentina and Uruguay do not currently have skills classification taxonomies, so we undertake our own classification.²² The first approach manually classifies skills keywords into a skills taxonomy. Skills categories are predetermined based on existing literature and then skills are mapped to the categories based on keywords. The second approach presents an experimental methodology based on machine learning that allows the data to

²¹ In Argentina, 86 percent of the unique skills are categorized by Lightcast. In Uruguay, 92 percent are.

²² Argentina has system for certifying competencies (*Sistema Nacional de Certificación de Competencias y Formación Continua*) that provides competence guidelines for around 300 job roles. However, skills are not harmonized and listed.

speak for itself in determining both skills categories and the skills that should appear in those categories.

4.1.1.1 Manual classification

We develop an original skills taxonomy to create skills categories that are relevant to and recognizable by training and employment services practitioners as well as jobseekers. To do so, we build on existing approaches taking advantage of Lightcast's categorization where practical. We build on several recent efforts, each of which includes skills keywords that we use to assist our categorization. Deming and Noray (2020) uses keywords and phrases to classify skills requirements in online job postings into 14 non-exhaustive categories. Similarly, Alekseeva et al. (2021) uses a list of keywords to tag artificial intelligence skills identified in online job postings. Bennett et al. (2022) also classifies skills data from online postings on job portals in four countries into three categories (cognitive, social, and manual) and 14 subcategories using a list of keywords, synonyms, and phrases. Finally, Cunningham et al. (2022) uses Lightcast's own basic categorization to classify skills into 7 categories: digital (low, medium, high); cognitive; socioemotional; language; and technical.

We define five categories and 46 subcategories of skills. These are cognitive, socioemotional, digital, manual, and technical. The categories are meant to be exhaustive of all skills. **Appendix 2** provides a more detailed explanation of the categories. The five categories are divided into more granular subcategories as follows:

- *Cognitive* (11 subcategories): thinking, mathematics, communication skills, financial skills, business systems, quality control, business analysis, project management, data analysis, language skills and adaptability.
- *Digital* (3 subcategories): basic, intermediate, and advanced.
- *Socioemotional* (8 subcategories): teamwork, communication, general social skills, organizational skills, character, customer service, people management, and creativity.
- *Manual* (5 subcategories): finger dexterity, hand-foot-eye coordination, driving, flying, and physical skills.
- *Technical* (19 subcategories): product lifecycle management, STEM, health, care services, social sciences, law, education, administrative support, security, sales, management, machinery repair, media production, art, environment, energy, construction, military, and agriculture.

The categorization occurs in two steps. First, the categorization of skills keywords in the job postings data follows the approach of Cunningham et al. (2022). Cunningham et al. (2022) create a dictionary of keywords in online job postings data collected by Lightcast in Malaysia and use them to categorize keywords into four categories (cognitive, socioemotional, digital, and manual) following established skills taxonomies. We replicate this approach in the first step of our categorization, which results in categorization of 60 percent of the 21,779 unique skills in the job postings data.²³ In the second step, the subcategory name provided by Lightcast is used to classify the remaining skills in the first four categories and to create the technical category for those skills related to specific sector or occupational knowledge that could not be classified in the first four categories. This results in the classification of 90 percent (19,402) of the unique skills.

²³ The unique list is compiled for Argentina, Chile, and Uruguay. The taxonomy is then applied to the United States.

The skills taxonomy is used to categorize all of the unique skills that appear in the job postings into the five nonoverlapping categories (**Table 10**). In all countries, technical skills account for the largest share of unique skills (around 40 percent), followed by digital skills (around 30 percent) and cognitive skills (around 15 percent).

Table 10: Distribution of skills by categories in Argentina, Chile, and Uruguay

Skill	Argentina		Chile		Uruguay	
	#	%	#	%	#	%
Cognitive	2,413	14	2,210	14	1,769	16
Digital	6,067	35	5,260	34	3,497	31
Social	1,066	6	1,015	6	835	7
Manual	868	5	784	5	588	5
Technical	7,013	40	6,424	41	4,512	40
Total	17,427	100	15,693	100	11,201	100

Note: Unclassified skills are not shown. 1,917 skills (10 percent) are not categorized in Argentina, 1,787 (10 percent) in Chile, and 900 (7 percent) in Uruguay.

Source: Lightcast.

4.1.1.2 Classification using machine learning²⁴

The machine learning classification approach allows the data to define skills categories. Significant advancements in natural language processing (NLP), particularly in generative pre-trained transformers – the type of model on which ChatGPT is built – allow for data-driven approaches to categorizing text, including skills keywords. These approaches allow the data itself to determine which skills keywords are related to each other and so should be classified into the same category of a taxonomy. The machine learning approach, in our formulation, transforms words and phrases into vectors (embeddings), then uses clustering approaches to classify “similar” vectors into categories and subcategories, and, finally, uses thematic topics generated from the vectors themselves to define labels for the categories and subcategories. While researchers make key decisions and define parameters, categories and labels are largely derived endogenously. The manual approach, in contrast, involves defining categories *ex ante* and then mapping keywords to categories, relying on previous categorizations and researchers’ knowledge and research to link keywords with categories. The former approach allows for discovery of new categories that researchers may not be aware of and avoids researchers’ biases, but also has the potential to result in categories that are less relevant or familiar to policymakers.

Our approach builds on and extends several recent papers that apply natural language processing to skills classification and other labor market questions. Several recent papers have applied machine learning approaches to labor market problems, including the classification of skills in job postings (Ao et al. 2023; Djumalieva et al. 2018; Gallagher et al. 2022; Lassébie et al. 2021; Shehu and Gjika 2024). Gallagher et al. (2022), which is the most similar to our efforts, aims to extract skill names from online job postings and generate hierarchical skill clusters based on embedded skill sentences.²⁵ Our methodology takes a similar approach. However, our study advances this and other previous projects in two significant ways. First, we incorporate domain-specific text (that is, job postings) in the language model training process to enhance the accuracy of identification of skills embeddings and cluster identification. Second, rather than manually assigning names to clusters, we use a

²⁴ A more detailed description of this methodology is included in the companion paper Dandanayak et al. (2024).

²⁵ See https://github.com/nestauk/ojd_daps_skills.

generative pre-trained transformer (OpenAI's GPT-3.5) to automatically generate interpretable labels.

Our methodology for generating the skills taxonomy from job postings and skill descriptions data involves four steps. We begin by implementing an automated procedure for translating job postings at scale from Spanish into English. Many postings contained irrelevant HTML elements and stop words like "the" and "is," so we also purged each posting of these. Due to computational constraints, we translate and process a (representative) sample of 500,000 job postings for use in most of the subsequent analysis. Once cleaned, we undertake the following four steps using the BERTopic package²⁶ for steps two through four (**Figure 7**).

- 1) *Improve the language model used to identify skills.* BERT (Bidirectional Encoder Representations from Transformers), introduced by Google in 2018, is a pre-trained language model designed to understand the context of words in a sentence bidirectionally by looking at the surrounding text. BERT can capture complex dependencies in language, making it highly suitable for generating skills embeddings (skills represented in vector form), a task in which understanding nuanced relationships between terms is crucial. However, domain-specific pre-training of Large Language Models (LLMs) has been shown to be beneficial particularly for tasks requiring precise handling of expert jargon (e.g., medical terms) that is not commonly included in the default corpuses on which LLMs are pre-trained (Gu et al. 2021; Ji et al. 2023). Thus, we extend the pre-training of a BERT variant (specifically, DistilRoBERTa²⁷) using the text of the job postings from Argentina, Chile, and Uruguay (this is called continued pre-training). This new model is then used in the second step below to generate skill embeddings.
- 2) *Generate skill embeddings.* After improving the language model, the next step is to identify skill embeddings, that is, skills represented in vector form. The BERTopic package, enhanced with the pre-training on the text of the job postings, is used to generate the skill embeddings. These embeddings are generated from the first sentence of text descriptions of the unique skills in the postings data that are provided by Lightcast. These text descriptions provide much more detail than the single word or phrase skill names. Stop words and common but uninformative keywords (for example, "professional" and "certificate") are eliminated from the text descriptions.
- 3) *Identify skills clusters.* The next step is to cluster the skill embeddings (skill vectors) to obtain meaningful skill categories. To do so, we use HDBSCAN (Hierarchical Density Based Spatial Clustering of Applications with Noise), a nonparametric clustering algorithm that identifies clusters in noisy data by isolating groups of points that are mutually close to each another.²⁸

²⁶ Available at <https://pypi.org/project/bertopic/>. Grootendorst (2019) provides a comprehensive overview of BERTopic, a topic modeling technique with an associated implementation package that integrates BERT (Bidirectional Encoder Representations from Transformers) and TF-IDF as its core methodological components.

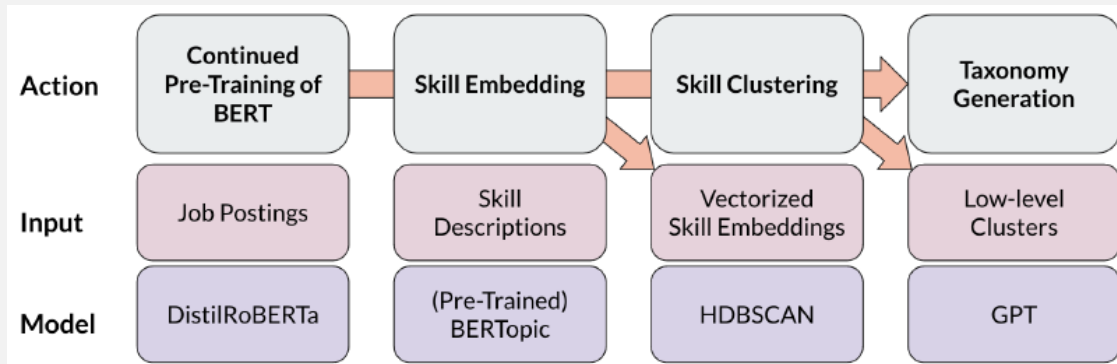
²⁷ Available at <https://huggingface.co/distilbert/distilroberta-base>. DistilRoBERTa, based on DistilBERT, which was introduced by Hugging Face in 2019, is a smaller and more efficient version of BERT, making programs that use it faster and more reproducible due to its lighter model size with insignificant loss of performance for computationally lighter tasks.

²⁸ At a high level, the algorithm works as follows. First, the algorithm defines a mutual reachability distance between vectors that is larger for either pairs of distant vectors or for vectors in non-dense areas. Based on this mutual reachability distance, the algorithm then constructs a minimum-distance spanning tree of the dataset

We prefer the nonparametric HDBSCAN method over parametric methods such as k-means clustering because it performs better in noisy settings with many outliers, which the skills dataset exemplifies. Additionally, HDBSCAN is able to identify clusters in the data organically without being forced to create an arbitrary number of clusters (as k-means does, for example). Finally, the implementation of HDBSCAN is particularly useful for hierarchical clustering, which is needed for the classification of skills. The BERTopic package allows for integration of HDBSCAN clustering with our trained embedding model to produce clusters.

- 4) *Generate the taxonomy.* Generating the skills taxonomy consists of two main steps. First, we construct sensible names for the clustered skills to form our taxonomy’s lowest-level (most granular) skill categories. This was achieved through topic modeling in the BERTopic package together with GPT-3.5’s word representations.²⁹ As part of the clustering process, the BERTopic package produces a collection of words that define each cluster based on the common semantic meaning of the words contained in the cluster. By feeding these topic words and skill descriptions (the same as those used in the Step 2 above) into GPT 3.5, we obtain interpretable names for each skill cluster obtained from HDBSCAN. Second, we organize these lowest-level skill categories into a hierarchical taxonomy, which is where the hierarchical component of HDBSCAN is useful. HDBSCAN is able to automatically generate this hierarchy by virtue of its clustering method, which allows us to identify which clusters are closest to each other and iteratively build the hierarchy. The BERTopic package supports hierarchical topic modeling for these hierarchical clusters, which allows us to build a complete taxonomy with human-interpretable terms for the resulting higher-level (less granular) categories.

Figure 7: Overview of skills classification pipeline



The resulting skills taxonomy has 50 categories in its most granular classification with three additional levels of categories. The raw hierarchical taxonomy is complex with a large number of

to identify the smallest distances needed to connect vectors to find points that naturally group together. Next, based on a predetermined “minimum cluster size” HDBSCAN traverses down the tree and sequentially splits the data into smaller and smaller clusters until the minimum size is reached. The algorithm then determines which of the resulting split clusters are robust enough to be extracted as standalone clusters and which ought to be recombined into larger clusters. Points that are too far away from a robust cluster to be assigned cluster membership in this step are determined to be outliers. We employed a specific form of HDBSCAN called soft clustering that, in addition to producing clusters, assigns a vector of probabilities to every point where each component denotes the probability that the point belongs to the component’s corresponding cluster. We used this to deal with outlier points, as outlined in the next section.

²⁹ See https://maartengr.github.io/BERTopic/getting_started/representation/llm.html.

intermediate layers. To resolve this, we manually reorganize the hierarchy into a 4-layer categorization, which maintains the hierarchical relationships established by our clustering while flattening and assigning more interpretable names to the levels. The final taxonomy is divided between digital and non-digital skills; then into human-oriented, technical, software engineering and development, and data science subcategories; and, finally, into 12 additional subcategories and the 50 low-level (most granular) categories (**Table 11**). Two important points stand out. First, the highest-level division is between digital and nondigital skills, indicating that digital skills represent a highly divergent category of skills that cannot be closely approximated by any non-digital skill group. Second, within the non-digital skills the taxonomy distinguishes between “human-oriented” or soft skills and technical skills. Finally, data science skills, which encompass statistics and its mathematical applications, are categorized differently from other digital skills (that is, skills related to software engineering and development), suggesting that these skills are distinct from other software skills.

