Escudero, Verónica; Liepmann, Hannah; Podjanin, Ana

Book Part

Using Online Vacancy and Job Applicants' Data to Study Skills Dynamics

Reference: In: Big Data Applications in Labor Economics, Part B (2024). Emerald Publishing Limited, S. 35 - 99.
https://doi.org/10.1108/S0147-91212024000052B023.
doi:10.1108/S0147-91212024000052B023.

USING ONLINE VACANCY AND JOB APPLICANTS' DATA TO STUDY SKILLS DYNAMICS

Verónica Escudero[a,b], Hannah Liepmann[a,c] and Ana Podjanin[a]

[a]*International Labour Organization, Switzerland*
[b]*Center for Affective Global Action, USA*
[c]*IZA, Germany*

ABSTRACT

Outside of Europe and the United States, the knowledge on skills dynamics is scarce due to a lack of data. We therefore assess whether online data on vacancies and applications to a job board can help fill this gap. We propose a taxonomy with three broad categories – cognitive, socioemotional, and manual skills – and 14 commonly observed subcategories, which we define based on unique skills identified through keywords and expressions. The taxonomy is comprehensive but succinct, suitable for developing and emerging economies, and adapted for online data. We then develop a text-mining approach to implement the taxonomy. Based on Uruguayan data from the job board BuscoJobs, we find that our model is able to assign skills to 64% of applicants' employment spells and 94% of vacancies. While online data are usually skewed toward highly qualified work, we show that our data include meaningful numbers of vacancies and applicants of intermediate and even lower qualification levels. Our approach relies on data that are currently available in many countries, thereby allowing for country-specific analysis that does not assume that occupational skills are constant across countries. This is key as we find considerable differences between our findings and those using US O-NET data. Finally, we end with an illustration of how our approach can inform the

Big Data Applications in Labor Economics, Part B
Research in Labor Economics, Volume 52B, 35–99

ISSN: 0147-9121/doi:10.1108/S0147-91212024000052B023

analysis of skills dynamics. To our knowledge, we are the first to explore this approach in the context of emerging economies.

Keywords: Online big data; job board; skills dynamics; skills taxonomy; natural language processing

1. INTRODUCTION

Major transformative phenomena, such as technological progress and trade, are shaping labor markets. Skills are an important factor in this process as they significantly influence how labor market transformations affect employment opportunities. For example, the demand for specific skills depends on whether these are complementary to or can be substituted by a newly introduced technology. As a result, workers' skills determine the comparative resilience of certain groups in contemporary labor markets, while affecting the relative vulnerability of others.

A rich literature has examined skills dynamics in Europe and, in particular, the United States. Computer technology has been found to replace tasks that can be routinized and complement nonroutine analytical and interactive tasks (Atalay et al., 2020; Autor et al., 2003), although the rising demand for nonroutine analytical tasks has been reversed more recently (Beaudry et al., 2016). Some studies predict that artificial intelligence and robotics will also replace nonroutine analytical labor and lead to large-scale job destruction (Frey & Osborne, 2017). Other analyses, in contrast, demonstrate the importance and nonreplicability of interactive skills within occupations, projecting smaller net job losses (Arntz et al., 2016). Given systematic differences in labor markets, it would be misleading to extrapolate these findings directly to other countries. Yet, the knowledge on skills dynamics is scarce outside of Europe and the United States, which is largely driven by the absence of adequate sources of data. Such knowledge is urgently needed today as it would contribute to understanding which policy responses are required to better prepare workers and respond to employers' needs in fast-changing labor markets.

In this article, we assess whether online data on vacancies and applications to a job board (or job portal) are a suitable source for studying skills dynamics outside of the high-income economies that the literature has so far focused on. The advantage of this approach is its reliance on data that is currently available in many countries across the world, thereby allowing for country-specific analyses that do not need to assume that occupational skills are the same across countries. Another distinguishable feature is the ability to study detailed skills dynamics across time, representative of both labor demand and supply. This results from the panel nature and granularity of the data, which is distinct from currently available survey data from emerging and developing economies. Finally, to the best of our knowledge, we are the first to explore this approach in the context of an emerging economy.

To devise our taxonomy, we build on the growing literature that uses online vacancy data from the United States to study skills dynamics, in particular those key contributions that have proposed seminal taxonomies to categorize skills (Deming & Kahn, 2018; Deming & Noray, 2020; Hershbein & Kahn, 2018). For our purposes, a limitation of such taxonomies is that they are specifically tailored to professional workers in the United States. Therefore, we complement such efforts, by further incorporating manual skills and broadening the underpinnings of socioeconomic skills, with a view to expanding the scope of possible analyses of online data beyond workers with high formal qualification levels. This is particularly relevant when studying the supply and demand of skills in emerging economies, and it is also relevant for high-income countries as it allows for analyses that are tailored to a more comprehensive set of workers. Specifically, to obtain a taxonomy that is comprehensive but succinct, we systematically aggregate three broad categories of skills – cognitive, socioemotional, and manual – and 14 commonly observed and recognizable skills subcategories. We identify these based on predefined unique tasks characterized through keywords and expressions stemming from our literature review and complemented by our own analysis. Each skills subcategory is therefore grounded in the academic literature, where we combine insights from seminal studies in economics (including also the literature on skill-biased technological change, see, for example, Acemoglu & Autor, 2011) and psychology (see, for example, Almlund et al., 2011).

We then develop a text-mining methodology that allows implementing our skills taxonomy in big online vacancy and applicants' data. This methodology provides a tool for researchers and practitioners to (i) preprocess raw information in textual data, and (ii) detect and classify the skills contained in a given country-specific dataset using natural language processing (NLP) techniques. We test this approach empirically, by exploiting information from the Uruguayan job board *BuscoJobs*. Thanks to the availability of these rich data, the setting of Uruguay is well suited for our purposes. In more detail, *BuscoJobs* is the biggest private job-search portal in the country and our data include the entire content uploaded between 2010 and 2020, capturing both the supply side (through applicants sharing their current profile and labor market biographies) and the demand side (in the form of vacancies posted by firms). Uruguay also represents an informative setting, since it allows studying a labor market – and skills supplied and demanded – that are clearly distinct from the cases the literature has so far focused on. To illustrate this, Uruguay is classified as a high-income country by the World Bank, but its current GDP per capita amounts to only 27% of the US American GDP per capita (see World Bank, 2023). In 2019, informal employment rates were at 24.5% in Uruguay, which is more than twice as high as the estimated 9.8% for Northern America (ILO, 2023). Moreover, the Uruguayan labor market regulation is comparatively strong, given the decisive influence of tripartite collective bargaining, which can impact the speed of labor market transformations (ILO, 2014).

To test our methodology, we use five types of metrics: first, the degree of representativeness and bias of these data in comparison to the overall labor market and how stable this is across time; second, the share of observations (vacancies and applicants' job spells) classified in terms of their skills content; third, the possibility of capturing a broader set of skills vis-à-vis taxonomies from seminal studies that have analyzed similar data in a high-income context and neglected, in particular, manual skills; fourth, the difference in results obtained when using Uruguayan data compared to what we would have obtained when relying instead on US O-NET data under the assumption that occupational skills are the same across these two countries; and finally, the applicability and suitability of our approach to study skills dynamics.

To preview our empirical results, we find that our NLP model is able to characterize the vast majority of vacancies and applicants' job spells in terms of the skills requirements and the skills that people have, respectively. The results significantly improve once we consider synonyms in addition to the keywords and expressions included in our predefined skills taxonomy. Then, 64% of applicants' job spells and 94% of vacancies are assigned to at least one of our 14 skills subcategories. While we lack a direct benchmark to compare these results (due to the limited availability of supply-side data on job boards and aggregators), we believe our model allows for a meaningful classification of skills that should enable credible analyses. This is especially notable considering the substantial heterogeneity in the quality of information available in the free-text descriptions we use to code the skills variables and gives hope to the possibility of replicating our methodology in similar data sources of other countries.[1]

We also find that, although our data are not fully representative of the country's labor force, they allow for meaningful analyses of labor supply and demand pertaining to intermediate and even lower qualification levels. Based on the Uruguayan household survey, we show, for example, that applicants tend to be younger, more educated, and more likely to live in the capital than the overall labor force. They are also more likely to work in clerical support occupations and are less represented in craft and other trades. This is inherent to online vacancy and applicants' data and may require focusing the analysis on specific segments of the labor market and/or using weighting techniques to account for specific biases (Fabo & Kureková, 2022). At the same time, we show that in addition to highly qualified work, the data include meaningful numbers of vacancies requiring and applicants having intermediate and even lower qualification levels, which represents a characteristic that is not commonly found in online vacancy datasets from the US. We also establish that the bias in the *BuscoJobs* data remains stable across time for both vacancies and applicants' job spells, since the discrepancy between the occupational distribution in the *BuscoJobs* data and the household survey data does change significantly from 2010 through 2020. This means that the data are suitable for studying dynamics over time (for a similar argument, see Deming & Noray, 2020; Hershbein & Kahn, 2018).

We also show the importance of capturing a mixture of different skills, which might be especially relevant when analyzing skills dynamics outside of Europe and the United States. A large share of classified skills (61% in the case of the applicants' data) is attributable to keywords and expressions from studies analyzing online data, which neglect manual skills. However, additional sources also play a meaningful role. For the applicants' data, 31% of classified skills thus relate to keywords and expressions from studies using non-online sources and 9% to complementary keywords and expressions from the pilot version of O-NET Uruguay.

We demonstrate the importance of employing country-specific data when assessing skills dynamics outside of Europe and the United States, in line with recent literature (Lewandowski et al., 2020, 2022; Velardez, 2021). Within occupations we find considerable differences when comparing our findings to what we would have obtained, had we used O-NET data from the United States. Looking at the relative importance of the three broad skills categories (i.e., manual, socioemotional, and cognitive skills), we find that manual skills play a larger role in the composition of occupations using our Uruguayan data, than using the US O-NET. The importance of cognitive skills varies more strongly in our data from 60% for professionals to below 20% for plant and machine operators, while all occupations have in general lower scores in the socioemotional skills category in our approach than according to the US O-NET data.

Finally, we end with an illustration of how our approach can inform the analysis of skills dynamics. We illustrate the evolving supply and demand for specific skills narrowing our focus to a subset of the data, where we consider only applicants who applied to at least one vacancy and vacancies that received applications from at least one applicant. Notably, we observe a consistent gap between the skills sought by job vacancies and those explicitly indicated by applicants, and this gap widens over time. We posit that this gap may be related to employers' increasing, more complex, or more exigent demands placed on prospective employees, as reflected in the lengthier job postings. While these findings are not meant to offer conclusive results, they provide a first foundation for future research utilizing big data from job boards and our methodology.

In addition to relating to the literature for high-income countries as discussed further above, this study advances the literature on skills dynamics in emerging and developing countries. While the relative scarcity in empirical evidence in these economies is in large part due to the absence of appropriate data, some studies have addressed this data challenge in innovative ways. One strand of the literature studies skills dynamics in low- and middle-income countries by imputing occupational task information from the US O-NET data (Almeida et al., 2017; Bhorat et al., 2018; Reijnders & de Vries, 2018). This requires making the strong assumption that the task content of occupations is invariant across countries, which we do not need to make in our approach. A second strand of the literature measures tasks directly in the countries studied, combining survey data from the Programme for the International Assessment of Adult Competencies (PIAAC) and the Skills Measurement Program (STEP) with other longitudinal data sources. These studies identify significant differences in the skill composition of jobs and occupations across countries and highlight the importance of analyzing these questions using country-specific data (Carbonero et al., 2023; Caunedo et al., 2023; Lewandowski et al., 2020, 2022; Lo Bello et al., 2019). For example, Lewandowski et al. (2020) show that even within the same occupations, emerging and developing countries rely more strongly on routine work than developed economies. Moreover, Zapata-Román (2021) documents, in contrast to advanced economies, increases in earnings for routine occupations in Chile, while Xing (2021) finds the opposite for China.[2] Unfortunately, for developing and emerging economies, PIAAC and STEP surveys are currently available as single cross-sections only.[3] In contrast to our study, this strand of the literature therefore neglects skills dynamics that occur within occupations over time, although findings for high-income countries show that within-occupational skills changes can be significant (Atalay et al., 2020; Spitz-Oener, 2006). Moreover, skills variables in the PIAAC and STEP surveys are limited in number, making it difficult to assess skills in a comprehensive manner. This stands in contrast to the granularity of the data that we propose.

This article is structured as follows. Section 2 develops the skills taxonomy and explains how we classify unique skills within categories and subcategories. Section 3 describes the *BuscoJobs* data. Section 4 summarizes our methodology to implement the skills taxonomy in the *BuscoJobs* data. It describes the process we follow to create skills variables through the identification of keywords and expressions and the use of NLP techniques. It also provides an illustration of how our approach can inform the analysis of skills dynamics. Section 5 concludes with a brief discussion of how online job data could be used in the future to analyze a range of research questions for emerging and developing countries.

2. ASSESSING SKILLS OUTSIDE OF THE UNITED STATES: A SKILLS TAXONOMY FOR RESEARCH PURPOSES

We now devise our taxonomy. The discussion first motivates our objectives and general approach. Grounded in the literature from labor economics and psychology, we then present the conceptual underpinnings and definitions of our broad and more detailed skills categories.

Our taxonomy has to strike a balance between different objectives. To begin with, we want to categorize skills in a comprehensive way, capturing broadly defined skills with a view to understanding major labor market trends. In addition, we want these broadly defined skills categories to be suitable for a Latin American context and, more generally, to be representative of an emerging and developing-country labor market. This requires adjustments in comparison to approaches that were developed for a North American or European country context. Finally, our taxonomy needs to be suitable for an implementation using online data on job vacancies and applicants. This requires adapting the taxonomy to the mode of expression and vocabularies specific to this type of data.

With these objectives in mind, our taxonomy consists of the three broad categories of cognitive skills, socioemotional skills, and manual skills that we disaggregate into 14 subcategories (Table 1). Recent research confirms that these three broad skills categories represent vastly different attributes. Across workers' employment biographies, each category entails unique learning and adjustment patterns and therefore produces distinct returns (Lise & Postel-Vinay, 2020). This motivates our choice to organize the taxonomy around these three categories.

We deliberately focus on skills, rather than occupations. This is based on the realization that, first, skills are a central dimension for understanding how major transformative phenomena, such as technological progress or trade, affect employment opportunities (e.g., Acemoglu & Autor, 2011; Autor et al., 2003). Second, occupations are characterized by complex bundles of skills that change across time (e.g., Arntz et al., 2016; Atalay et al., 2020; Spitz-Oener, 2006). This implies that

Table 1. Categorization of Skills, Keywords, and Sources.

Category	Source of the Category	Keywords/Expressions	Source of Keywords/Expressions
		Cognitive Skills	
Cognitive skills (narrow sense)	DK (2018)	Problem solving, research, analytical, critical thinking, math, statistics	DK (2018)
		Mathematics, adaptability, direction, control, planning	ALM (2003) (from nonroutine analytic tasks)
		Data analysis, data engineering, data modelling, data visualization, data mining, data science, predictive analytics, predictive models	DN (2020)
		Analyse, design, devising rule, evaluate, interpreting rule, sketch	S-O (2006), APST (2020) (from nonroutine analytic tasks)
		Calculation	ALM (2003) (from routine analytic tasks)
		Bookkeeping, correcting, measurement	S-O (2006), APST (2020) (from routine cognitive tasks)
		Information processing, decision making, generation of ideas, memory	O-NET Uruguay
Computer (general) skills	DK (2018)	Computer, spreadsheets, common software, Excel, PowerPoint	DK (2018)
		Computer literacy, Internet skills, Word, Outlook, Office, Windows	DN (2020)
Software (specific) skills and technical support	DK (2018) & DN (2020)	Programming language or specialized software, Java, SQL, Python	DK (2018)
		Computer installation, computer repair, computer maintenance, computer troubleshooting, web development, site design	DN (2020)
Machine Learning and Artificial Intelligence	DK (2018) DN (2020)	Artificial intelligence, machine learning, decision trees, apache hadoop, Bayesian Networks, Automation Tools, Neural Networks, Support Vector Machines (SVM), Supervised learning, TensorFlow, MapReduce, Splunk, Convolutional Neural Network (CNN), Cluster Analysis	DN (2020)