Table 11: Skills categories resulting from the machine learning classification approach

Level 1	Level 2	Level 3	Level 4	
Digital (59%)	Data Science (40%)	Statistics	Mathematical Problem Solving & Optimization	
			Statistical Methods for Data Analysis & Predictive Modeling	
		Data Engineering	Database Management Systems & Query Tools for Data Handling	
			Data Analysis & Visualization in Business Intelligence Software	
	Software Engineering & Development (19%)	Digital Creation	Search Engine Optimization & Website Searching	
			Digital Content Management Platforms	
			Multimedia Production Skills	
		Cyber Security	Real-time Communication & Collaboration Platforms	
			Identity Access Management & Security Solutions	
			Cybersecurity Measures for Protecting Information Systems	
			Network Communication Technologies & Protocols	
		IT Development	IT Development	Storage Devices & Data Management
				Operating Systems - Linux, Mac OS X, Windows Mobile
				Virtualization Technology Comparison in Operating Systems
	Cloud Computing Platforms & Specialized Skills			
	Web Application Development Frameworks & Languages			
	Non-Digital (41%)	Technical (31%)	Medical	Software Testing Automation Framework
				Respiratory Diseases & Therapies
Cardiovascular Health & Medical Procedures				
Urinary System Surgical Procedures				
Medical Imaging Techniques & Radiology in Diagnosing & Treating				
Neurological assessment & treatment of related conditions				
Preventing & Managing Infectious Diseases				
Chemical & Life Sciences			Chemical & Life Sciences	Molecular Biology Techniques for Genetic Analysis
				Chemistry & Its Specialized Fields
				Soil & Water Interaction Studies in Geotechnical Engineering
		Ecology of Living Organisms & Their Interactions with the Environment		
Mechanical		Mechanical	Industrial Automation & Programmable Logic Controllers	
			Electrical Power Systems Design & Maintenance	
			Equipment & Machinery Repair & Maintenance Skills in Industry	
			Oil & Gas Drilling Specialized Skills	
			Metalworking Processes & Specialized Skills	
			CAD Software for Modeling & Drawing in Engineering	
			Language Abilities & Communication Skills	
	Human-Oriented (10%)		Language	Specialized Writing Skills in Journalism & Media Coverage
Marketing Strategies for Various Products & Services				
Specialized skills in Education & Learning				
Improving Business Efficiency & Productivity Through Enhanced Processes & Task Management				

	Environmental Performance Evaluation & Compliance Management
	Traffic & Transportation Systems Management
Safety & Wellness	Workplace Safety & Hazard Management Frameworks
	Specialized Healthcare Skills for Patient Care
	Mental Health Interventions & Specialized Care Focus
	Fitness & Exercise Science Specializations
Supply Chain Management	Enterprise Resource Planning Software Solutions for Business
	Inventory Management in Supply Chain & Procurement
	Tax Policy & Compliance Analysis for Businesses & Individuals
Financial & Legal Analysis	Financial Accounting & Management for Analysis & Reporting
	Legal Regulations in Information Technology & Case Law
	Insurance Policies & Legal Aspects

Note: The percentages in parentheses are the share of unique skills in a given category.

In sum, the machine learning-based skills classification methodology succeeds in producing a sensible classification of skills into a skills taxonomy. The categories are well formed and group together skills that seem similar. This is encouraging evidence that skills can be extracted from job postings data and categorized using an unsupervised and nonparametric model. Nevertheless, our approach has some drawbacks that future researchers will have to address. In particular, there are a large number of outliers in the initial clustering step: nearly half (47 percent) of the unique skills are not classified. We took steps to address this, but no method provided significant confidence that the outliers were being categorized correctly. The outliers might indicate skills that are either caught between skill categories or skills that are highly divergent and belong to no single category. Our current methodology is not capable of identifying and differentiating between these cases, which is a task for future research.

4.1.1.3 Comparing the two methodologies

The two methodologies result in very different skills taxonomies. To compare the two methodologies, we calculate a mutual information score, which measures how much one can learn about one variable (in our case a manually created skills category) based on knowing a second variable (in our case a machine learning-generated skills category). The mutual information score between the two lowest-level skills taxonomy categories is very low, meaning that the taxonomies are quite different.³⁰

The differences in the taxonomies are related to differences in approach, rather than problems inherent in either one. The categories in the manual classification aim for generality and breadth of coverage with the broad categories having little overlap and aiming to cover as many different conceivable skills as possible (for example, through categories such as “Business” and “Education and Training”). However, the taxonomy based on the machine learning approach emerges from the skills that are represented in the dataset, leading it to focus on finer distinctions between highly represented skills and ignore skills with less representation (and so omit related skill categories from the taxonomy). For example, the manually created taxonomy has a single category to represent “Information Technology” whereas the machine learning-generated taxonomy has more than 10, which occurs because software skills are highly represented in the dataset. In contrast, the machine learning taxonomy struggles to classify certain types of (non-digital) skills such as socioemotional skills. For example, 44 percent of skills classified as “social” in the manual taxonomy are not classified in the machine learning-based taxonomy.

These differences suggest different potential use cases. The manually created taxonomy has the advantage of being rooted in research on skills, achieves comprehensiveness across skill categories

³⁰ **Appendix 4** provides more details about calculating the mutual information score.

even if certain skills are underrepresented in the postings data, and is developed with end users (in our case, policymakers and labor market actors) in mind. The machine learning-generated taxonomy, in contrast, is sensitive to sampling bias in the skills in the postings data. However, this taxonomy is grounded in currently relevant skills, which is valuable for keeping labor market information up to date with new skills (and skills keywords) that may be emerging in areas of demand. Indeed, technological progress is likely to result in the creation of new types of jobs and tasks that require new skills that researchers creating skills taxonomies may not be aware of ex ante. The machine learning approach could thus be used to prompt researchers to add new terms to the manual taxonomy. At the same time, the machine learning approach itself can be improved by experimenting with alternative approaches to incorporating the range of skills included in the manually created taxonomy.

4.1.2 Results

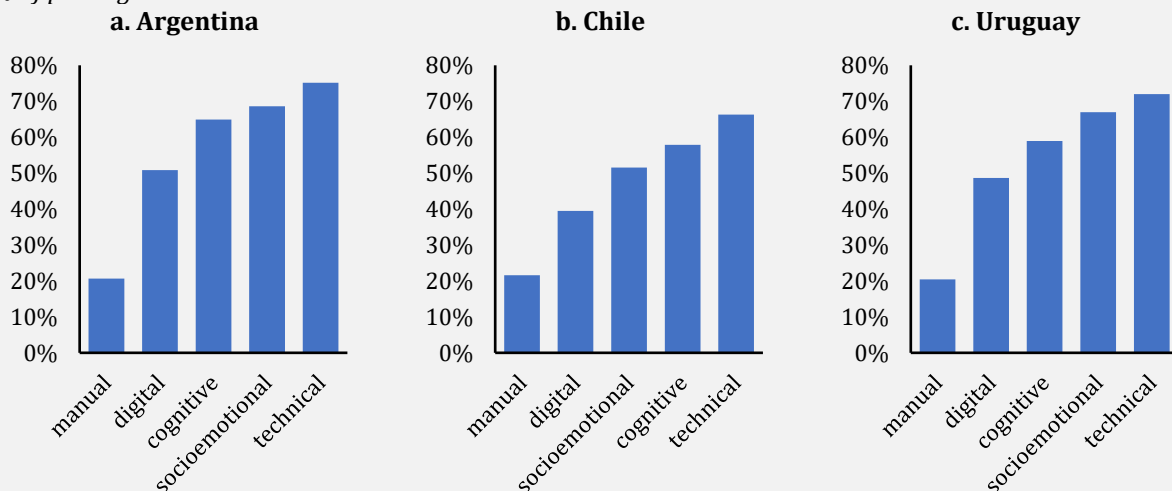
Once classified, the demand for each skill can be measured taking into account the frequency with which each skill appears across the job postings in each country. Given the experimental nature of the machine learning approach, the manually generated skills taxonomy is used to present results on skills demand. The demand for each skill is measured using relative frequency, specifically the ratio of the number of job postings that mention a skill to the total number of job postings.

4.1.2.1 Overall demand for skills by category and subcategory

More than half of vacancies in Argentina, Chile, and Uruguay request cognitive, socioemotional, and technical skills (**Figure 8**). Digital skills are frequently requested, but less so than these other three categories. Manual skills are less commonly requested. In Argentina, 75 percent of job postings require at least one skill classified as technical, followed by socioemotional (69 percent), cognitive (65 percent), digital (51 percent) and manual (21 percent) skills. In Uruguay, the pattern is very similar. The pattern is also similar in Chile, though cognitive skills are more frequently requested than socioemotional skills and all skills other than manual ones are less frequently requested. The median number of skills requested per vacancy is 6 in Argentina, 3 in Chile, and 4 in Uruguay.

Figure 8: Skill categories in demand in Argentina, Chile, and Uruguay, 2021-23

% of postings



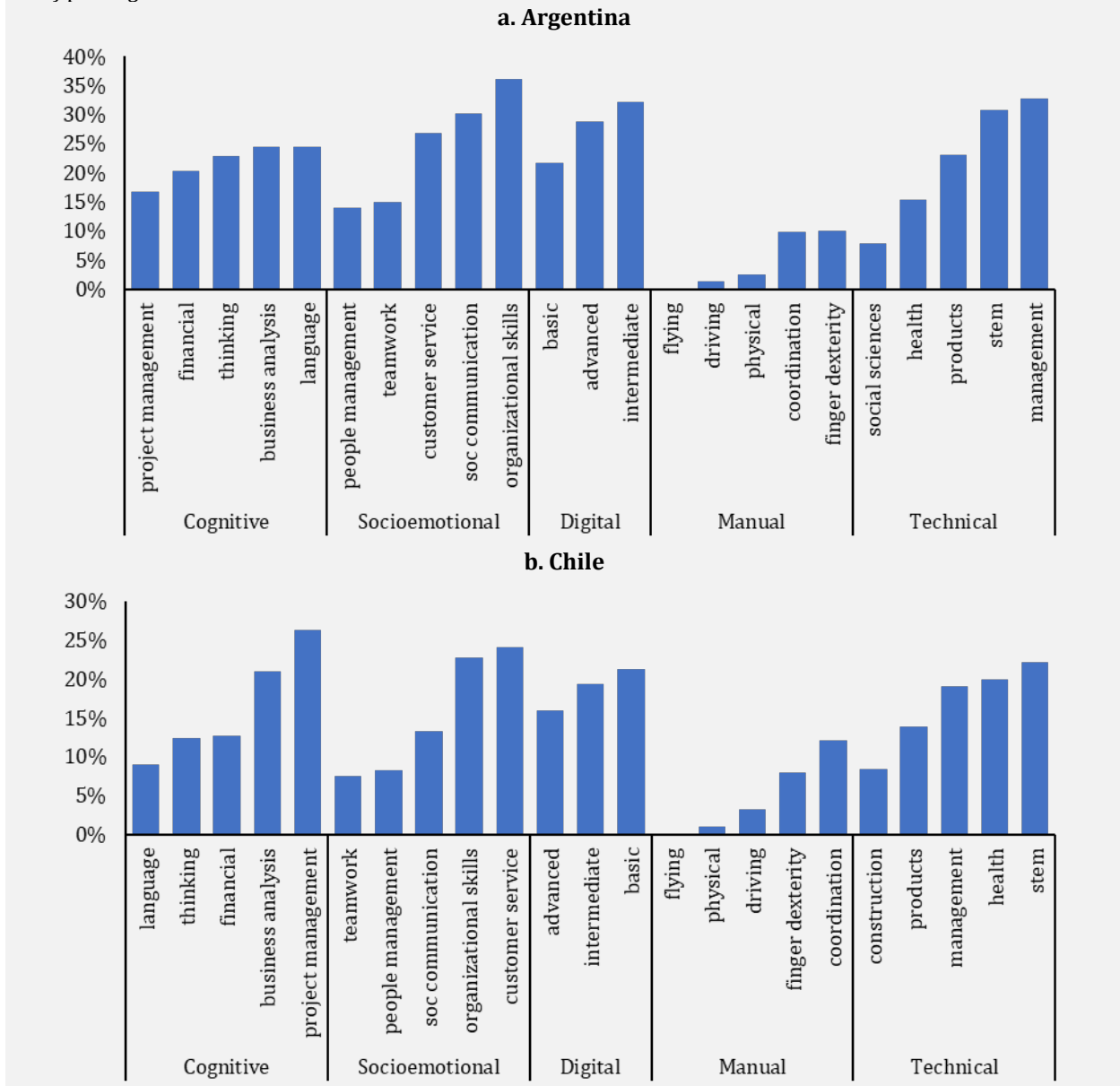
Note: The period is January 2021 to December 2023.

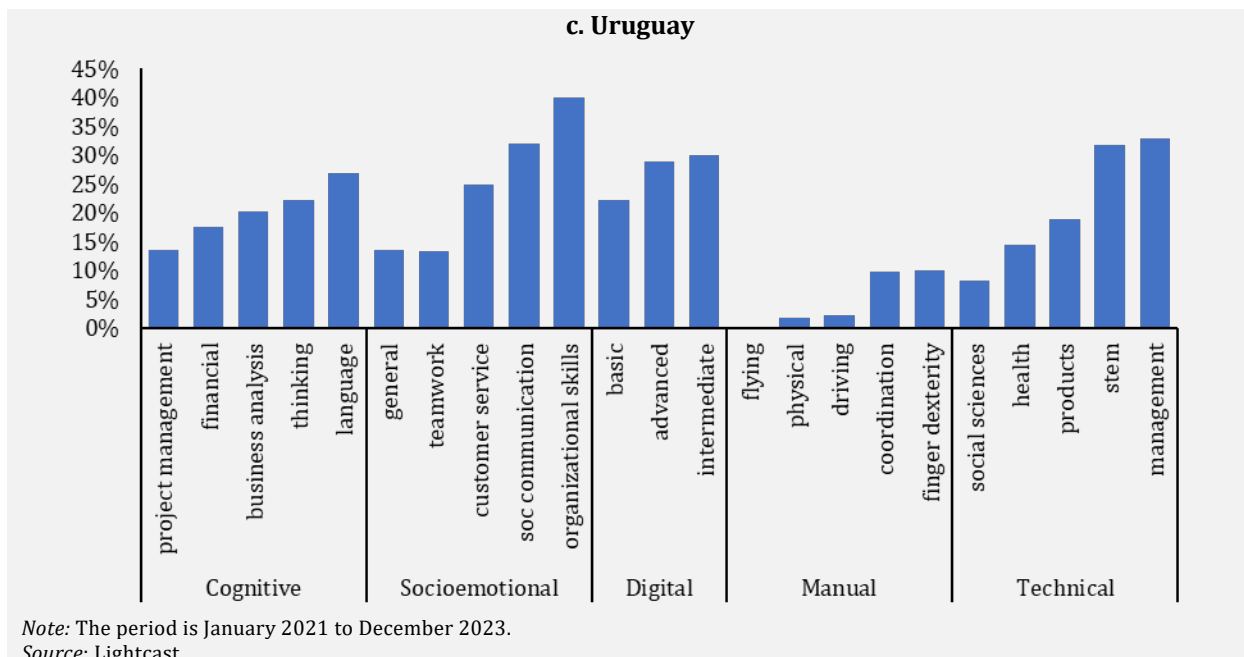
Source: Lightcast.

The kinds of skills in demand are quite similar across the three countries. **Figure 9** provides a comprehensive overview of the most sought-after subcategories of skills.

- *Cognitive skills.* Language skills are in highest demand in Argentina and Uruguay, appearing in a quarter of all job postings. In Chile, in contrast, project management skills are the most valued, appearing in a quarter of all postings. Business analysis skills are highly demanded across all three countries, appearing in more than a fifth of all postings.
- *Socioemotional skills.* Organizational and communication skills are among the top two skills in both Argentina and Uruguay. As is the case for cognitive skills, the demand for socioemotional skills is different in Chile, where customer service is the most demanded skill and communication is relegated to third place.
- *Digital skills.* In both Argentina and Uruguay, intermediate digital skills are in highest demand while in Chile basic digital skills are in highest demand.
- *Manual skill.* Manual skills are less in demand overall. Finger dexterity and coordination are the most in demand across all three countries.
- *Technical skills.* The top technical skills show more variation across countries. STEM skills are in demand in all three countries, while management skills are important in Argentina and Uruguay and health skills in Chile.

Figure 9: Top skills by skill category in Argentina, Chile, and Uruguay, 2021-23
% of postings





4.1.2.2 Coappearance of skills

Certain types of skills tend to appear together in job postings suggesting complementarity between skillsets. There is evidence of growing complementarity between social and cognitive skills (Deming and Kahn 2018; Weinberger 2014) and of complementarity between digital and other skills including cognitive and socioemotional skills, including in Latin America (OECD 2016, 2019; Grundke et al. 2018; Deming and Kahn 2018; Cunningham et al. 2022). The postings data provides evidence of complementarity (**Table 12**). Cognitive skills have the strongest correlations with all other categories while manual skills have the weakest. In Argentina, there is a moderate correlation between cognitive and digital skills (0.22), and between cognitive and socioemotional skills (0.20). When digital skills are disaggregated by level of expertise, cognitive skills appear to be more associated with higher levels of digital skills (0.18 for both intermediate and advanced digital skills). **Table 17** in **Appendix 5** provides the correlations with the detailed skills levels. In Uruguay, the correlation between cognitive skills and digital skills is substantially higher (0.2/0.3) and is also higher with socioemotional (0.26) and technical (0.25) skills. Stronger correlations are also seen between technical skills and digital (0.13/0.16) and socioemotional (0.25) skills. The strongest correlation is with intermediate digital skills in the case of cognitive skills, with basic digital skills in the case of socioemotional skills, and with intermediate digital skills in the case of technical skills. At the occupational level, complementarity appears to be stronger between certain subcategories and also between a particular subcategory and another type of category. **Table 18** and **Table 19** in **Appendix 5** provide the correlations for Software Developers and Administrative and Executive Secretaries, respectively, as an example.³¹

³¹ For Software Developers in Argentina, thinking skills (cognitive category) have a high correlation with other cognitive subcategories such as business analysis (0.17), and with other skill categories such as social skills (0.19) and technical skills (0.17). In Uruguay, the correlation between thinking and other cognitive subcategories applies to language (0.11) and data analysis (0.10), but the correlation between thinking and social skills is even higher (0.27). In addition, cognitive skills are highly correlated with organizational skills (0.13 in Argentina and 0.24 in Uruguay) and intermediate digital skills (0.16 in Argentina and 0.26 in Uruguay).

Table 12: Correlations of skill requirements, 2020-23*Bivariate correlations*

a. Argentina					
Variables	(1)	(2)	(3)	(4)	(5)
(1) cognitive	1.000				
(2) digital	0.219	1.000			
(3) socioemotional	0.200	0.079	1.000		
(4) manual	0.014	0.006	0.047	1.000	
(5) technical	0.140	0.124	0.062	0.059	1.000

b. Chile					
Variables	(1)	(2)	(3)	(4)	(5)
(1) cognitive	1.000				
(2) digital	0.123	1.000			
(3) socioemotional	0.146	0.130	1.000		
(4) manual	-0.026	-0.061	0.010	1.000	
(5) technical	0.076	0.043	0.000	-0.001	1.000

c. Uruguay					
Variables	(1)	(2)	(3)	(4)	(5)
(1) cognitive	1.000				
(2) digital	0.401	1.000			
(3) socioemotional	0.260	0.141	1.000		
(4) manual	0.054	0.071	0.105	1.000	
(5) technical	0.252	0.199	0.248	0.084	1.000

Note: Bivariate correlations across all skill categories and digital subcategories at the firm level. Only firms with non-missing ID and with more than 10 postings are included.

Source: Lightcast.

4.1.2.3 Skills profiles by occupation

The skills taxonomy can be used to create skills profiles at the occupation level. As an example, for each country we select the four occupations with the highest share of job postings to represent occupations in high demand. We then present the top five skills requested in each occupation (**Figure 10**). This demonstrates the kind of detailed skills information that can be generated for specific occupations of interest. Comparing across the occupations within countries also reveals interesting patterns. Occupations like software developers, commercial sales representatives, shop sales assistants, and contact center information clerks have one or two dominant skills that are requested across most (more than 70 percent) of job postings. The other in-demand occupations – management and organization analysts and administrative and executive secretaries – require a more diverse set of skills that varies from country to country. An addendum to this paper includes the top skills requirements for every occupation in Argentina, Chile, and Uruguay. Finally, we include the skills requirements in the United States for each of the in-demand occupations to benchmark each country’s requirements against those of an advanced economy (**Figure 10d**). Several interesting

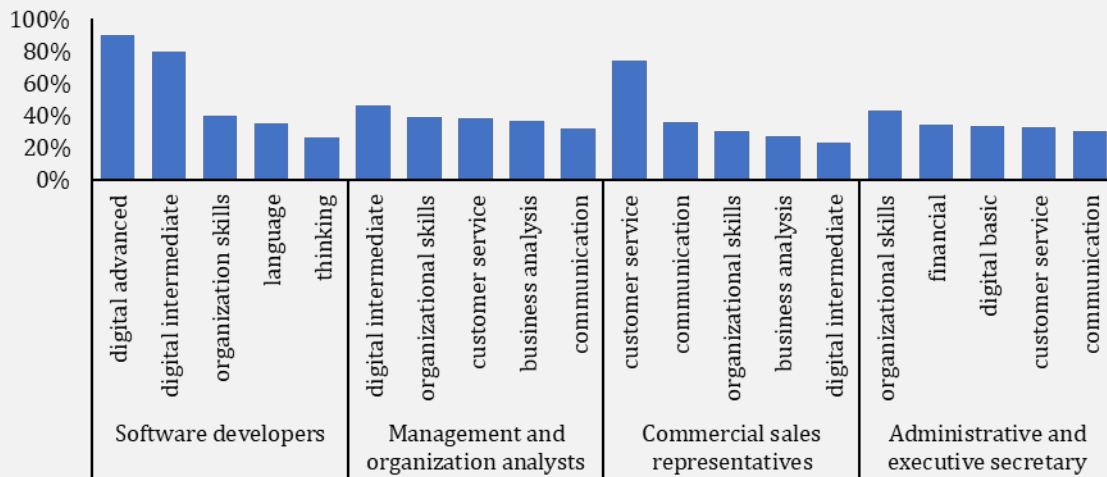
For Administrative and Executive Secretaries, the pairwise correlations between thinking and other cognitive subcategories are stronger in the case of language skills. Cognitive skills are also strongly correlated with organizational skills (social skills category) and basic digital skills.

patterns stand out. For example, communication skills are among the top 5 skills required in the United States across each of the in-demand occupations while communication skills are much less frequently required in Argentina, Chile, and Uruguay. On the other hand, advanced digital skills are the top skill for software developers in the United States as in Argentina and Uruguay.

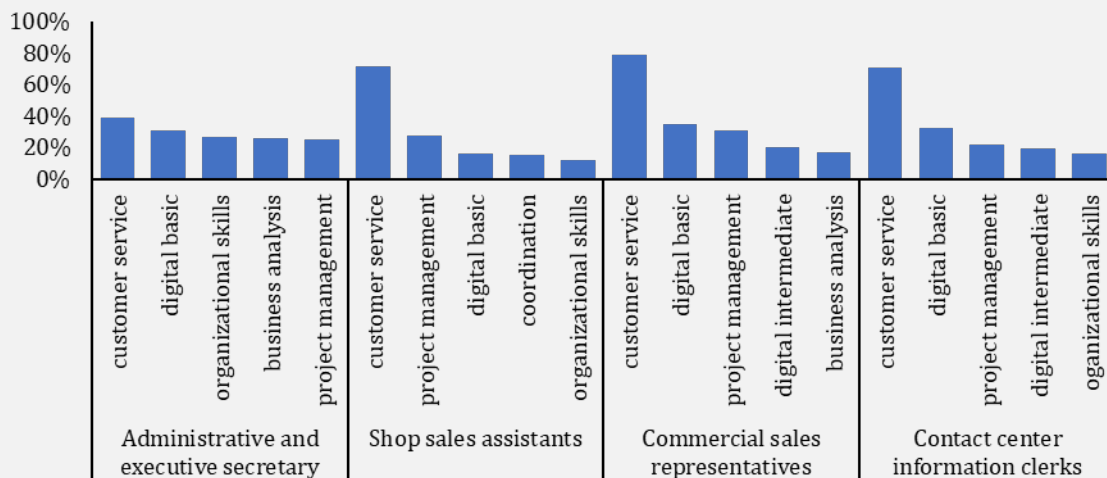
Figure 10: Top skills requested for the four most in-demand occupations in Argentina, Chile, and Uruguay with the United States as a Comparator, 2021-23

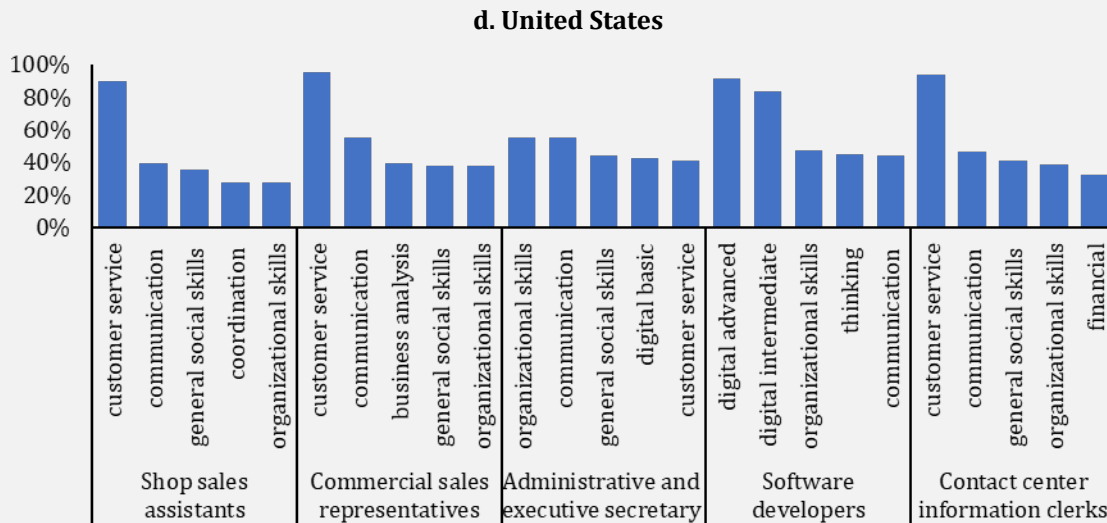
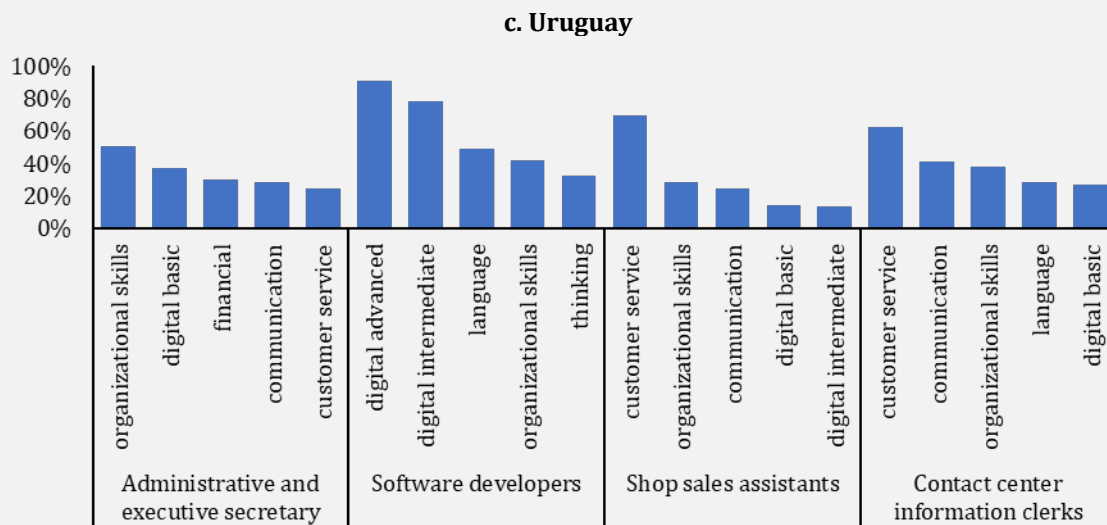
% of postings

a. Argentina



b. Chile





Note: The period is January 2021 to December 2023 for Argentina, Chile, and the United States. Given the much larger number of postings collected, the period is December 2023 for the United States.
Source: Lightcast.