Table 1. *(Continued)*

Category	Source of the Category	Keywords/Expressions	Source of Keywords/Expressions
Financial skills	DK (2018)	Budgeting, accounting, finance, cost	DK (2018)
Writing skills	DK (2018)	Writing	DK (2018)
		Editing, reports, proposals	DN (2020)
Project management skills	DK (2018)	Project management	DK (2018)
		Socioemotional Skills	
Character skills (conscientious-ness, emotional stability and openness to experience)	DK (2018)	Organized, detail oriented, multitasking, time management, meeting deadlines, energetic	DK (2018)
		Self-starter, initiative, self-motivated	DN (2020)
		Competent, achieving, hardworking, reliable, punctual, resistant to stress, creative, independent	KBHT (2016), HK (2012)
Social skills (including agreeableness and extraversion)	DK (2018)	Communication, teamwork, collaboration, negotiation, presentation	DK (2018)
		Team, persuasion, listening	DN (2020)
		Flexibility, empathy, assertiveness	KBHT (2016), HK (2012)
		Advice, entertain, lobby, teaching	S-O (2006), APST (2020) (from nonroutine interactive tasks)
		Interact with others, verbal abilities	O-NET Uruguay
People management skills	DK (2018)	Supervisory, leadership, management (not project), mentoring, staff	DK (2018)
		Staff supervision, staff development, performance management, personnel management	DN (2020)

Customer service skills	DK (2018)	Customer, sales, client, patient	DK (2018)
		Persuading, selling	ALM (2003) (from nonroutine analytic and interactive tasks)
		Advertise, sell, buy, purchase	S-O (2006), APST (2020) (from nonroutine interactive tasks)
		Repetitive customer service	ALM (2003) (from routine analytic and interactive tasks)
		Manual Skills	
Finger-dexterity skills	ALM (2003), under routine manual tasks	Picking, sorting, repetitive assembly, mixing ingredients, baking ingredients, sewing and decorative trimming, operating tabulating machines, packing agricultural produce	ALM (2003)
		Control, equip, operate	S-O (2006), APST (2020)
		Repetitive movements	O-NET Uruguay
Hand-foot-eye coordination skills	ALM (2003), under nonroutine manual tasks	Attending cattle, attending other animals, driving to transport passengers, *driving to transport charge*, piloting airplanes, pruning and treating ornamental and shade trees, performing gymnastic feats, *performing other sports requiring skill and balance*	ALM (2003)
		Accommodate, renovate, repair, restore, serving, *cleaning*	S-O (2006), APST (2020)
		Reaction on time, fine manipulations	O-NET Uruguay
Physical skills	O-NET Uruguay	Resistance, time dedicated to walking and running, *carrying heavy loads*	O-NET Uruguay

Source: Authors' elaboration.

Notes: ALM (2003) stands for Autor et al. (2003), APST (2020) for Atalay et al. (2020), DK (2018) for Deming and Kahn (2018), DN (2020) for Deming and Noray (2020), HK (2018) for Hershbein and Kahn (2018), HK (2012) for Heckman and Kautz (2012), KBHT (2016) for Kureková et al. (2016), and S-O (2006) for Spitz-Oener (2006). The O-NET Uruguay pilot project, which so far captures 22 selected occupations only, is detailed in Ministerio de Trabajo y Seguridad Social (2020) and Velardez (2021). The keywords used are meant to be the most comprehensive possible to provide appropriate definitions for categories and subcategories, but synonyms are not yet included at this stage. Even if some skills subcategories are closely related, the words used to categorize them are mutually exclusive. This allows for a unique identification of skills categories. Note also that within subcategories, some keywords are redundant (like "math" and "mathematics"). This simply means that more than one of the sources used that word to categorize the subcategory and has no repercussion for the implementation of our taxonomy as long as the repetition occurs within subcategories. Words in *italics* indicate keywords included by the authors of this article to complete existing definitions.

workers – across qualification levels – perform a combination of cognitive, socio-emotional, and manual skills.[4] The composition of these skills bundles is expected to shape how mega-trends (like technology and trade) affect workers' labor market situations. In a similar vein, ongoing policy debates emphasize the importance of skills that are portable across jobs and occupations (see ILO, 2021).[5]

As shown in Table 1, our taxonomy builds on Deming and Kahn (2018), who augment the task-based approach of Autor et al. (2003) to adapt it to the specific characteristics of online job data (BurningGlass data for the United States in their case). We have extended the approach of Deming and Kahn (2018) by adding information provided by a range of different studies (Atalay et al., 2020; Autor et al., 2003; Deming & Noray, 2020; Heckman & Kautz, 2012; Hershbein & Kahn, 2018; Kureková et al., 2016; Spitz-Oener, 2006) and O-NET Uruguay.[6] Specifically, within the categories of cognitive and socioemotional skills, we include additional keywords that are meant to broaden the scope of descriptions used. This also allows us to capture terms whose popularity changed over time, but which refer to the same skill (see Deming & Noray, 2020).[7] In addition, we add the entire category of "manual skills." Manual skills were not the focus of Deming and Kahn (2018) but are important for our goal of obtaining a comprehensive skills taxonomy. Manual skills are, moreover, particularly relevant in the context of an emerging country's labor market, such as that of Uruguay. Overall, we make use of the various keywords used by seminal articles to be the most comprehensive possible in the definition of categories and subcategories, while devising our conceptual framework.[8]

The 14 subcategories of our taxonomy capture both skills that pertain to tasks that workers perform on the job and skills that refer to individuals' personal attributes.[9] Both types of skills are important to provide a consistent representation for labor market developments in a wide range of contexts. In combination, they encompass skills that employers demand in vacancies and skills that workers supply and present in their online profiles. It is important to note that even if some skills subcategories are closely related, the words used to categorize them are mutually exclusive.[10] This allows for the unique identification of skills categories in our granular data.

Our categorization has the advantage of being both complete and succinct. The comprehensiveness of this approach is necessary for it to be relevant to different countries' realities and is possible thanks to the granularity of online data. Meanwhile, its compactness into three broad categories and only 14 subcategories is tailored to research purposes, where overly detailed taxonomies, ontologies, or uncategorized lists of skills could introduce complexity that would be challenging to utilize effectively.

Within the category of *cognitive skills*, we define the subcategory "cognitive skills (narrow sense)" as the abilities or qualities needed to perform tasks that require analysis and calculation, problem-solving, intuition, flexibility, and creativity (Acemoglu & Autor, 2011, p. 1076; Autor et al., 2003, p. 1284). A large literature has studied the relative demand for cognitive skills and their returns in high-income countries, and cognitive skills are recognized in policy debates as being central for workers' resilience to labor market transformations (ILO, 2021).

One strand of this literature thus looks at the polarization of occupations by skill level vis-à-vis technological change and automation.[11] This literature finds that occupations largely requiring cognitive skills have thrived with the development of computer technology, both in terms of job creation and wage growth. This is driven by cognitive skills that are difficult to routinize (notably, those associated with non-repetitive cognitive tasks), where occupations relying on such tasks have been more resilient to technological shocks. Meanwhile, low-skilled occupations (also non-substitutable by technology, as discussed below) have likewise witnessed a broad-based increase in employment, relative to middle-skilled occupations. More recent studies suggest that the rise of high-skilled occupations has evolved, documenting a reversal in the demand for cognitive skills in the United States in 2000 (Beaudry et al., 2016). While the share of highly educated college graduates has increased, occupations intensive in cognitive skills have seen little wage and employment growth after 2000. The finding is consistent with the Information Technology (IT) revolution, and hence the introduction of IT as a general-purpose technology, having reached a maturity stage. This implied that high-skilled workers moved into occupations relying less strongly on cognitive skills (Beaudry et al., 2016). The reversal in the demand for cognitive skills therefore has consequences for the wage structure of occupations (see also Roys & Taber, 2019).

Another strand of the literature complements this knowledge by focusing on skill changes within occupations.[12] These studies take advantage of richer data sources with greater granularity, such as vacancy data. They find substantial variation in skill requirements within occupations. While cognitive skills are an important driver of this variation, socioemotional skills are taking on a key role (Kureková et al., 2016). In this context, Deming and Kahn (2018) show that a greater emphasis on cognitive skills, especially when they are demanded in combination with socioemotional skills, is associated with higher wages and firm productivity. Therefore, the considerable within-occupational variation in the demand for such skills contributes to explaining a significant share of existing patterns of wage inequality. In addition, changes in skills over time play a decisive role also within occupations (Atalay et al., 2020; Spitz-Oener, 2006). As one example, STEM occupations have seen pronounced changes in cognitive skills requirements – for example, STEM vacancies requiring skills linked to machine learning and artificial intelligence rose by 460% in the decade before 2017 – which can be attributed to the rapid diffusion of technological innovation in STEM (Deming & Noray, 2020). Against this background, the words we choose to identify cognitive skills aim to match the analytical job tasks defined by Autor et al. (2003) that are used by most articles studying routine-biased technological change and employment polarization. We add additional words suggested by Atalay et al. (2020), Deming and Kahn (2018), Deming and Noray (2020), Spitz-Oener (2006), and O-NET Uruguay.

Finally, we complete cognitive skills by adding five additional subcategories from Deming and Kahn (2018): computer skills, software skills, writing skills, financial skills, and project management skills. We also add one additional category suggested by Deming and Noray (2020), which refers to machine learning and artificial intelligence. These subcategories are intricately linked to the cognitive skills described before and represent topical subcategories that organize cognitive skills around themes. We categorize them separately because they are commonly listed in a wide range of online job vacancies and applicants' work experiences. The list of topical subcategories targets in particular white-collar jobs (Deming & Kahn, 2018), which means that this list can be consulted at the detailed level or condensed, depending on the focus of a given study on a particular segment of the labor market.

Our second broad category of *socioemotional skills* adds to the categorization a set of personal attributes that involve intellect, but more indirectly and less consciously than cognitive skills. The literature uses a range of expressions to refer to such skills, including noncognitive skills, soft skills, socioemotional skills, or personality traits. These different terms suggest different properties.[13] First, we avoid the use of the term "noncognitive skills," as all skills require some sort of cognition. Second, while "traits" gives a sense of immutability, "skills" connotes the possibility of learning (Kautz et al., 2014). A long debate in the psychology literature has discussed whether socioemotional skills are stable across different situations at a fixed point in time, as well as the degree of malleability of these skills over the life course (see Almlund et al., 2011, sec. 2). Today, most psychologists and economists support the notion of a stable personality across situations, which is supported by a large body of evidence showing that stable socioemotional skills exist and predict a variety of behaviors (Kautz et al., 2014).[14] However, as pointed out by Heckman and Kautz (2012) "while traits are relatively stable across situations, they are not set in stone. They change over the life cycle." The consensus that socioemotional skills are malleable and can be learned, is key in terms of policy. If this were not the case, there would be little room for training initiatives. Another decisive question is whether socioemotional skills can still significantly change during adult life. Disciplines diverge on this question. Economists like Carneiro and Heckman (2005) and more recent publications by Heckman and coauthors put emphasis on the malleability of these skills until adolescence and early adulthood. From this perspective, parenthood, education policies, and workplace-based internships and apprenticeships are the central means for developing socioemotional skills. Meanwhile, social psychologists argue that at least some of these attributes (e.g., emotional intelligence) can be learned at any age, which places greater emphasis on training, learning, and education policies for adults.[15] Sociologists and management scholars agree, emphasizing the indispensability of experiential learning for some socioemotional skills such as problem solving, teamwork, and other social skills (see Green et al., 2001, sec. 2).

Measuring and classifying socioemotional skills is a challenge (Humphries & Kosse, 2017). Yet, the psychology literature has arrived at a well-accepted categorization of these skills, called the "five-factor personality model" or "big five" (McCrae & Costa, 2008). It includes agreeableness, conscientiousness, emotional stability, extraversion and autonomy, and openness to experience. A large literature on the labor market returns of so-called noncognitive skills has established that different noncognitive skills are important predictors of life outcomes, such as education and labor market success (Heckman & Kautz, 2012). Some of these personal attributes are correlated with cognitive skills (Kureková et al., 2016), but their explanatory power on labor market and education outcomes goes beyond this correlation.[16] A study for the United States found that improvements in noncognitive skills are much more important for future earnings and employment than similar improvements in cognitive skills (Heckman et al., 2006). Similarly, a study for Sweden showed that both cognitive and noncognitive skills are strong predictors of future earnings, but that noncognitive skills have a stronger effect for people at the low end of the earnings distribution (Lindqvist & Vestman, 2011). In the emerging and developing world, four studies have assessed the causal effects of socioemotional skills on labor market outcomes, by experimentally evaluating the effect of job training programs focused solely on socioemotional skills and different categories of workers. The evidence is more mixed, with two studies finding positive effects in India and Togo (Adhvaryu et al., 2023; Campos et al., 2017), another study from the Dominican Republic finding positive effects for women but not men (Acevedo et al., 2020), and a final study for Jordan finding null effects (Groh et al., 2016).

In terms of our taxonomy, we divide the five-factor model between the subcategories "character skills" and "social skills." Our subcategory "character skills" follows Deming and Kahn (2018), deviating somewhat from their definition (see Table 1). It includes conscientiousness (i.e., "the tendency to be organized, responsible, and hardworking," American Psychological Association, 2020), openness to experience (i.e., "the tendency to be open to new aesthetic, cultural, or intellectual experiences," American Psychological Association, 2020), and emotional stability (i.e., this is the contrary to "neuroticism." "Emotional stability is predictability and consistency in emotional reactions, with absence of rapid mood changes," American Psychological Association, 2020). It also encompasses dimensions such as being relaxed, independent, self-confident, and the degree of vulnerability to stress (Brunello & Schlotter, 2011; Heckman & Kautz, 2012).

There is no single agreed way to aggregate the five-factor personality model into broader categories. Most researchers bundle together various socioemotional skills depending on their specific research questions. We aggregate them into "character skills" and "social skills," as distinguishing between all five categories separately would increase the probability of making mistakes at this finer level of aggregation, without necessarily adding considerable value to the analysis. Instead, the two groups of "character skills" and "social skills" are based on the similarities in their predictive nature that have been found by the empirical literature. As such, we bundle conscientiousness, emotional stability, and openness to experience together within our subcategory "character skills," because these three factors appear to share the ability to predict education and labor market outcomes, although not with equal strength (Almlund et al., 2011; Brunello & Schlotter, 2011). For example, conscientiousness stands out as the most predictive trait of the five-factor model on future outcomes. It has been found to predict educational attainment, health, and labor market outcomes, in some cases, as strongly as measures of cognitive ability.[17] While conscientiousness predicts performance and wages across a wide spectrum of jobs, the predictive power of cognitive skills decreases with job complexity (Almlund et al., 2011; Kautz et al., 2014). Meanwhile, attributes related to emotional stability – especially internal locus of control, or the belief that one can determine one's success as opposed to believing that outcomes are the result of fate or luck – positively predict earnings (Brunello & Schlotter, 2011) and job search effort (Almlund et al., 2011). Finally, openness to experience predicts finer measures of educational attainment, such as class attendance (Almlund et al., 2011) but also years of education (Borghans et al., 2008). It has also been associated with transversal competencies, such as sense of initiative and entrepreneurship, which are important factors for education and labor market success (Brunello & Schlotter, 2011). The words we use to define such "character skills" come from Deming and Kahn (2018), but also from Deming and Noray (2020), Kureková et al. (2016), and Heckman and Kautz (2012).

We complete the category of *socioemotional skills* with three of Deming and Kahn (2018)'s subskills groups, which capture social skills (as mentioned before), people management skills, and customer service skills. Within "social skills" we include the original words suggested by the authors (e.g., communication, teamwork, collaboration). We add the remaining two categories from the five-factor personality model, namely agreeableness and extraversion. Agreeableness is the tendency to be cooperative toward others (American Psychological Association, 2020). Extraversion is the orientation people show toward the outer world and social contact (American Psychological Association, 2020). In our taxonomy, we characterize these two personal attributes with words such as cooperation, flexibility, empathy, and assertiveness.[18] We also add the subcategories of "people management skills" and "customer service skills," which are determined by a large set of socioemotional skills, but usually listed separately in online job vacancies and workers' job experiences. As before, we use various references as sources for keywords with a view to arriving at a complete taxonomy and a more comprehensive use of words to define each category (see the last column of Table 1).

The importance the literature attributes to social skills and other socioemotional skills has considerably increased during the last decade and policy debates emphasize the core relevance of these skills across workers' life cycle (ILO, 2021). Research shows that employers in the United States and Europe rank some social skills above cognitive ones, particularly in low-skilled labor markets (Bowles et al., 2001; Kureková et al., 2016). Moreover, studies point to an increasing complementarity between cognitive skills and social skills (Arntz et al., 2016; Borghans et al., 2014; Deming & Kahn, 2018; Weinberger, 2014). As computers substitute for a wider set of noninteractive tasks, interpersonal skills are becoming increasingly central in a wide range of professional jobs (Autor, 2014; Lu, 2015). Interestingly, according to research in Europe, the increased demand for interactive skills does not substitute formal education, but appears in addition to it (Kureková et al., 2016). The implications for workers in emerging and developing countries are potentially significant, as large shares of employment in these countries are concentrated in sectors and occupations (e.g., services) that require disproportionately interpersonal skills.

Our third category, *manual skills* takes as point of departure how Autor et al. (2003) categorize manual tasks. In their analysis, manual tasks are divided between routine and nonroutine ones. The former are activities that necessitate finger-dexterity skills and the latter activities requiring "situational adaptability, visual and language recognition, and in-person interactions" (Acemoglu & Autor, 2011, p. 1077). Broadly speaking, routine manual tasks are more prevalent in production and operative occupations and nonroutine manual tasks in service occupations, but also in some production and operative positions requiring physical or situational adaptability (Acemoglu & Autor, 2011).