4.2 Use case #2: Identifying similar occupations

Understanding the similarity between jobs is an essential component of labor market information systems. Identifying similar jobs allows employment services agencies to inform displaced jobseekers of potential new job opportunities based on their skills and to advise them of what training may be necessary to fill gaps in these skills. Understanding such similarities is important to manage the risk of both idiosyncratic shocks like one-off job loss and systemic shocks like climate change and the transition to a low-carbon economy.

4.2.1 Methodology

Two approaches to identifying occupational similarity are undertaken. The first approach uses an established metric, relative comparative advantage, to assess similarity based on the skills demanded

in each occupation. The second, experimental approach uses semantic similarity in job postings to assess their similarity.

4.2.1.1 Relative comparative advantage

Occupational similarity is assessed by comparing the skills required to perform different jobs. The methodology follows Apella and Zunino (2022b) in using a procedure that constructs an “occupation space,” which is an adaptation of the methodology proposed by Hidalgo and Hausmann (2009) to build a product space for goods and services. The adaptation of the product space methodology draws on the concept of relative comparative advantage (RCA).³² We assume that a particular skill exhibits a revealed comparative advantage for an occupation when the average requirement of that skill within the occupation, measured as the relative frequency with which that skill appears in occupation, is higher than the average requirement across all occupations. The specialization of occupations according to their skills – their RCA – is defined by constructing a binary matrix with the number of rows equal to the number of occupations (O) and the number of columns equal to the number of skills (H). From the binary matrix of revealed comparative advantages, the frequency with which two occupations jointly exhibit RCA is calculated. That is, the number of skills whose intensity of use is above average in both occupations is calculated. This allows for the identification of occupations that might require the same capabilities.

Results are represented in an occupation map. The similarity scores between the occupations are represented in a matrix in the occupation space (with occupations in both rows and columns) where each element indicates the frequency with which two occupations intensively use skills simultaneously relative to the frequency with which the skill is used intensively across occupations. An occupation space map is created by extracting the most significant connections that keep all the nodes in the graph connected.³³ Once the proximity of occupations is identified and represented in this occupation map, additional information such as the characteristics of occupations can be incorporated. In the occupation map, this second dimension is visualized using colors.

4.2.1.2 Semantic similarity³⁴

We also undertake an experimental approach to measuring occupational similarity by comparing the text of jobs postings across occupations. Recent literature has explored the potential to use embeddings and other data-driven methods to compare jobs with some evidence suggesting that these approaches could be superior to methods relying on the frequency with which keywords appear (Cao et al. 2021; Seegmiller, Papanikolaou, and Schmidt 2023; and Zhu, Javed, and Ozturk 2016). This may be the result of being able to compare similar words and phrases that are not exactly the same. For this method, we assume that occupations that are similar also have postings that are

³² Balassa (1977) asserts that comparative advantage is revealed through observed trade patterns, specifically high shares of export markets. In our case, comparative advantages are determined by observing the labor market. Previous research on classifying occupations based on skills has overwhelmingly used revealed comparative advantage (Alabdulkareem et al. 2018; Dawson et al. 2021; José-García et al. 2023). See **Appendix 6** for a more detailed explanation of the methodology.

³³ This is done using a Maximum Spanning Tree algorithm as in Hidalgo et al. (2007). A spanning tree refers to the different subsets of a graph (in this case, the occupation space) that include all attributes (occupations) with the minimum number of edges. When the edges have weights (in this case, given by the distances between occupations), it is possible to obtain the Maximum Spanning Tree, which is the spanning tree representation that maximizes the sum of the weights of the edges.

³⁴ A more detailed description of this methodology is included in the companion paper Dandanayak et al. (2024).

semantically similar. We use text from a subset of the sample of postings created for the skills classification task. We generate vector embeddings (that is, number representations of the job postings text) for each of the postings' job descriptions using an embeddings model from OpenAI.³⁵ We then aggregate the embeddings by occupation to obtain an "average" embedding for each. We use cosine similarity on these aggregated vectors to compare how similar one occupation's average posting description is to another. This approach results in ranked lists of occupational similarity that can be mapped in a manner similar to the RCA approach.

4.2.1.3 Comparing the two methodologies

The lists generated by the experimental approach and by the RCA approach (applied to the same subset of postings) are quite different. The two lists exhibit only 30 percent overlap for most occupations.³⁶ A low level of agreement between the two systems is somewhat surprising. However, it may indicate that the semantic approach is picking up similarities in how postings are written rather than similarities in what is being described. Additionally, there is evidence that the differences are driven largely by "remainder" occupations that aggregate jobs "not elsewhere classified" within the ISCO classification scheme. Given the lack of theoretical underpinnings for the semantic similarity approach, in contrast to the skills-based RCA approach, we set aside this method as an area for further research and present the results based on the RCA analysis.

4.2.2 Results

This section presents the occupation maps constructed for each country. In all cases, each node in the graph represents an occupation and the distance between the nodes indicates the proximity of the occupations in terms of the skills required for their performance. We look first at the distance between the occupations alone and then at two different occupational characteristics: labor market demand proxied by the number of vacancies and an occupation's greenness. Note that the proximity results are the same across the two representations for each country so the shape of the map and the distance between the different nodes are identical across the representations.

4.2.2.1 The similarity of occupations

Looking at the distance between the occupations in the occupation maps show several interesting results (**Figure 10**). First, in both countries there are areas of the map with high occupation density, indicating a significant core of commonly required skills. This is particularly the case in Uruguay where there is a very dense cluster of occupations, perhaps reflecting the smaller size of Uruguay's labor market. Conversely, each map includes some less dense areas of occupations with more specific skill profiles. The existence of high-density areas on the map is relevant for both youth employment policies and reskilling and upskilling policies for experienced workers. Given the high rates of unemployment among young people in Argentina and Uruguay, policies targeting young people's first employment experience may be better targeted by promoting placements in occupations located in dense areas of the map where the experience gained allows for the development of skills transferable to many occupations. Reskilling and upskilling policies may also be more effective when targeted to dense areas on the map.

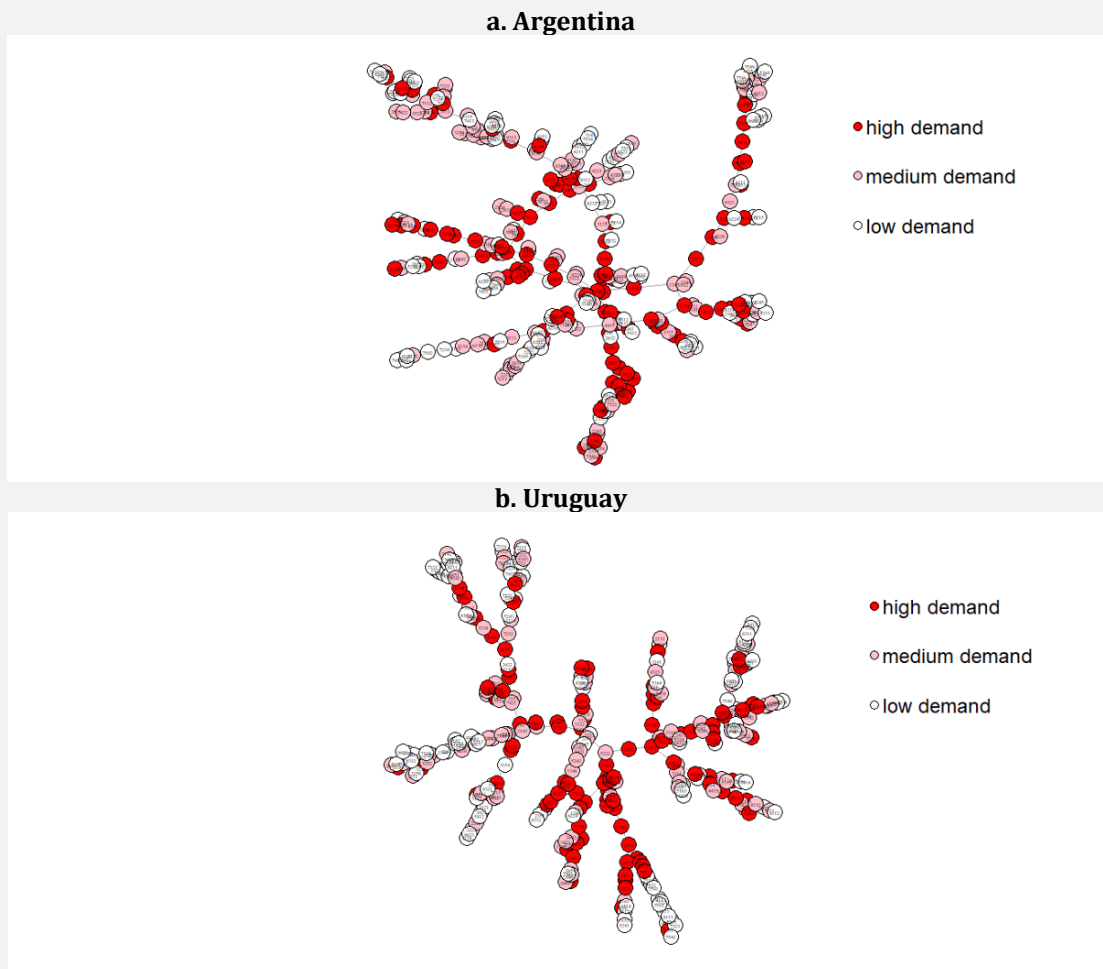
³⁵ We use the text-embedding-3-small model. See <https://platform.openai.com/docs/guides/embeddings>.

³⁶ We use Jaccard similarity to measure this overlap.

4.2.2.2 The similarity of low-, medium-, and high-demand occupations

High-, medium- and low-demand occupations cluster together, indicating that workers in low-demand occupations have options to switch to medium- and high-demand occupations with similar skillsets. The demand profile for each occupation is overlaid on each occupation map (**Figure 11**). The number of postings for each occupation are calculated for each country to proxy demand. The occupations are then divided into three major groups: i) the first third with the fewest postings, labeled as low-demand occupations; ii) the second third with fewer postings, labeled as medium-demand occupations; and iii) the third with the most postings, labeled as high-demand occupations. Notably, the different regions of the map do not show defined clusters related to the demand profile of occupations. That is, in various areas of the map occupations with high, medium, and low demand coexist. This result, which will change as occupational demand changes, is favorable for the design of successful upskilling and reskilling policies, as these should preferably be oriented towards occupations that are dynamic in terms of demand. The outlook for upskilling and reskilling policies would be more challenging if all low-demand occupations were close to each other and distant from medium- or high-demand occupations.

Figure 11: Occupational similarity and demand for occupations in Argentina and Uruguay



Note: Each node represents an occupation. The distance between the nodes indicates the proximity of the occupations in terms of the skills required.

Source: Lightcast.

4.2.2.3 The similarity of green occupations

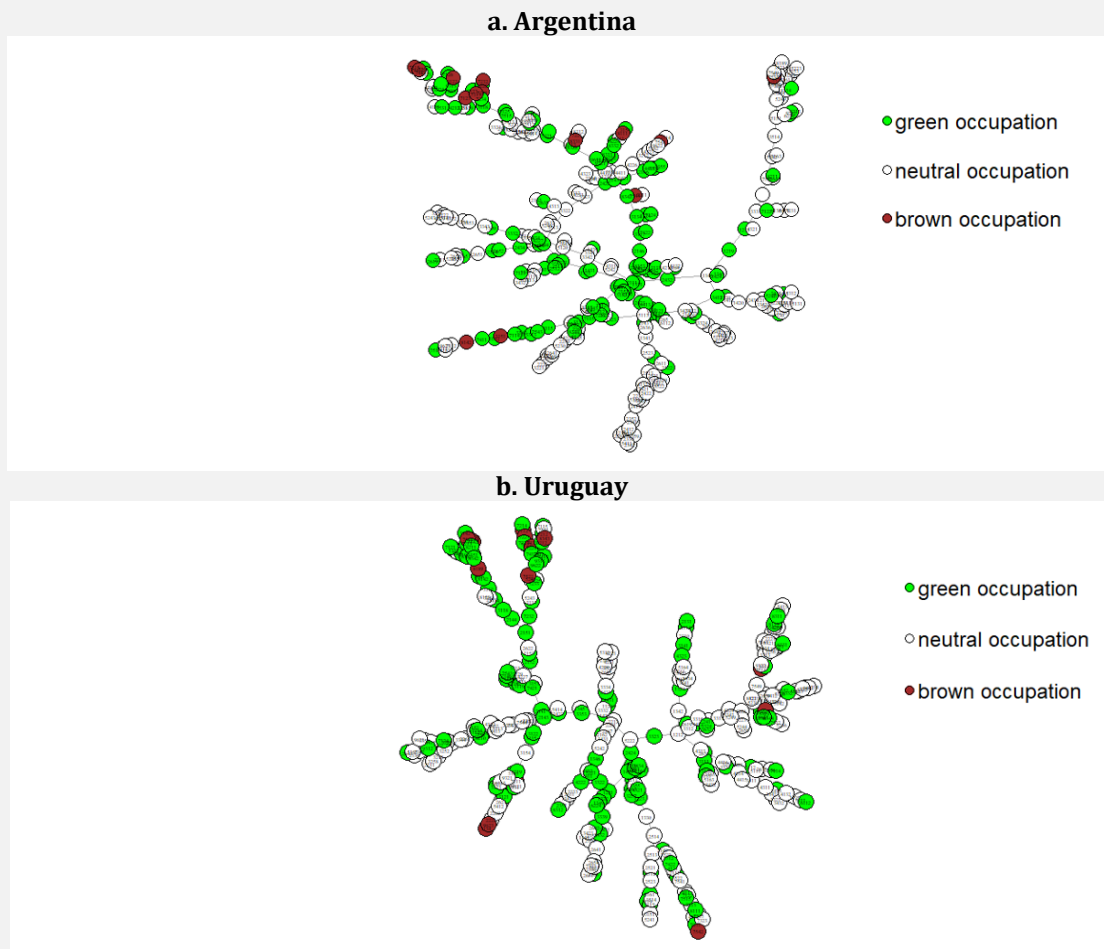
The third representation focuses on a sustainable environmental transition in the labor market by identifying green jobs and brown jobs. This approach follows Arakaki et al. (2022)'s background work for the World Bank's *Argentina Country Climate and Development Report* (World Bank 2022) in using the US O*NET's identification of green jobs and Vona et al. (2018)'s identification of brown jobs.³⁷ A correspondence table is used to match the green and brown occupations, categorized in the United States' Standard Occupational Classification (SOC), to the ISCO-08 classification scheme used for the job postings.³⁸

Similar to the previous representations, different regions of the map do not show defined clusters of green and brown occupations (**Figure 12**). That is, in various areas of the map, green, brown, and neutral (non-green, non-brown) occupations coexist. This result indicates that brown occupations, despite having the potential to decline during a transition to more environmentally sustainable activities, are related to green occupations in terms of the skills required for their performance. This is a favorable finding for designing successful upskilling and reskilling policies, as it provides greater opportunities for transitioning workers from brown jobs to green jobs.