The model by Autor et al. (2003) captures the vulnerability of routine manual labor in light of automation and outsourcing. We implement this category in our taxonomy using words such as picking or sorting, repetitive assembly, and mixing or baking ingredients. However, not all manual tasks enter this category. Several tasks require motor processing and visual capabilities that, given the contemporary state of technological progress, cannot easily be programmable (e.g., driving a car through traffic, cleaning, and housekeeping). In the United States, lower-skilled workers have thus moved into service occupations that are associated with such tasks and have witnessed employment growth due to a concurrent change in consumer preferences that favored low-skilled services (Autor & Dorn, 2013; Lee et al., 2022). Outside of high-income countries, excess supply of labor is disproportionately absorbed by low-value-added urban services, which contributes to an underdeveloped manufacturing sector (e.g., Rodrik, 2018). Nonroutine manual skills relate to tasks such as accommodating, serving, or cleaning, but also attending cattle in farms, pruning and treating plants, driving, and even performing gymnastics and other sports requiring balance.

Finally, we complement the categorization of manual skills with a "physical skills" subcategory that pertains specifically to personal attributes, physical strength and effort, such as resistance, as well as related tasks such as those requiring walking and running, and carrying heavy loads.

Manual skills have been less studied in the literature than cognitive ones. One reason is that these tasks, especially the part that can be routinized, have been declining in the advanced world (Spitz-Oener, 2006). Another reason is that manual skills, and more generally low-skill jobs, are underrepresented in online sources, such as the BurningGlass data, and are therefore not the focus of studies using these data (Hershbein & Kahn, 2018). Yet, there are some findings from the advanced world that shed light on the importance of integrating this skills category into a research-oriented taxonomy, given that these tasks are important when considering workers at the lower end of the wage distribution (Autor & Dorn, 2013). While research is scant outside the United States, we expect these tasks to be at least as important in the emerging and developing world, if not more important. Recently, Lise and Postel-Vinay (2020) studied the returns of manual skills requirements using O-NET data from the United States. They find that manual skills have in general moderate returns, and that they are easily accumulated on the job but also relatively fast lost when not employed. Roys and Taber (2019) explore the payoff to different skills for low-skilled workers in the United States over the life cycle. They find that while the payoff related to interpersonal skills has increased over time, the returns of manual skills remain the most relevant factor determining the wages of low-skilled workers. From a policy perspective, the finding suggests that investments in manual skills are key for improving the wages of these workers. Importantly, however, while there is a clear understanding that education and training is a path to develop cognitive skills, the literature has not sufficiently explored how to develop manual skills further. Meanwhile, policy practitioners, through their experience implementing these policies, point to the importance of work-based learning to develop technical skills, including manual ones (Cedefop, 2015; European Commission, 2013; Kis & Catriona Windisch, 2018; ILO, 2017). Therefore, understanding these developments better has profound implications for low-skilled workers. This extends to large segments of workers who perform manual work in emerging and developing countries, including in informal labor markets. Yet, it is also relevant for high-income countries as it allows for analyses that are tailored to a more comprehensive set of workers.

Finally, as a cross-cutting dimension, the routine intensity of skills has been decisive for understanding the substitutability of routine-manual and routine-cognitive tasks by technology and hence the polarization of employment structures in high-income countries (Autor et al., 2003; Goos et al., 2014; Spitz-Oener, 2006). Research outside of high-income countries relies on survey data that is available for some countries, while imputing routine intensity for others. It shows that low- and middle-income countries have seen no, or a comparatively small, decline in routine tasks between 2000 and 2017 (Lewandowski et al., 2020). In addition, middle-skilled jobs are not necessarily intense in routine tasks, such that a study for Argentina finds some evidence of a reallocation of employment toward middle-skilled jobs (Maurizio & Monsalvo, 2021).

Empirically capturing routine intensity within our skills categorization would require implementing it as a cross-cutting dimension, as we expect that routine intensity interacts differentially with the various skills categories. The existing literature conventionally relies on keyword-based identification to discern terms associated with routine skills (see Atalay et al., 2020; Autor et al., 2003; Spitz-Oener, 2006). This approach, however, presents complexities that our available data and methodology so far could not definitely resolve. In particular, the nuanced nature of certain skills, such as financial skills which may manifest varying degrees of routine-intensity, requires rigorous validation. Developing a robust methodology to capture skills' intensity effectively in our framework, thus remains a promising focus for future research.

3. DATA AND DESCRIPTIVE STATISTICS

To implement our skills taxonomy empirically, we rely on online data from the Uruguayan, private online job board *BuscoJobs*, which contains detailed information on (i) workers searching for jobs, (ii) vacancies posted by firms, and (iii) applications jobseekers have made to the vacancies posted. Our data cover the entire content uploaded on the job board from January 2010 through December 2020 in Uruguay.[19] In this section, we start by presenting the features of the job board (Section 3.1), then discuss conceptually issues of representativeness in job board data (Section 3.2), and finally assess the representativeness empirically, for both the applicants' and vacancy data (Sections 3.3 and 3.4).

3.1 Features of the BuscoJobs Data

To post online vacancies through *BuscoJobs*, enterprises need to create an account on the portal and pay a small fee. The portal offers several types of enterprise subscriptions, which cover different time spans and allow firms to publish between three and 600 vacancies. Firms can view users' curriculum vitaes (CVs) for the duration of their chosen subscription.

The data capture detailed information associated with each job vacancy. This includes characteristics of the firm posting the vacancy (e.g., name and location), firms' expectations and characteristics of the vacancy (e.g., desired age and sex of the candidate, required work experience and formal qualifications, the wage range offered), and a detailed text description of the job advertised. *BuscoJobs* generated upon request a variable indicating firms' industry, classified at the four-digit level of the International Standard Industrial Classification (ISIC), Revision 4.[20] Finally, we created a variable capturing occupational categories, following the two-digit ISCO-08 classification and using a machine-learning approach (see Section 4 for details).

As a distinctive feature, *BuscoJobs* not only allows firms to post vacancies but also enables jobseekers to register and apply for these vacancies directly through the portal. Jobseekers can create an account for free and include in their profile basic personal information (e.g., sex, birth dates, and places of residence), information about their educational attainment, the technical skills they possess, the languages they master, and, importantly, their entire history of employment (e.g., job spells' dates, position, location, and a detailed open-text description of the position). This allows us to observe rich longitudinal data on individuals' characteristics and employment biographies. Jobseekers can also express their job preferences, whether they are actively seeking a job, whether they are employed or unemployed, and whether they want their profiles to be fully visible for employers.[21] As for the vacancies, we created a variable capturing the occupational categories pertaining to applicants' employment spells, where we followed the two-digit ISCO-08 classification and employed a machine-learning approach.

Another advantage of the *BuscoJobs* database, in comparison to other job boards and job aggregators in the country, is that *BuscoJobs* does not include duplicate job postings. Duplicate job postings are a common characteristic of most job boards as these republish vacancies to increase applications, and of most job platforms as these scrape postings to improve their coverage (Office for National Statistics, 2021). Jijkoun (2016) estimates that a typical vacancy is reposted between two and five times with the proportion of duplicates ranging as high as 50 to 80% in different countries. Such duplication inflates the number of job advertisements, and as these are often not identical, they cannot be easily identified. Duplication thus undermines the representativeness and usability of vacancy data (Evans et al., 2023). In contrast, *BuscoJobs* has a policy of not reposting. Instead, it publishes vacancies for as long as firms require, for example, a few remain open for up to two months.

BuscoJobs also presents a low volatility over time (Equipos Consultores, 2020), covering a comparatively long time series from 2010 through 2020. It captures vacancies, applicants' employment histories, and applications. In contrast, other major job boards and aggregators in Uruguay and other data sources that are prominently used in the literature, such as the BurningGlass data, only contain vacancy data.

3.2 Coverage and Applicability of Applicants' and Vacancy Data to Identify and Measure Skills Variables

When assessing the demand and supply of skills and pursuing related analyses, online data have two general advantages, compared to other traditionally used data sources, such as labor force surveys (see also Hershbein & Kahn, 2018).

First, online vacancy and applicants' data have a high degree of granularity. This allows identifying detailed trends in skills and skills needs by generating variables of workers' tasks performed on the job and the tasks requested in vacancies. These statistical indicators can be computed without recurring to proxies, such as occupations or industries, which do not capture changes to skills within the groups of these categorizations (Hershbein & Kahn, 2018). Furthermore, online vacancy and applicants' data tend to provide detailed information on geographical location, occupations, and industries. These variables can often be generated in a robust way at a highly disaggregated level, thanks to large sample sizes (ILO, 2020).