³⁷ Green jobs involve activities such as reducing fossil fuel use, decreasing pollution, and increasing energy efficiency. Brown occupations are those most likely to be in pollution-intensive industries.

³⁸ Given that the correspondence between SOC and ISCO codes is not one-to-one, and SOC provides a more detailed classification (8 digits for green jobs and 6 digits for brown jobs), the classification of ISCO 4-digit occupations as green or brown is an approximation.

Figure 12: Occupational similarity and green and brown jobs in Argentina and Uruguay



Note: Each node represents an occupation. The distance between the nodes indicates the proximity of the occupations in terms of the skills required.

Source: Encuesta Permanente de Hogares (Argentina); Encuesta Continua de Hogares (Uruguay); Lightcast; US O*NET; Vona et al. (2018).

4.2.2.4 A real world application

To better understand the practical applications of the occupation maps, we demonstrate the case of shifting a worker out of a low-demand, brown occupation. In Argentina and Uruguay, Upholsterers and Related Workers are characterized by low demand and are classified as a brown occupation. These characteristics suggest that transitioning upholsterers to other occupations is both expected (given low demand) and desirable (given the environmental profile). The occupation map can be used to identify for each country the closest occupations in terms of skills demanded. In all three countries, it is possible to identify occupations with more than 50 percent similarity in skills but with a more favorable profile in terms of demand (medium or high demand) and greenness (a green or at least neutral profile) (**Table 13**). This tool also allows us to identify the skills necessary for reskilling: the skills that are needed in the occupation to which the worker wishes to transition but that are not needed in the occupation from which the worker is transitioning. To illustrate this use, **Table 14** shows the reskilling needs for Upholsterers and Related Workers to transition to high-demand, environmentally sustainable jobs: Assemblers Not Elsewhere Classified for Argentina and Air Conditioning and Refrigeration Mechanics for Uruguay. Technical skills are prominent among reskilling needs.

Table 13: The jobs most similar to Upholsterers and Related Workers in Argentina and Uruguay

a. Argentina				
Occupation	Proximity	Demand	Greenness	
Upholsterers and Related Workers	NA	Low	Brown	
Print Finishing and Binding Workers	0.62	Low	Neutral	
Assemblers Not Elsewhere Classified	0.60	High	Green	
Forestry Technicians	0.60	Low	Green	
Metal Polishers, Wheel Grinders and Tool Sharpeners	0.55	Low	Brown	

c. Uruguay				
Title	Proximity	Demand	Greenness	
Upholsterers and Related Workers	NA	Low	Brown	
Air conditioning and refrigeration mechanics	0.78	High	Green	
Blacksmiths, hammersmiths and forging press workers	0.67	Median	Brown	
Shotfirers and blasters	0.67	Low	Brown	
Meter readers and vending-machine collectors	0.67	Low	Brown	

Source: Lightcast.

Table 14: Reskilling needs for Upholsterers and Related Workers in Argentina and Uruguay

Argentina: Assemblers Not Elsewhere Classified	Uruguay: Air conditioning and refrigeration mechanics
<i>Cognitive: Adaptability</i>	<i>Cognitive: Business systems</i>
<i>Social: People Management</i>	<i>Technical: Energy</i>
<i>Technical: Management</i>	
<i>Technical: Social Studies</i>	

Source: Lightcast.

4.3 Use case #3: Using job vacancy postings as an indicator of labor market health

Job vacancy postings are a potential leading indicator of labor market health. As shown in Section 3, there is a significant correlation over time between the job postings and labor market indicators like the employment and unemployment rates. Because the data are collected and can be reported on a real time basis, unlike the sources of the other labor market indicators, this correlation suggests that the data could be used as a leading indicator of labor market performance.

The data currently available provide a first step in analyzing whether postings can be used in this way. At this stage, there are two important limitations to evaluating whether the data can be used as a leading indicator of employment or unemployment rates. First, the available time series of postings (about 3.5 years) is still somewhat short for constructing robust time series models, in particular to create models that are robust to seasonal fluctuations. Second, as noted in the previous section, there is a general increase in the number of postings over time in the three countries, which may be associated with improvements in the job posting collection process (for example, incorporating new sources) or even an increasing use of online posting. These phenomena are not necessarily related to actual employment or unemployment trends.

4.3.1 Methodology

Simple time series models are estimated to use job postings to predict future employment and unemployment rates. The models regress the unemployment and employment rates on the total number of monthly postings using the time lag that presents the best fit. Monthly data are used in all cases, so seasonal dummies are included as controls. The period considered spans from August 2020 to December 2023. The models are constructed considering stationary transformations of the included variables. An analysis of unit roots is first conducted on the series using the Augmented Dickey-Fuller (ADF) test. Finally, for the postings both the original series and a moving average transformation are used to smooth the significant volatility of the raw data. Thus, four models are tested: the association between the raw number of online job postings and the employment rate (Model 1) and the unemployment rate (Model 2) and the association between the moving average of online job posting and the employment rate (Model 3) and the unemployment rate (Model 4).

4.3.2 Results

The estimated models show encouraging results for using the job postings data as a leading indicator of labor market health, though the results are not yet sufficiently robust. The cases of Argentina and Uruguay are very similar (see **Table 15** in **Appendix 1**). In both countries in all of the estimated models the coefficients have the expected sign: the coefficient on the job postings variable is positive for employment rates and negative for unemployment rates. However, the estimated coefficients are not always significant. In neither country do the online job postings show a statistically significant correlation with the unemployment rate when both variables are considered in stationary transformations. In the case of the employment rate, no significant relationship is observed when working with the raw posting data, but there is a significant relationship when the smoothed series is included as a determinant. In this case, postings lagged by two periods appear as a significant variable for predicting the behavior of the employment rate. This result aligns with the average duration of postings in both countries, which is approximately 60 days. This finding shows significant potential for use as a leading indicator, as the posting information anticipates employment behavior by two months, while employment data are published with at least a one-month delay. This also shows that the job postings data can provide useful insight into labor markets generally even in places like Argentina where informality rates are high.

Conclusions

Job postings data can complement existing data sources about the labor market in Argentina and Uruguay. The job postings data for Argentina and Uruguay offers detailed information about labor market demand, including skills demand, on an ongoing basis, which fills a gap in the labor market information systems in both countries. Checks of the quality of the job postings data show that they are a quality source of information about the labor market. However, biases are present, particularly a bias towards high-skilled, likely formal jobs. Still, the volume of data available means that insights can still be made into areas where the postings data has less labor market coverage.

The postings data offer new analytical insights and open possibilities to improve policy in a range of labor market areas. The paper has shown the potential to put the job postings data to work to describe labor market health, identify skills demand, and identify similar occupations. The first use can help policymakers understand where labor markets are headed more quickly, perhaps two months in advance, of current survey data. The second two uses have implications for employment and training services by providing demand-based insight into which skills should be prioritized and

recommended to which types of workers. Online job postings offer these potential uses at a substantially lower cost than that required for labor force surveys.

The paper also offers a methodological contribution in outlining a machine learning pipeline for identifying and classifying skills without a predetermined skills taxonomy. While the machine learning-based methodology does not offer the comprehensive approach to skills identification and categorization that literature-based manual classification methods provide, the machine learning method excels in emphasizing in the taxonomy the skills that are in highest demand and in identifying in-demand skills that researchers may overlook. Using the machine learning pipeline in coordination with a manual classification could help keep the manual classification up to date with current labor market demand by uncovering new skills and skills keywords to incorporate. This is, in fact, the approach undertaken by the United States in maintaining its occupational database O*NET.

References

- Alabdulkareem, Ahmad, Morgan R. Frank, Lijun Sun, Bedoor AlShebli, César Hidalgo, and Iyad Rahwan. 2018. "Unpacking the Polarization of Workplace Skills." *Science Advances* 4(7).
- Albertini, Julien, Arthur Poirier, and Danilo R. Trupkin. 2019. "A Job Vacancy Rate for Argentina." Mimeo.
- Alekseeva, Liudmila, JoséAzar, Mireia Ginéa, Sampsa Samila, and Bledi Taska. 2021. "The Demand for AI Skills in the Labor Market." *Labour Economics* 71.
- Almlund, Mathilde, Angela Lee Duckworth, James Heckman, and Tim Kautz. 2011. "Personality Psychology and Economics." In *Handbook of the Economics of Education*, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 1–181. Amsterdam: North Holland.
- Ao, Ziqiao, Gergely Horváth, Chunuan Sheng, Yifan Song, and Yutong Sun. 2023. Skill Requirements in Job Advertisements: A Comparison of Skill-Categorization Methods based on wage Regressions." *Information Processing Management* 60 (2).
- Apella, Ignacio and Gonzalo Zunino. 2022. "El cambio tecnológico y las tendencias del mercado laboral en América Latina y el Caribe: un análisis basado en las tareas." *Revista de la CEPAL* 136:65-88.
- Apella, Ignacio and Gonzalo Zunino. 2022b. "Espacio de Ocupaciones en Uruguay. Un instrumento para apoyar el diseño de políticas de reconversión laboral." World Bank, Washington, D.C.
- Apella, Ignacio and Gonzalo Zunino. 2017. "Technological Change and the Labor Market in Argentina and Uruguay." Policy Research Working Paper 8215, World Bank, Washington, D.C.
- Apella, Ignacio, Rafael Rofman, and Helena Rovner. 2020. *Skills and the Labor Market in a New Era: Managing the Impacts of Population Aging and Technological Change in Uruguay*. Washington, D.C.: World Bank.
- Arakaki, Agustín, Mariana Conte Grand, Fabián González, Penny Mealy, Lourdes Rodríguez Chamussy, and Julie Rozenberg. 2022. "Background Note 8: Transition from Brown to Green Jobs: Its Potential Poverty and Distributional Impacts in Argentina." World Bank, Washington, D.C.
- Atalay, Enghin, Phai Phongthientham, Sebastian Sotelo, and Daniel Tannenbaum. 2020. "The Evolution of Work in the United States." *American Economic Journal: Applied Economics* 12(2):1-34.
- Autor, David. 2022. "The Labor Market Impacts of Technological Change: From Unbridled Enthusiasm to Qualified Optimism to Vast Uncertainty." Working Paper 30074, National Bureau of Economist Research, Cambridge.
- Avitabile, Ciro and Rafael de Hoyos. 2018. "The Heterogeneous Effect of Information on Student Performance: Evidence from a Randomized Control Trial in Mexico." *Journal of Development Studies* 135:318-48.

- Azar, José, Ioana Marinescu, Marshall Steinbaum, and Bledi Taska. 2020. "Concentration in US Labor Markets: Evidence from Online Vacancy Data." *Labour Economics* 66.
- Balassa, Bela. 1977. "Revealed Comparative Advantage Revisited: An Analysis of Relative Export Shares of the Industrial Countries, 1953-1971." *The Manchester School* 45(4):327-44.
- Beam, Emily A. 2016. "Do Job Fairs Matter? Experimental Evidence on the Impact of Job-Fair Attendance," *Journal of Development Economics* 120:32-40.
- Beam, Emily A., David Mckenzie, and Dean Yang. 2016. "Unilateral Facilitation Does Not Raise International Labor Migration from the Philippines," *Economic Development and Cultural Change* 64(2):323-68.
- Belot, Michèle, Philipp Kircher, and Paul Muller. 2022. "Do the Long-Term Unemployed Benefit from Automated Occupational Advice during Online Job Search?" Discussion Paper Series 15452, IZA, Bonn.
- Belot, Michèle, Philipp Kircher, and Paul Muller. 2019. "Providing Advice to Job Seekers at Low Cost: An Experimental Study on On-Line Advice." *The Review of Economic Studies* 86(4):1411-47.
- Bennett, Fidel, Verónica Escudero, Hannah Liepmann, and Ana Podjanin. 2022. "Using Online Vacancy and Job Applicants' Data to Study Skills Dynamics." Working Paper 75, ILO, Geneva.
- Borgonovi, Francesca, Flavio Calvino, Chiara Criscuolo, Julia Nania, Julia Nitschke, Layla O'Kane, Lea Samek, and Helke Seitz. 2023. "Emerging Trends in AI Skill Demand Across 14 OECD Countries." OECD Artificial Intelligence Papers No. 2, OECD, Paris.
- Brancatelli, Calogero, Alicia Marguerie, and Stefanie Brodmann. 2020. "Job Creation and Demand for Skills in Kosovo: What Can We Learn from Job Portal Data?" Policy Research Working Paper 9266, World Bank, Washington, D.C.
- Brown, Jennifer and David A. Matsa. 2020. "Locked by Leverage: Job Search During the Housing Crisis," *Journal of Financial Economics* 136(3):623-48.
- Brynjolfsson, Erik, Tom Mitchell, and Daniel Rock. 2023. "Quantifying the Distribution of Machine Learning's Impact on Work." Mimeo.
- Brynjolfsson, Erik, Danielle Li, and Lindsey R. Raymond. 2023. "Generative AI at Work." Working Paper No. 31161, National Bureau of Economic Research, Cambridge, MA.
- Cammeraat, Emile and Mariagrazia Squicciarini. 2021. "Burning Glass Technologies' Data Use in Policy-relevant Analysis: An Occupation-level Assessment." OECD Science, Technology and Industry Working Papers 2021/05, OECD, Paris.
- Cao, Lina, Jian Zhang, Xinquan Ge, and Jindong Chen. 2021. "Occupational Profiling Driven by Online Job Advertisements: Taking the Data Analysis and Processing Engineering Technicians as an Example." *PLoS One* 16(6):e0253308.