Second, online vacancy and applicants' data are typically collected at high frequency, making it possible to avoid time gaps associated with more common data collection methods, such as survey data, which are sampled in sometimes long intervals. Some of the delays associated with the processing and cleaning of survey data can also be avoided (ILO, 2020), in particular if the initial database associated with an online job board or job aggregator is well structured.

One of the major drawbacks associated with the use of online job board and aggregator data is their representativeness. As they are not based on random sampling, it is difficult to draw generalized conclusions about the overall working age population or the universe of firms. The data tend to be biased toward certain segments of the labor market, in particular higher-skilled occupations (Cedefop, 2021; Fabo & Kureková, 2022; Hershbein & Kahn, 2018).

The representativeness depends on several characteristics inherent to the national context. This includes the internet penetration rate. In Uruguay, 83.4% of the population used the internet in 2019.[22] This is considerably above the global average, estimated at 53.7%, and the regional level, reaching 68.8% for Latin America and the Caribbean (LAC) (see World Bank, 2023). These figures suggest that exploiting web-based big data in the Uruguayan context could be associated with less bias, compared to other countries.

Nonetheless, also among more advanced countries where a high share of individuals uses the internet on a regular basis (e.g., the European Union, where 82.8% of the population used the internet in 2019, World Bank, 2023), there are major differences in the share of vacancies posted online. For example, the proportion of job vacancies published online in 2017 varied from below 50% in countries such as Romania or Denmark to close to 100% in Finland, Sweden, and Estonia (ILO, 2020). One explanatory factor is the share of the population living in urban or rural areas. In urban areas, the services sector tends to play a larger role and the incidence of online job advertisements is thus higher. In contrast, print media and word-of-mouth communication seem to be more relevant in rural areas (ILO, 2020). In Uruguay, a large majority of the working-age population (94.7%) resided in urban areas in 2019 (authors' calculation based on the household survey introduced in the next subsection). This is related to the prominent role of the capital Montevideo and the high degree of centralization in Uruguay.

Online vacancies also tend to be advertised by larger firms (e.g., international firms) and less by smaller firms, or firms operating in the construction, agricultural, and hospitality sectors (ILO, 2020). The incidence of informality also matters, since informal jobs often do not appear on public online job boards or aggregators (Cedefop, 2021). Uruguay has a significantly lower rate of informality than the LAC region as a whole. In 2019, 24.5% of all employment was informal in Uruguay compared to an estimated 53.6% in LAC (ILO, 2023).

Uruguay's labor market characteristics are thus amenable to the use of online portals by firms and jobseekers. Beyond this observation, it is important to empirically assess the degree and direction of biases of our data, which we do next.

3.3 Representativeness of the BuscoJobs Applicants' Data

The *BuscoJobs* applicants' database covers around 650,000 user profiles and a total of approximately 1,200,000 job spells. As shown in Fig. 1 (line with round markers), the data coverage is good from 2011 onward. The number of individuals joining the job board tends to increase up until 2017, to slightly decrease again until 2020.

To assess the representativeness of the database, we compare selected characteristics of the job board's users with information from the Uruguayan household survey, the Encuesta Continua de Hogares, which is representative of the country's population. Throughout this section, we present averages over annual results from 2010 through 2020.[23] The respective samples are restricted to individuals aged 15 and older, who are in the labor force, and live in Uruguay. Moreover, we transform the *BuscoJobs* employment spell data into a panel, where we consider each job board user's situation during the month of July in a given year.

Between 2010 and 2020, 47.4% of the applicants in our data are women, which compares well with the Uruguayan labor force (45.7%). However, 71.8% of the applicants are located in the Montevideo area, compared to only 41.6% of the

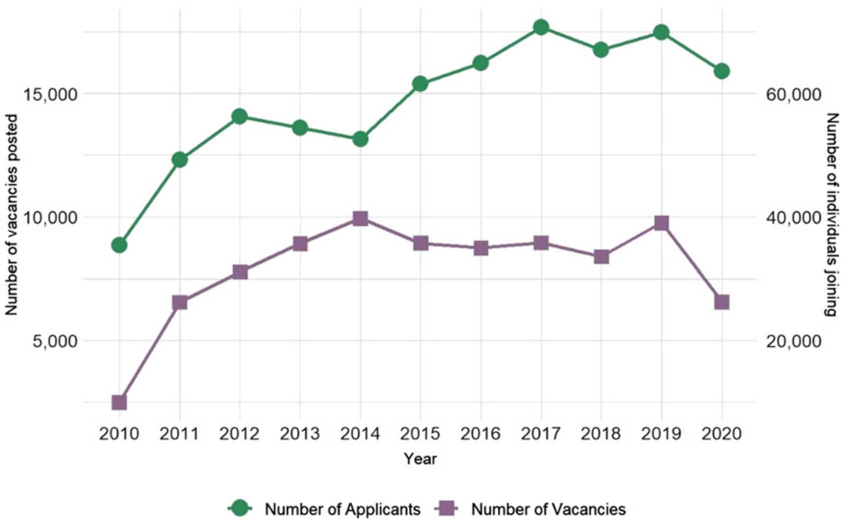

Fig. 1. Absolute Number of Individuals Joining *BuscoJobs* and Vacancies Posted in a Given Year, 2010–2020.

Notes: Authors' compilation based on *BuscoJobs* data. We restrict the sample to individuals aged 15 and older as well as to vacancies and individuals located in Uruguay. Round markers refer to individuals (scale shown on the right) and squared markers to vacancies (scale shown on the left).

overall labor force. The overrepresentation of the Montevideo area in the *BuscoJobs* applicants' database is likely related to the different job-search methods used in large metropolitan areas, compared with smaller urban areas or rural areas. This is consistent with the previously discussed finding of the higher relevance of print media and word-of-mouth communication in rural areas (ILO, 2020). According to the household survey, with 68.2%, a comparatively large share of those reporting to have used the internet to find a job, indeed lived in the Montevideo area during our period of analysis.

The applicants' data also contain a disproportionately high share of younger applicants (Fig. 2). This discrepancy is most pronounced among individuals aged 20–24 and 25–29, who taken together account for 48.5% of *BuscoJobs* users, compared to 22.3% of the Uruguayan labor force. More similar representations are observed for the age category 35–39 that accounts for around 11% of *BuscoJobs* users and the national labor force. The tendency is inverted as older age categories are considered.

The larger share of young portal users is intuitive given their stronger familiarity and use of IT tools, compared with older workers. It can also be rationalized by younger workers entering the labor market for the first time, and thus searching more actively for employment. For comparison, Marinescu and Skandalis (2021) report an average age of 31.8 for individuals using the online job-search platform of the French public employment services. This is very

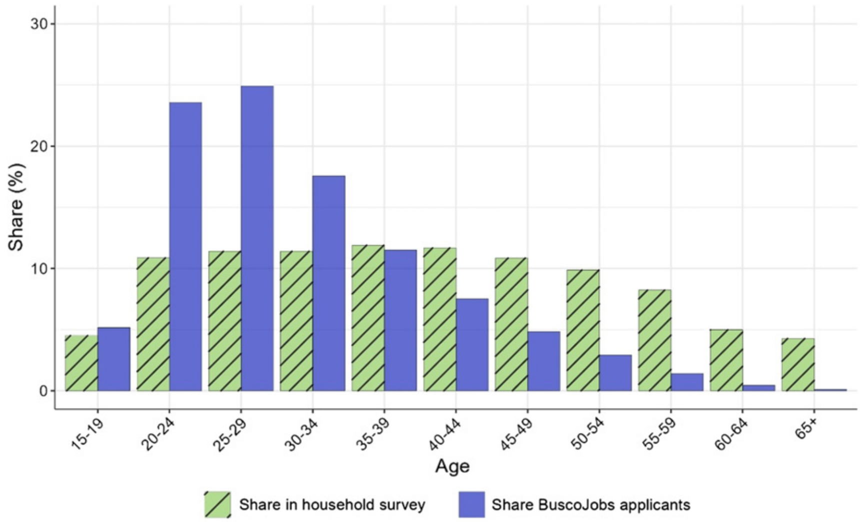

Fig. 2. Age Distribution, *BuscoJobs* Applicants' Database Compared to the Labor Force according to Household Survey Data, 2010–2020.
Notes: Authors' calculations based on the Uruguayan household survey and the *BuscoJobs* applicants' database. The latter exploits the panel dimension of the applicants' data, considering the situation of each individual in July of a given year.

similar to the average age we find (31.1 years). Other studies directly focus on young individuals, perhaps because these are frequent users of online job-search engines when applying for jobs (Barbarasa et al., 2017; Kureková & Žilinčíková, 2018).

When looking at formal qualification levels, we find that the overall qualification structure compares well between applicants on the *BuscoJobs* portal and the national labor force (Fig. 3). Large shares of both populations report having completed neither a technical and vocational degree nor a university degree. Nevertheless, this is less likely among applicants (68.1%) compared with the national labor force (74.5%). Applicants are slightly less likely to have completed a technical and vocational degree, while being more likely to have obtained university degrees. Taken together, the findings indicate that applicants have above-average qualification levels.

In terms of the occupational distribution, Fig. 4 shows a certain divergence when comparing *BuscoJobs* applicants' employment spells with the national distribution. Notably, a disproportionately large share of the *BuscoJobs* users was employed in a clerical support profession (35.2%), while none of the users indicated any work experience in an agricultural profession.

This suggests that when analyzing the applicants' data, it may be appropriate to focus on certain groups of workers that are well represented without making inferences to the whole population and/or to use weighting techniques to

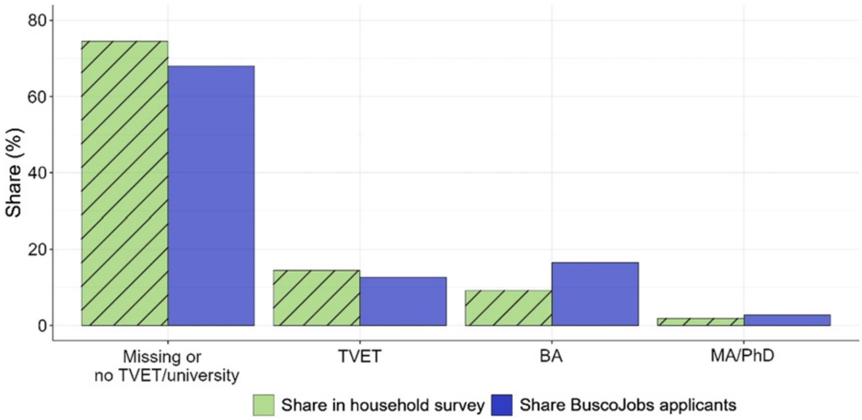

Fig. 3. Highest Formal Qualification Level, *BuscoJobs* Applicants' Database Compared to Household Survey Data, 2010–2020.
Notes: Authors' calculations based on the Uruguayan household survey and the *BuscoJobs* applicants' database. The latter exploits the panel dimension of the applicants' data, considering the situation of each individual in July of a given year. "TVET" stands for technical and vocational education and training at any level, that is, this includes the lower secondary, upper secondary, and short-cycle tertiary education levels. We only consider completed qualifications.

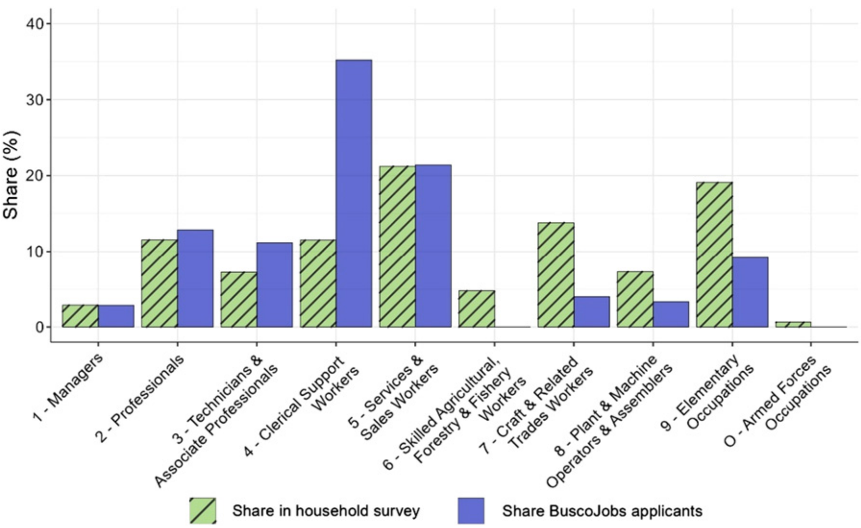

Fig. 4. Occupational Distribution, *BuscoJobs* Applicants' Database Compared to Household Survey Data, 2010–2020.
Notes: Authors' calculations based on the Uruguayan household survey and the *BuscoJobs* applicants' database. The latter exploits the panel dimension of the applicants' data, considering the situation of each individual in July of a given year. We include employed individuals only and display one-digit ISCO 08 occupational categories.

improve representativeness. As a matter of comparison, while many studies exploiting online applicants' data do not provide information on individuals' occupations, Marinescu and Rathelot (2018) show that there is significant overlap − but also some divergence − between the online data from the job board CareerBuilder and national survey data for the United States.[24]

Yet, while the applicants' dataset is not fully representative of the national occupational distribution, it has several comparative advantages. The dataset includes reasonably large samples for occupations requiring intermediate or lower levels of formal qualification, such as services and sales workers, clerical support workers, and even workers from elementary occupations. This suggests that the data can be used for analyses that extend beyond highly qualified workers, which one might not have thought at first glance.[25] Moreover, as mentioned above, the rich and granular information – in particular the information on tasks performed by workers – is an asset to obtain a better understanding of skills dynamics.

We are also able to document that the occupational biases are stable across time (Fig. 5). A primary concern when using these data is whether the representatives of the sample change over time; as this would be a threat to internal validity of any analysis exploiting a temporal dimension (Deming & Noray, 2020; Hershbein & Kahn, 2018). Fig. 4 shows that this is not the case. Based on Hershbein and Kahn (2018), the x-axis illustrates the *BuscoJobs* share in occupations in 2010 minus the share in the same occupation and year in the Uruguayan Household Survey, and the y-axis illustrates the same percentage point difference for each year from 2011 to 2020. Darker shades correspond to earlier years. The 45-degree line shows occupations where the shares between *BuscoJobs* and the household survey, did not change from 2010. The only group clearly deviating from this line are services and sales workers, where the year of 2010 seems to represent an outlier.

3.4 Representativeness of the BuscoJobs Vacancy Data

BuscoJobs is one of the leading job boards in Uruguay. According to the provider's calculations, it captures around 50% of online vacancies in Uruguay, providing information that is detailed and reliable. Vacancies are recent, have high turnover, and many of the fields available for each vacancy are complete (see Di Capua et al., 2020). As discussed before, the jobs ads posted are unique and are clean from repetitions, which is not a given among job portals. As other job portals and job aggregators in Uruguay post the same vacancies several times, it is estimated that the effective coverage of *BuscoJobs* is in fact closer to 60% of online vacancies.

Our data contain almost 87,000 vacancies, posted by more than 6,500 firms. As shown in Fig. 1 above (line with squared markers), the number of vacancies posted in the job board has increased constantly to reach a peak in 2014. While the observed numbers have remained at a high level also over the following years, the 2020 drop in the job postings might reflect the economic slowdown induced by the COVID-19 pandemic. It is important to note that one job posting can describe

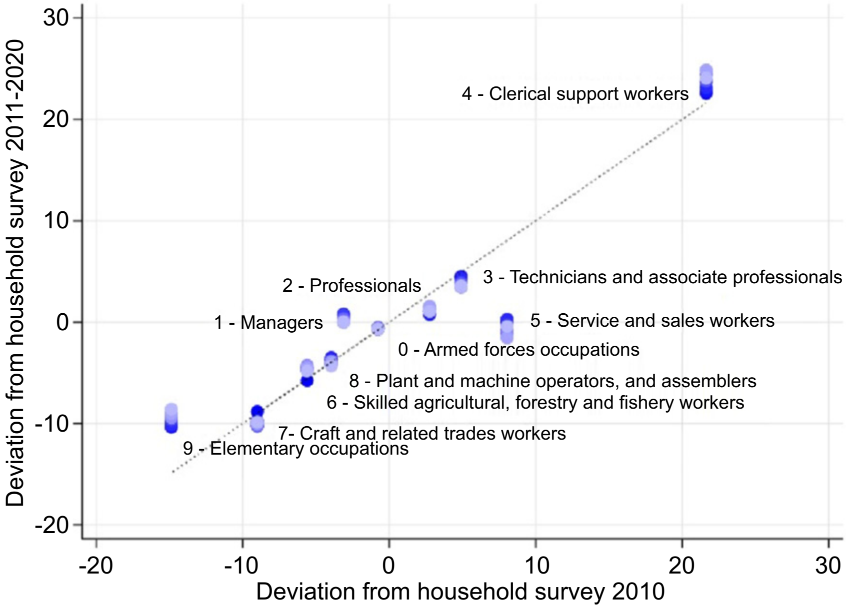

Fig. 5. Representativeness of *BuscoJobs* Occupations in Applicants' Data, Relative to Occupational Distribution in Uruguayan Household Survey Data (2010–2020).