- Carnevale, Anthony P., Tamara Jayasundera, and Dimitri Repnikov. 2014. "Understanding Online Job Ads Data: A Technical Report." Georgetown University Center on Education and the Workforce, Washington, D.C.
- Carranza, Eliana and David McKenzie. 2024. "Job Training and Job Search Assistance Policies in Developing Countries." *Journal of Economic Perspectives* 38(1):221-44.
- Caunedo, J., E. Keller, and Y. Shin. 2021. "Technology and the Task Content of Jobs across the Development Spectrum." NBER Working Paper 28681. Cambridge, MA: National Bureau of Economic Research.
- CEDEFOP, European Commission, ETF, ILO, OECD, and UNESCO. 2021. "Perspectives on Policy and Practice: Tapping into the Potential of Big Data for Skills Policy." CEDEFOP, Luxembourg.
- Che, Natasha. 2021. "Dissecting Economic Growth in Uruguay." IMF Working Paper WP/21/2, International Monetary Fund, Washington, D.C.
- Conzelmann, Johnathan G., Steven W. Hemelt, Brad Hershbein, Shawn M. Martin, Andrew Simon, and Kevin M. Stange. 2023. "Skills, Majors, and Jobs: Does Higher Education Respond?" Working Paper 31572, National Bureau of Economics Research, Cambridge, MA.
- CSC (Critical Skills Monitoring Committee). 2019. "Critical Occupations List 2018/2019: Technical Report." CSC, Putrajaya.
- Cunningham, Wendy, Harry Moroz, Noël Muller, and Aivin Solatorio. 2022. "The Demand for Digital and Complementary Skills in Southeast Asia." World Bank Policy Research Working Paper, World Bank, Washington, D.C.
- Dammert, Ana C., Jose Galdo, and Virgilio Galdo. 2015. "Integrating Mobile Phone Technologies into Labor-Market Intermediation: A Multi-Treatment Experimental Design." *IZA Journal of Labor & Development* 4(11).
- Dawson, Nikolas, Mary-Anne Williams, and Marian-Andrei Rizoiu. 2021. "Skill-driven Recommendations for Job Transition Pathways." *PLoS ONE* 16(8).
- Del Carpio, Ximena, Olga Kupets, Noël Muller, and Anna Olefir. 2017. *Skills for a Modern Ukraine*. Washington D.C.: World Bank.
- DE4A (Digital Economy for Africa). 2021. *Digital Skills: The Why, the What, and the How - Methodological Guidebook for Preparing Digital Skills Country Action Plans for Higher Education and TVET (V2.0)*. Washington, DC: World Bank.
- Deming, David J. 2017. "The Growing Importance of Social Skills in the Labor Market." *Quarterly Journal of Economics* 132(4): 1593–1640.
- Deming, David J. and Kadeem Noray. 2020. "Earnings Dynamics, Changing Job Skills, and STEM Careers." *The Quarterly Journal of Economics* 135(4):1965-2005.
- Deming, David J. and Lisa B. Kahn. 2018. "Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals." *Journal of Labor Economics* 36(S1):S337-69.

- Dicarlo, Emanuele, Salvatore Lo Bello, Sebastian Monroy-Taborda, Ana Maria Oviedo, Maria Laura Sanchez-Puerta, and Indhira Santos. 2016. "The Skill Content of Occupations across Low and Middle Income Countries: Evidence from Harmonized Data." Discussion Paper 10224, IZA, Bonn.
- Di Ionno, Michelle and Michael Mandel. 2016. "Argentina: The Road to the App Economy." Progressive Policy Institute, Washington, D.C.
- Djumalieva, Jyldyz, Antonio Lima, and Cath Sleeman. 2018. "Classifying Occupations according to Their Skill Requirements in Job Advertisements." Discussion Paper 2018-04, Economic Statistics Centre of Excellence, National Institute of Economic and Social Research, London.
- Dandanayak, Utkarsh, Kiran Duggirala, Ishaan Goel, Giyoung Kwon, Katherine Papen, and Prakhar Saxena. "A Data-Driven Approach to Job Skills and Occupations: Evidence from Latin America." Charles River Economics Labs, Chicago, IL.
- Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock. 2023. "GPTs Are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models." Mimeo.
- Escudero, Verónica, Hannah Liepmann, and Damián Vergara. 2023. "Directed Search, Minimum Wages, and Worker Characteristics: Evidence from an Online Job Board." Mimeo.
- Evans, David, Claire Mason, Haohui Chen, and Andrew Reeson. 2023. "An algorithm for Predicting Job Vacancies using Online Job Postings in Australia." *Humanities & Social Sciences Communications* 10:102.
- Faberman, R. Jason and Marianna Kudlyak. 2019. "The Intensity of Job Search and Search Duration." *American Economic Journal: Macroeconomics* 11(3):327-57.
- Fabo, Brian and Lucia M'ytina Kureková. 2022. "Methodological Issues Related to the Use of Online Labour Market Data." Working Paper 68, ILO, Geneva.
- Forsythe, Eliza, Lisa B. Kahn, Fabian Lange, and David Wiczer. 2020. "Labor Demand in the Time of COVID-19: Evidence from Vacancy Postings and UI Claims." *Journal of Public Economics* 189.
- Gallagher, Elizabeth, India Kerle, Cath Sleeman, and George Richardson. 2022. A New Approach to Building a Skills Taxonomy. Technical Report TR-16, Economic Statistics Centre of Excellence, National Institute of Economic and Social Research, London.
- Gmyrek, Pawel, Hernan Winkler, and Santiago Garganta. 2024. "Buffer or Bottleneck? Employment Exposure to Generative AI and the Digital Divide in Latin America." Policy Research Working Paper 10863, World Bank, Washington, D.C.
- Grootendorst, Maarten. 2019. "Bertopic: Neural Topic Modeling with a Class-based TF-IDF Procedure." Mimeo.
- Grundke, Robert, Luca Marcolin, The Linh Bao Nguyen, and Mariagrazia Squicciarini. 2018. "Which Skills for the Digital Era?: Returns to Skills Analysis." OECD Science, Technology and Industry Working Papers 2018/09. Paris: OECD.

- Gu, Yu, Robert Tinn, Hao Cheng, Michael Lucas, Naoto Usuyama, Xiadong Liu, Tristan Naumann, Jianfeng Gao, and Hoifung Poon. 2021. "Domain-specific Language Model Pretraining for Biomedical Natural Language Processing." *ACM Transactions on Computing for Healthcare* 3(1).
- Guerra, N., K. Modecki, and Wendy Cunningham. 2014. "Social-Emotional Skills Development across the Life Span: PRACTICE." World Bank Policy Research Working Paper 7123, World Bank, Washington, D.C.
- Hale, T., N. Angrist, R. Goldszmidt, B. Kira, A. Petherick, T. Phillips, S. Webster, E. Cameron-Blake, L. Hallas, S. Majumdar, and H. Tatlow. 2021. "A Global Panel Database of pandemic policies (Oxford COVID-19 Government Response Tracker)." *Nature Human Behaviour* 5: 529-538
- Hansen, Stephen, Peter John Lambert, Nicholas Bloom, Steven J. Davis, Raffaella Sadun, and Bledi Taska. 2023. "Remote Work across Jobs, Companies, and Space." Working Paper 31007, Cambridge, MA.
- Hershbein, Brad and Lisa B. Kahn. 2018. "Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings." *American Economic Review* 108(7):1737-72.
- Hidalgo, César A. and Ricardo Hausmann. 2009. "The Building Blocks of Economic Complexity." *PNAS* 106(26):10570-5.
- Hidalgo, C.A., B. Klinger, A.-L. Barabási, and R. Hausmann. 2007. "The Product Space Conditions the Development of Nations." *Science* 317(5837):482-7.
- IFC (International Finance Corporation). 2019. *Digital skills in sub-Saharan Africa: Spotlight on Ghana*. Washington, DC: IFC.
- ILO (International Labour Organization). 2023. "2022 Labour Overview: Latin America and the Caribbean." International Labour Organization, Geneva.
- ILO (International Labour Organization). 2020. "The Feasibility of Using Big Data in Anticipating and Matching Skills Needs." International Labour Organization, Geneva.
- ITU (International Telecommunication Union). 2020. *Digital Skills Assessment Guidebook*. Geneva: ITU.
- INDEC (Instituto Nacional de Estadística y Censos). 2018. "Correspondencias entre el CON-17 y la CIUO-08." INDEC, Buenos Aires.
- Jensen, Robert. 2012. "Do Labor Market Opportunities Affect Young Women's Work and Family Decisions? Experimental Evidence from India," *The Quarterly Journal of Economics* 127(2):753-92.
- Jensen, Robert. 2010. "The (Perceived) Returns to Education and the Demand for Schooling," *The Quarterly Journal of Economics* 125(2):515-48.

- Ji, Shaoxiong, Tianlin Zhang, Kailai Yang, Sophia Ananiadou, Erik Cambria, and Jörg Tiedemann. 2023. "Domain-specific Continued Pretraining of Language Models for Capturing Long Context in Mental Health." Mimeo.
- José-García, Adán, Alison Sneyd, Ana Melro, Anaïs Ollagnier, Georgina Tarling, Haiyang Zhang, Mark Stevenson, Richard Everson, and Rudy Arthur. 2023. "C3-ioc: A Career Guidance System for Assessing Student Skills Using Machine Learning and Network Visualisation." *International Journal of Artificial Intelligence in Education* 33(4):1092–1119.
- Katz, Lawrence F., Jonathan Roth, Richard Hendra, and Kelsey Schaberg. 2022. "Why do sectoral employment programs work? Lessons from WorkAdvance." *Journal of Labor Economics* 40(S1):S249-91.
- Kuhn, Peter and Kailing Shen. 2015. "Do Employers Prefer Migrant Workers? Evidence from a Chinese Job Board," *IZA Journal of Labor Economics* 4(22):1-31.
- Kuhn, Peter and Kailing Shen. 2013. "Gender Discrimination in Jobs Ads: Evidence from China." *The Quarterly Journal of Economics* 128/910:287-336.
- La Buonora, Lucía. 2017. "Homogeneización de la variable Ocupación en las Encuestas Continuas de Hogares de Uruguay: 1993-2016." Thesis, Universidad Católica de Uruguay.
- Lassébie, Julie, Luca Marcolin, Marieke Vandeweyer, and Benjamin Vignal. 2021. "Speaking the Same Language: A Machine Learning Approach to Classify Skills in Burning Glass Technologies Data." Social, Employment and Migration Working Papers No. 263, OECD, Paris.
- Lewandowski, Piotr, Albert Park, Wojciech Hardy, Yang Du, and Saier Wu. 2022. "Technology, Skills, and Globalization: Explaining International Differences in Routine and Nonroutine Work Using Survey Data." *The World Bank Economic Review* 36(3):687-708.
- Lin, Jeffrey. 2011. "Technological Adaptation, Cities, and New Work." *The Review of Economics and Statistics* 93(2):554-574.
- Lo Bello, S., M. L. Sánchez-Puerta, and H. Winkler. 2019. "From Ghana to America: The Skill Content of Jobs and Economic Development." World Bank Policy Research Working Paper 8758, Washington, DC: World Bank.
- MAC (Migration Advisory Committee). 2017. "Assessing Labour Market Shortages: A Methodology Update." Migration Advisory Committee, London.
- Marinescu, Ioana. 2017. "The General Equilibrium Impacts of Unemployment Insurance: Evidence from a Large Online Job Board," *Journal of Public Economics* 150:14-29.
- McKenzie, David. 2017. "How Effective are Active Labor Market Policies in Developing Countries? A Critical Review of Recent Evidence." *World Bank Research Observer* 32(2): 127-54.
- Modestino, Alicia Sasser, Daniel Shoag, and Joshua Ballance. 2016. "Downskilling: Changes in Employer Skill Requirements over the Business Cycle." *Labour Economics* 41:333-47.

- Molina, Eduardo Chávez, Franco Bernasconi, and José Rodríguez de la Fuente. 2020. "Propuesta de correspondencias entre CON y CIUO: Sintaxis para SPSS, Stata y R." Working Paper 6, Universidad de Buenos Aires, Buenos Aires.
- Moroz, Harry, JJ Naddeo, and Nga Thi Nguyen. 2021. "Digital Skills in a Digitizing Vietnam: Assessing the Demand for Digital Skills in Vietnam." World Bank, Washington, D.C.
- Muller, Noel and Abba Safir. 2019. "What Employers Actually Want: Skills in Demand in Online Job Vacancies in Ukraine." World Bank Social Protection & Jobs Discussion Paper 1932. Washington, DC: World Bank.
- Napierala, Joanna and Vladimir Kvetan. 2022. "Changing Job Skills in a Changing World," In *Handbook of Computational Social Science for Policy*, edited by Eleonora Bertoni, Matteo Fontana, Lorenzo Gabrielli, Serena Signorelli, and Michele Vespe, 243-59. Cham: Springer.
- Nayyar, Gaurav, Mary Hallward-Driemeier, and Elwyn Davies. 2021. *At Your Services? The Promise of Services-Led Development*. Washington, D.C.: World Bank.
- Nomura, Shinsaku, Saori Imaizumi, Ana Carolina Areias, and Futoshi Yamauchi. 2017. "Toward Labor Market Policy 2.0: The Potential for Using Online Job-Portal Big Data To Inform Labor Market Policies in India." Policy Research Working Paper 7966, World Bank, Washington, D.C.
- OECD (Organisation for Economic Co-operation and Development). 2022. *Skills for the Digital Transition: Assessing Recent Trends using Big Data*. Paris: OECD.
- OECD (Organisation for Economic Co-operation and Development). 2021. "An Assessment of the Impact of COVID-19 on Job and Skills Demand using Online Job Vacancy Data." OECD, Paris.
- OECD (Organisation for Economic Co-operation and Development). 2021b. *Effective Adult Learning Policies: Challenges and Solutions for Latin American Countries*. Paris: OECD.
- OECD (Organisation for Economic Co-operation and Development). 2019. *OECD Skills Outlook 2019: Thriving in a Digital World*. Paris: OECD.
- OECD (Organisation for Economic Co-operation and Development). 2016. "Skills for a Digital World: 2016 Ministerial Meeting on the Digital Economy Background Report." OECD Digital Economy Papers 250. Paris: OECD.
- OECD (Organisation for Economic Co-operation and Development) and ILO (International Labour Organization). 2018. "Global Skills Trends, Training Needs and Lifelong Learning Strategies for the Future of Work." G-20 Employment Working Group, Geneva.
- Parrilla, Sebastián. 2022. "Indicadores de empleo verde y azul en Uruguay." International Labour Organization and Ministry of Labor and Social Security, Montevideo.
- Pennings, Steven. 2020. "The Utilization-Adjusted Human Capital Index (UHCI)." Policy Research Working Paper 9375, World Bank, Washington, D.C.
- Samek, Lea, Mariagrazia Squicciarini, and Emile Cammeraat. 2021. "The Human Capital behind AI: Jobs and Skills Demand from Online Job Postings." OECD Science, Technology, and Industry Policy Papers No. 120, OECD, Paris.

- Sato, Misato, Leanne Cass, Aurélien Saussay, Francesco Vona, Leo Mercer and Layla O’Kane. 2023. *Skills and Wage Gaps in the Low-carbon Transition: Comparing Job Vacancy Data from the US and UK*. London: Grantham Research Institute on Climate Change and the Environment and Centre for Climate Change Economics and Policy, London School of Economics and Political Science. Grantham Institute on Climate.
- Seegmiller, Bryan, Dimitris Papanikolaou, and Lawrence D.W. Schmidt. 2023. “Measuring Document Similarity with Weighted Averages of Word Embeddings.” *Explorations in Economic History* 87.
- Shehu, Milena and Eralda Gjika. 2024. A Comprehensive Review of the Three Main Topic Modeling Algorithms and Challenges in Albanian Employability Skills. *European Scientific Journal*, 20(12):31-51.
- Shen, Kailing and Yanran Zhu. 2023. “Labor Force Transition Dynamics: Unemployment Rate or Job Posting Countrs?” Discussion Paper 16373, Institute of Labor Economics, Bonn.
- SkillsFuture Singapore. 2022. “Skills Demand for the Future Economy.” SkillsFuture Singapore, Singapore.
- Sorensen, Karen, and Jean Michel Mas. 2016. “A Roadmap for the Development of Labor Market Information System.” African Union/ FHI 360, US Agency for International Development, Washington, DC.
- Testaverde, Mauro and Josefina Posadas. 2021. *Toward a World-Class Labor Market Information System for Indonesia An Assessment of the System Managed by the Indonesian Ministry of Manpower*. Washington, D.C.: World Bank.
- Torres, Jose and Sidonia McKenzie. 2020. “Youth Unemployment in Uruguay.” IMF Working Paper WP/20/281, International Monetary Fund, Washington, D.C.
- Tsvetkova, Alexandra, Elettra D’Amico, Alexander Lembcke, Polina Knutsson, and Wessel Vermeulen, 2024. “How well do online job postings match national sources in large English speaking countries? Benchmarking Lightcast data against statistical sources across regions, sectors and occupations.” OECD Local Economic and Employment Development (LEED) Papers, OECD, Paris.
- Velardez, Miguel Omar. 2021. “Análisis de distancias ocupacionales y familias de ocupaciones en el Uruguay.” Economic Commission for Latin America and the Caribbean, Santiago.
- Vona, Francesco, Giovanni Marin, Davide Consoli, and David Popp. 2018. “Environmental Regulation and Green Skills: An Empirical Exploration.” *Journal of the Association of Environmental and Resource Economists* 5(4):713-53.
- UNESCO (United Nations Educational, Scientific and Cultural Organization). 2019. “TVET Policy Review: Malawi.” UNESCO, Paris.
- UNESCO (United Nations Educational, Scientific and Cultural Organization). 2019b. “TVET System Review: Myanmar.” UNESCO, Paris.

- UNESCO (United Nations Educational, Scientific and Cultural Organization). 2017. *Digital Skills for Life and Work*. Report by the Broadband Commission for Sustainable Development. Paris: UNESCO.
- Vermeulen, Wessel and Fernanda Gutierrez Amaros. 2024. "How Well Do Online Job Postings Match National Sources in European Countries? Benchmarking Lightcast Data against Statistical and Labour Agency Sources across Region, Sectors and Occupation." OECD Local Economic and Employment Development (LEED) Papers, OECD, Paris.
- Weinberger, Catherine J. 2014. "The Increasing Complementarity Between Cognitive and Social Skills." *Review of Economics and Statistics* 96: 849–861.
- Wiswall, Matthew and Basit Zafar. 2015. "How Do College Students Respond to Public Information about Earnings?" *Journal of Human Capital* 9(2):117-69.
- World Bank. 2023. *A New Growth Horizon: Improve Fiscal Policy, Open Markets, and Invest in Human Capital*. Washington, D.C.: World Bank.
- World Bank. 2022. *Country Climate Development Report: Argentina*. Washington, D.C.: World Bank.
- World Bank. 2022b. "Argentina Labor Market and Social Protection Diagnostic." World Bank, Washington, D.C.
- World Bank. 2022c. *Indonesia's Online Vacancy Outlook: From Online Job Postings to Labor Market Intelligence 2020*. Washington, D.C.: World Bank.
- World Bank. 2021. *Toward a World-Class Labor Market Information System for Indonesia*. Washington, D.C.: World Bank.
- World Bank. 2020. *The Human Capital Index 2020 Update: Human Capital in the Time of COVID-19*. Washington D.C.: World Bank.
- World Bank. 2019. *World Development Report 2019: The Changing Nature of Work*. Washington, D.C.: World Bank.
- World Bank. 2019b. "Monitoring Occupational Shortages: Lessons from Malaysia's Critical Occupations List." World Bank, Kuala Lumpur.
- World Bank. 2015. *Uruguay Systematic Country Diagnostic*. Washington, D.C.: World Bank.
- Zhu, Yun, Faizan Javed, and Ozgur Ozturk. 2016. "Semantic Similarity Strategies for Job Title Classification." *arXiv:1609.06268*.

Appendix 1: Using job vacancy postings as a leading indicator of labor market health

Table 15: The relationship between job postings and the unemployment and employment rates in Argentina and Uruguay, 2020-23

VARIABLES	Argentina				Uruguay			
	(1) UR	(2) ER	(3) UR	(4) ER	(1) UR	(2) ER	(3) UR	(4) ER
UR(-1)	-0.167 (0.173)		-0.146 (0.154)		-0.195 (0.163)		-0.126 (0.168)	
Postings	0.00501 (0.0624)	0.0645 (0.156)			0.260 (0.273)	-0.0916 (0.232)		
ER(-1)		-0.195 (0.223)		-0.117 (0.0995)		-0.175 (0.164)		-0.232 (0.153)
Smoothed Postings(-2)			-0.00226 (0.0111)	0.00717** (0.00318)			-0.0450 (0.0660)	0.0209** (0.00804)
Constant	-0.20*** (0.0698)	0.330** (0.134)	-0.020** (0.00854)	0.00428* (0.00235)	-0.101 (0.0819)	0.147** (0.0721)	-0.00835 (0.00996)	0.00153 (0.00123)
Observations	43	43	41	41	39	39	38	38
R-squared	0.030	0.021	0.024	0.131	0.056	0.036	0.029	0.196

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The period is August 2020 to December 2023. The unemployment and employment rates are included in differences across all models. The variables Postings and Smoothed Postings are included in logarithmic differences. The variable Postings reflects the total number of monthly job postings. The variable Smoothed Postings is smoothed using moving averages. Models 1 and 2 regress the unemployment and employment rates on the total number of monthly Lightcast posts, while Models 3 and 4 regress these rates on the smoothed series of the number of postings. The models in the table include the postings variable (raw and smoothed) using the time lag that presents the best fit. ER = Employment rate; UR = Unemployment rate.

Source: Encuesta Permanente de Hogares (Argentina); Encuesta Continua de Hogares (Uruguay); Lightcast.

Appendix 2: Description of the skill concepts in the manual classification approach

Skills are categorized into five categories: cognitive, socioemotional, digital, manual, and technical.

- *Cognitive skills* encompass the mental abilities to think, learn, and solve problems (Almlund et al. 2011). Cognitive skills include basic skills such as literacy and numeracy as well as more sophisticated skills such as critical thinking, problem solving, and time management. Thus, cognitive skills are subdivided into eleven subcategories based on the skills that employers commonly demand: thinking, mathematics, communication, financial skills, business systems, project management, data analysis, language skills, and adaptability.
- *Socioemotional skills* are the attitudes and behaviors used to manage personal and social situations (Guerra, Modecki, and Cunningham 2014). Socioemotional skills are divided into eight subcategories: teamwork, communication, general social skills, organizational skills, character, customer service, people management, and creativity. In this taxonomy, creativity was classified as a socioemotional skill following the approach of Cunningham (2022), Cunningham and Villaseñor (2016), Muller and Safir (2019).
- *Digital skills* are the skills needed to work with information and communications technology (ICT) software and devices (Cunningham et al. 2022). Although often associated with the skills of specialized ICT workers, digital skills are used across occupations and tasks (UNESCO 2017; IFC 2019; ITU 2020; DE4A 2021; Moroz, Naddeo, and Nguyen 2021). Based on UNESCO (2017), there are three subcategories of digital skills: basic, intermediate, and advanced (**Table 16**). Basic digital skills are the ability to access and use digital technologies to perform basic tasks. Intermediate digital skills are the ability to use professional software for analysis, creation, management, and design. Advanced skills are the ability to perform specialized ICT tasks.

Table 16: Digital skill categories

Level	Definition	Examples
Basic	Ability to access and use digital technologies to perform basic tasks	<ol style="list-style-type: none"> 1. Functional use of digital devices 2. Online communication via emails 3. Using software for presentations, basic spreadsheet use 4. Finding, managing, and storing digital information and content (e.g., social media)
Intermediate	Ability to use professional software for analysis, creation, management, and design	<ol style="list-style-type: none"> 1. Using professional software for analytics, accounting, project management 2. Digital marketing, social media analytics 3. Web design, graphic design
Advanced	Ability to perform specialized ICT tasks	<ol style="list-style-type: none"> 1. Computer programming 2. Cloud computing, network management 3. Artificial intelligences 4. Data science, big data analytics 5. Cyber security 6. Web development, search engine optimization

Source: Cunningham et al. (2022) based on UNESCO (2017) and IFC (2019).

- *Manual skills* refer to gross and fine motor motions and coordination (Bennet et al. 2022). Skills associated with this definition are divided into five subcategories: finger dexterity, hand-foot-eye coordination, driving, flying, and physical skills.
- *Technical skills* are specific abilities to undertake a job. In our case, technical skills are treated as a residual category capturing skills that are not categorized as cognitive, socioemotional, digital, or manual and whose formulation is related to a specific degree or occupational knowledge. This category is divided into 19 subcategories by applying Lightcast's subcategorization to those skills not grouped in the above categories. Examples of technical skills include security services, the arts, and health sector skills.

Appendix 4: Comparing the manual and machine learning classifications

To get a basic sense for the similarity between the manual and machine learning taxonomies, we compare the 50 lowest-level skills categories in the machine learning taxonomy to the 33 lowest-level skills categories in the manual approach. The main metric we employ is the mutual information score, which for two distinct clusterings $U = \{U_1, \dots, U_n\}$ and $V = \{V_1, \dots, V_m\}$ of the same N datapoints is defined as:

$$MI(U, V) = \sum_{i=1}^n \sum_{j=1}^m \frac{|U_i \cap V_j|}{N} \log \left(\frac{N|U_i \cap V_j|}{|U_i||V_j|} \right)$$

This metric essentially measures the extent to which the clusterings U and V may be modeled as arising from independent random variables, where if the clusterings are completely independent, $MI(U, V)$ will be near 0. However, when there are a large number of clusters (as there are in this case), this metric reports clusterings as more similar than they actually are due to random chance alignments. Therefore, we consider the *adjusted* mutual information score:

$$AMI(U, V) = \frac{MI(U, V)}{\frac{H(U) + H(V)}{2} - \mathbf{E}[MI(U, V)]}$$

where $H(U)$ denotes the entropy (i.e. total information content) of U .

The *adjusted* mutual information score is 0.00225, which indicates that the taxonomies are very different at the lowest-level skills category.

Appendix 5: Correlations of skill requirements including skill subcategories categories and specific occupations, 2020-23

Table 17: Correlations of skill requirements including detailed digital skill categories, 2020-23

Bivariate correlations

a. Argentina							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) cognitive	1.000						
(2) digital basic	0.143	1.000					
(3) digital intermediate	0.182	0.177	1.000				
(4) digital advanced	0.182	0.076	0.413	1.000			
(5) socioemotional	0.200	0.096	0.073	0.046	1.000		
(6) manual	0.014	-0.010	0.062	0.075	0.047	1.000	
(7) technical	0.140	0.071	0.114	0.114	0.063	0.058	1.000

b. Uruguay							
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) cognitive	1.000						
(2) digital basic	0.211	1.000					
(3) digital intermediate	0.307	0.158	1.000				
(4) digital advanced	0.284	0.043	0.484	1.000			
(5) socioemotional	0.261	0.139	0.112	0.048	1.000		
(6) manual	0.052	0.037	0.137	0.096	0.106	1.000	
(7) technical	0.254	0.130	0.159	0.155	0.249	0.084	1.000

Note: Bivariate correlations across all skill categories and digital subcategories at the firm level. Only firms with non-missing ID and with more than 10 postings are included.

Source: Lightcast.

Table 18: Correlations of skill requirements for Software Developers, 2020-23

Bivariate correlations

a. Argentina

Skills categories and cognitive skills subcategories

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1) thinking	1.000													
(2) math	0.069	1.000												
(3) communic.	0.074	0.034	1.000											
(4) financial	0.111	0.030	0.045	1.000										
(5) bus. systems	0.131	0.045	0.061	0.078	1.000									
(6) quality contr.	0.059	0.010	0.053	0.031	0.086	1.000								
(7) business anal.	0.166	0.060	0.084	0.104	0.195	0.072	1.000							
(8) project mgt	0.132	0.049	0.085	0.049	0.101	0.047	0.116	1.000						
(9) data analysis	0.106	0.063	0.035	0.044	0.121	0.010	0.098	0.067	1.000					
(10) language	0.085	0.030	0.060	0.008	0.035	0.038	0.063	0.090	0.089	1.000				
(11) adaptability	0.041	-0.003	0.012	0.012	-0.001	0.008	0.012	0.012	-0.001	0.004	1.000			
(12) digital	0.048	0.011	0.000	0.005	0.040	0.086	0.010	-0.020	0.052	0.083	-0.010	1.000		
(13) social	0.187	0.026	0.099	0.055	0.088	0.126	0.135	0.154	0.039	0.113	0.037	0.013	1.000	
(14) technical	0.174	0.069	0.098	0.089	0.126	0.102	0.176	0.107	0.132	0.089	0.019	0.004	0.216	1.000

Skills categories and digital skills subcategories

Variables	(1)	(2)	(3)	(4)	(5)	(6)
(1) cognitive	1.000					
(2) digital-basic	0.064	1.000				
(3) digital-intermediate	0.163	0.088	1.000			
(4) digital-advanced	0.105	0.020	0.249	1.000		
(5) social	0.231	0.091	0.171	0.033	1.000	
(6) technical	0.231	0.092	0.141	0.004	0.216	1.000

Skills categories and social skills subcategories

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) cognitive	1.000										
(2) digital	0.075	1.000									
(3) teamwork	0.103	0.031	1.000								
(4) communication	0.167	-0.006	0.145	1.000							
(5) general social	0.130	0.010	0.097	0.200	1.000						
(6) organizational	0.134	0.070	0.133	0.126	0.042	1.000					
(7) character	0.078	0.010	0.072	0.123	0.190	0.043	1.000				
(8) customer service	0.096	-0.056	0.045	0.107	0.071	0.027	0.071	1.000			
(9) people management	0.133	0.027	0.083	0.172	0.121	0.068	0.079	0.113	1.000		
(10) creativity	0.071	0.018	0.116	0.045	0.135	0.024	0.102	0.044	0.042	1.000	
(11) technical	0.231	0.004	0.105	0.163	0.140	0.118	0.077	0.145	0.121	0.062	1.000

b. Uruguay

Skills categories and cognitive skills subcategories

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1) thinking	1.000													
(2) math	-0.002	1.000												
(3) communic.	0.086	0.013	1.000											
(4) financial	0.105	0.008	0.041	1.000										
(5) bus. systems	0.058	0.152	0.033	0.081	1.000									
(6) quality contr.	0.050	0.024	0.104	0.008	0.053	1.000								
(7) business anal.	0.104	0.008	0.143	0.064	0.108	0.078	1.000							
(8) project mgt	0.106	0.091	0.159	0.025	0.126	0.097	0.158	1.000						
(9) data analysis	0.103	0.066	0.029	0.009	0.047	0.035	0.125	-0.015	1.000					
(10) language	0.113	-0.098	0.000	-0.043	-0.109	0.007	0.007	-0.074	0.092	1.000				
(11) adaptability	0.010	0.006	-0.003	0.012	0.035	0.006	-0.007	-0.010	0.060	-0.003	1.000			
(12) digital	0.091	0.041	0.040	0.010	0.019	0.077	0.025	0.023	0.073	0.121	-0.005	1.000		
(13) social	0.266	0.033	0.149	0.045	0.097	0.142	0.154	0.145	0.085	0.141	0.037	0.141	1.000	
(14) technical	0.079	0.110	0.079	0.083	0.130	0.105	0.092	0.134	0.148	-0.021	0.002	0.106	0.220	1.000

Variables	(1)	(2)	(3)	(4)	(5)	(6)
(1) cognitive	1.000					
(2) digital-basic	0.058	1.000				
(3) digital-intermediate	0.255	0.107	1.000			
(4) digital-advanced	0.190	0.042	0.293	1.000		
(5) social	0.346	0.101	0.225	0.123	1.000	
(6) technical	0.165	0.118	0.227	0.141	0.220	1.000

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) cognitive	1.000										
(2) digital	0.205	1.000									
(3) teamwork	0.118	0.078	1.000								
(4) communication	0.181	0.048	0.294	1.000							
(5) general social	0.186	0.065	0.397	0.387	1.000						
(6) organizational	0.242	0.095	0.140	0.164	0.101	1.000					
(7) character	0.069	0.027	0.063	0.134	0.177	0.051	1.000				
(8) customer service	0.088	0.006	0.038	0.129	0.122	0.037	0.093	1.000			
(9) people management	0.167	0.097	0.231	0.310	0.276	0.069	0.091	0.129	1.000		
(10) creativity	0.067	0.023	0.085	0.013	0.018	0.009	0.084	0.049	-0.025	1.000	
(11) technical	0.165	0.106	0.123	0.155	0.147	0.206	0.028	0.089	0.103	0.075	1.000

Note: Bivariate correlations across all skill categories and digital subcategories at the firm level. Only firms with non-missing ID and with more than 10 postings are included.

Source: Lightcast.

Table 19: Correlations of skill requirements for Administrative and Executive Secretary, 2020-23

Bivariate correlations

a. Argentina

Skills categories and cognitive skills subcategories

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1) thinking	1.000													
(2) math	0.048	1.000												
(3) communic.	0.074	0.038	1.000											
(4) financial	0.055	0.027	-0.028	1.000										
(5) bus. systems	0.063	0.015	0.041	0.013	1.000									
(6) quality contr.	0.064	0.078	0.023	0.016	0.053	1.000								
(7) business anal.	0.120	0.042	0.046	0.113	0.103	0.065	1.000							
(8) project mgt	0.093	0.022	0.016	0.001	0.052	0.050	0.064	1.000						
(9) data analysis	0.108	0.051	0.053	0.046	0.054	0.061	0.065	0.056	1.000					
(10) language	0.149	0.036	0.106	0.024	0.087	0.061	0.121	0.109	0.111	1.000				
(11) adaptability	0.041	0.009	-0.003	0.009	0.000	-0.003	0.000	0.032	0.013	0.004	1.000			
(12) digital	0.099	0.044	0.089	0.061	0.047	0.061	0.046	0.035	0.101	0.134	0.014	1.000		
(13) social	0.128	0.008	0.087	0.020	0.058	0.039	0.089	0.102	0.041	0.079	0.047	0.006	1.000	
(14) technical	0.128	0.037	0.075	0.058	0.075	0.054	0.157	0.032	0.068	0.106	0.015	0.112	0.096	1.000

Skills categories and digital skills subcategories

Variables	(1)	(2)	(3)	(4)	(5)	(6)
(1) cognitive	1.000					
(2) digital-basic	0.086	1.000				
(3) digital-intermediate	0.032	0.147	1.000			
(4) digital-advanced	0.087	0.076	0.153	1.000		
(5) social	0.114	0.008	0.022	0.062	1.000	
(6) technical	0.130	0.046	0.113	0.140	0.096	1.000

Skills categories and social skills subcategories

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) cognitive	1.000										
(2) digital	0.089	1.000									
(3) teamwork	0.057	0.040	1.000								
(4) communication	0.148	0.069	0.113	1.000							
(5) general social	0.110	0.069	0.102	0.171	1.000						
(6) organizational	0.167	0.105	0.108	0.099	0.108	1.000					
(7) character	0.047	0.036	0.048	0.067	0.100	0.047	1.000				
(8) customer service	0.025	0.005	0.052	0.148	0.067	-0.034	0.027	1.000			
(9) people management	0.059	0.019	0.053	0.109	0.119	0.052	0.039	0.066	1.000		
(10) creativity	0.053	0.054	0.106	0.076	0.110	0.035	0.042	0.056	0.037	1.000	
(11) technical	0.130	0.112	0.051	0.143	0.089	0.125	0.026	0.094	0.094	0.050	1.000

b. Uruguay

Skills categories and cognitive skills subcategories

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1) thinking	1.000													
(2) math	0.012	1.000												
(3) communic.	0.056	0.014	1.000											
(4) financial	0.038	0.070	-0.001	1.000										
(5) bus. systems	0.056	0.074	-0.008	0.041	1.000									
(6) quality contr.	-0.027	-0.011	0.015	-0.027	-0.008	1.000								
(7) business anal.	0.117	-0.016	0.010	0.101	0.075	0.110	1.000							
(8) project mgt	0.151	0.013	0.045	0.025	0.127	0.046	0.130	1.000						
(9) data analysis	0.063	0.004	0.019	0.013	0.077	-0.005	0.092	0.064	1.000					
(10) language	0.193	-0.004	0.028	0.050	0.111	-0.015	0.196	0.123	0.100	1.000				
(11) adaptability	0.072	-0.008	0.025	-0.011	0.002	-0.010	0.016	0.053	-0.015	0.028	1.000			
(12) digital	0.148	0.030	0.063	0.177	0.086	-0.034	0.124	0.088	0.097	0.151	0.028	1.000		
(13) social	0.136	0.029	0.098	0.089	0.039	0.008	0.101	0.090	0.053	0.127	0.044	0.154	1.000	
(14) technical	0.096	0.021	0.123	0.044	0.067	0.032	0.112	0.095	0.047	0.089	0.035	0.109	0.251	1.000

Skills categories and digital skills subcategories

Variables	(1)	(2)	(3)	(4)	(5)	(6)
(1) cognitive	1.000					
(2) digital-basic	0.175	1.000				
(3) digital-intermediate	0.137	0.129	1.000			
(4) digital-advanced	0.111	0.047	0.129	1.000		
(5) social	0.196	0.108	0.110	0.075	1.000	
(6) technical	0.169	0.059	0.112	0.117	0.251	1.000

Skills categories and social skills subcategories

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) cognitive	1.000										
(2) digital	0.251	1.000									
(3) teamwork	0.087	0.113	1.000								
(4) communication	0.144	0.116	0.090	1.000							
(5) general social	0.145	0.128	0.121	0.161	1.000						
(6) organizational	0.267	0.250	0.032	0.127	0.102	1.000					
(7) character	0.077	0.052	0.061	0.125	0.221	0.084	1.000				
(8) customer service	0.072	0.031	-0.032	0.168	-0.005	0.029	0.038	1.000			
(9) people management	0.122	0.078	0.007	0.125	0.158	0.097	0.135	0.046	1.000		
(10) creativity	0.046	0.026	0.120	0.033	0.128	-0.004	0.089	0.018	0.028	1.000	
(11) technical	0.169	0.109	0.055	0.216	0.125	0.107	0.049	0.122	0.134	0.047	1.000

Note: Bivariate correlations across all skill categories and digital subcategories at the firm level. Only firms with non-missing ID and with more than 10 postings are included.

Source: Lightcast.

Appendix 6: Methodology for calculating revealed comparative advantage

The adaptation of the product space of goods (Hidalgo and Hausmann 2009) to the labor market first required adapting the concept of revealed comparative advantage (RCA). In the present case, the advantages are determined based on observations of the labor market. To this end, it is proposed that a specific skill has a revealed comparative advantage for the development of an occupation when the average requirement for that skill within the occupation is higher than the average requirement across all occupations.

To analyze this, the relative use of skill h in occupation i is examined through the indicator CA:

$$CA_{i,h} = \frac{I_{h,i}}{\sum_{i=1}^{i=O} I_{h,i} / O}$$

where the numerator $I_{h,i}$ indicates the intensity of use of skill h in occupation i and the denominator shows the average intensity of use of skill h across the total (O) occupations analyzed.

Using this tool, the specialization basket is defined in terms of occupations for each skill through a binary matrix with a number of rows equal to the number of occupations (O) and a number of columns equal to the number of skills (H). The elements of this matrix can be defined as follows:

$$RCA = \begin{cases} 1 & \text{if } VC_{i,h} \geq u \\ 0 & \text{if } VC_{i,h} < u \end{cases}$$

In this study, u is defined as equal to 1, which means a usage intensity of skill h that is greater than or equal to the average across occupations. This threshold could take different values, for example, greater than 1, in order to set a higher standard when defining comparative advantage. Based on the binary VCR matrix, the frequency with which two occupations have RCA together is calculated, meaning the number of skills whose usage intensity is above the average in both occupations. This allows for the identification of occupations that might require similar skills.

Next, a matrix is constructed in the Occupation Space ($O \times O$), where each element indicates the frequency with which two occupations intensively use the same skill relative to the frequency with which the skill is intensively used.

$$P(VCR(o_j|o_i)) \text{ or } P(VCR(o_i|o_j))$$

Depending on whether it is divided by the marginal row or column.

Finally, the proximity between two occupations i and j is defined as the minimum of the conditional probability of having a comparative advantage in one of the two, given that it is present in the other occupation. That is:

$$\varphi_{ij} = \operatorname{argmin} \{P(VCR(o_j|o_i)); P(VCR(o_i|o_j))\}$$

To construct the Occupation Space graph, the proximity matrix is simplified by extracting the most important connections that keep all nodes in the graph connected. This is done using a Maximum Spanning Tree algorithm, as in Hidalgo et al. (2007).³⁹

³⁹ A spanning tree refers to different subsets of the graph (in this case, the Occupation Space) that include all the attributes (occupations) with the minimum number of edges. When edges have weights (in this case given by the distances between occupations), it is possible to obtain the Maximum Spanning Tree, which is the spanning tree that maximizes the sum of the edge weights.