Notes: Author's calculations based on *BuscoJobs* applicants' database and Uruguayan household survey data, where we focus on employed individuals only. Based on Hershbein and Kahn (2018), the *x*-axis illustrates the *BuscoJobs* share in occupations in 2010 minus the share in the same occupation and year in the Uruguayan household survey. The *y*-axis illustrates these differences between the two samples for each year from 2011 to 2020. Dark shades correspond to earlier years. The 45-degree line shows occupations where the shares between *BuscoJobs* and the household survey did not change from 2010. Occupations are defined according to the one-digit ISCO 08 classification.

several vacant positions. In fact, 81.9% of the job advertisements posted on the job board are associated to one vacancy only, while 8.6% are associated to two vacancies, and 9.5% are linked to three or more open vacancies. Throughout the article, we conduct the analysis at the level of job postings.

For the assessment of the representativeness of the *BuscoJobs* online vacancy dataset, we again rely on the comparison with the national household survey data and present averages over annual results for years 2020 to 2010.[26] As before, the household survey analysis focuses on individuals aged 15 and older, who belong to the labor force. We also restrict our attention to individuals and vacancies located in Uruguay. Our comparison is indicative only, because vacancy data depend on turnover rates in the labor market and these may differ across regions, occupations, and economic sectors. In contrast, the household survey data represent a snapshot of workers' characteristics in a given moment. Nevertheless, this comparison is the best possible way to assess the representativeness of the *BuscoJobs* vacancy data.[27]

We find that 43.0% of the vacancies are located in the capital of Montevideo, compared to 41.6% of the overall labor force.[28] Moreover, *BuscoJobs* online vacancies include a disproportionally high share of professional categories, and underrepresent occupations requiring lower formal qualification levels (Fig. 6). With 84.4%, the broad ISCO-08 categories 2–5 (i.e., professionals, technicians and associate professionals, clerical support workers, and service and sales workers) dominate in the *BuscoJobs* vacancy data. This suggests that the data allow for meaningful analysis especially of these groups of workers, which also cover a significant share of national employment (51.5%). Elementary occupations, which typically require low formal qualification levels, are underrepresented in the *BuscoJobs* data. Yet they still account for 7.3% of the vacancies (compared with 19.1% in the household survey). Furthermore, skilled agricultural occupations are absent from the *BuscoJobs* online vacancy database, while they account for 4.8% of national employment.

Overall, these patterns are not surprising as national employment distributions across occupations deviate in similar ways from other online vacancy sources studied in the literature. This includes the BurningGlass data, which capture vacancies for professional jobs in a more comprehensive way than other types of

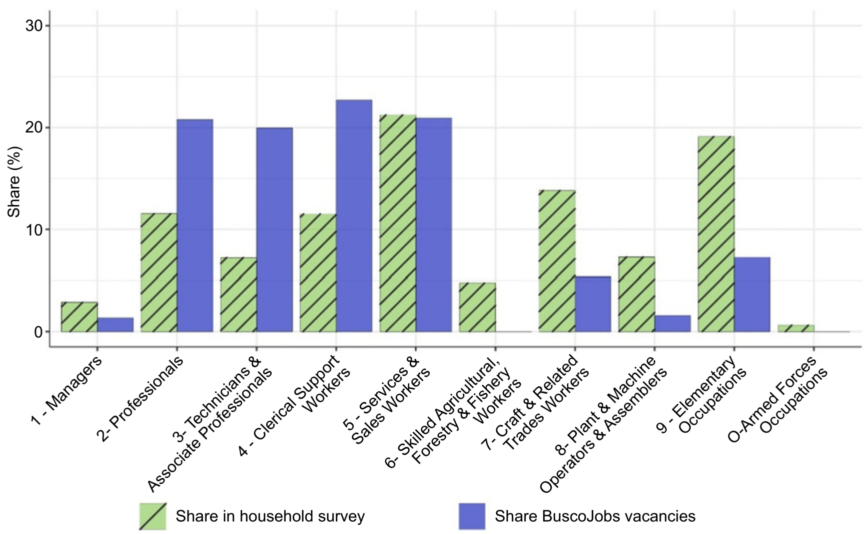

Fig. 6. Occupational Distribution, *BuscoJobs* Vacancy Database Compared to Household Survey Data, 2010–2020.

Notes: Authors' calculations based on the Uruguayan household survey and the *BuscoJobs* vacancy database. We display one-digit ISCO 08 occupational categories. The household survey analysis focuses on employed individuals only.

vacancies (Deming & Kahn, 2018; Hershbein & Kahn, 2018; ILO, 2020). We also note that the *BuscoJobs* vacancy data nevertheless cover a meaningful number of jobs requiring medium and even low formal qualification levels. Similar conclusions arise from the distribution across industrial sectors (see Table A1 and Fig. A1 in Appendix A).

As discussed above, we again look at changes in the representativeness of the *BuscoJobs* vacancy data over time to ensure the internal validity of temporal analyses. Reassuringly, Fig. 7 illustrates no major changes in the occupational distribution over time, when comparing it with the same distributions from the Uruguayan household survey. The only exceptions are again services and sales workers in 2010.

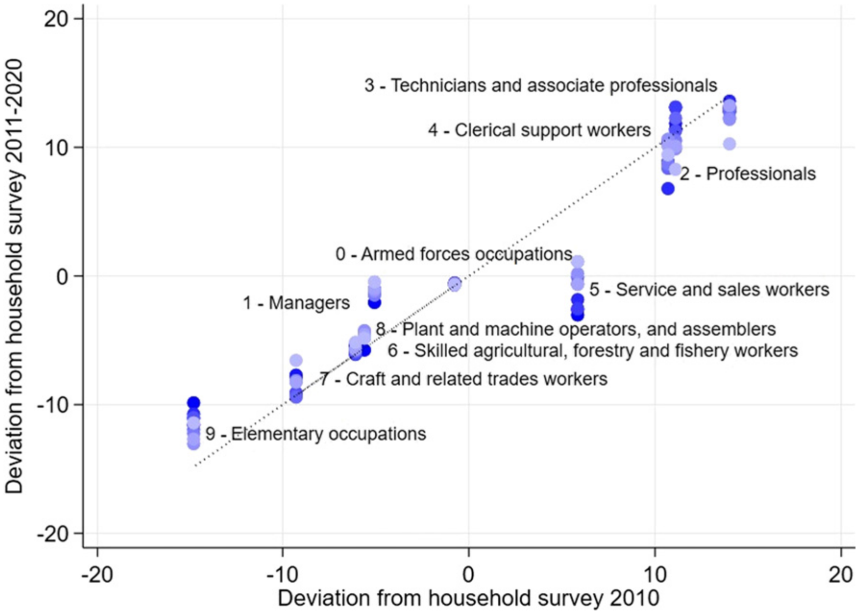

Fig. 7. Representativeness of *BuscoJobs* Occupations in Vacancy Data, Relative to Occupational Distribution in Uruguayan Household Survey Data (2010–2020).

Notes: Author's calculations based on *BuscoJobs* vacancy database and Uruguayan household survey data, where we focus on employed individuals only. Based on Hershbein and Kahn (2018), the *x*-axis illustrates the *BuscoJobs* share in occupations in 2010 minus the share in the same occupation and year in the Uruguayan household survey. The *y*-axis illustrates these differences between the two samples for each year from 2011 to 2020. Dark shades correspond to earlier years. The 45-degree line shows occupations where the shares between *BuscoJobs* and the household survey did not change from 2010. Occupations are defined according to the one-digit ISCO 08 classification.

4. EMPIRICAL IMPLEMENTATION OF THE TAXONOMY

We next describe our method for creating skills variables (Section 4.1). Then, we evaluate the method's performance in four ways: first, by assessing the share of vacancies and applicants' employment spells that the model is able to classify (Section 4.2); second, by evaluating the relevance of relying on various sources of information to devise our taxonomy (Section 4.3); third, by assessing the significance of employing our approach instead of using US O-NET data to study skills dynamics in Uruguay (Section 4.4); and fourth, by showing an application of our approach to assessing skills dynamics (Section 4.5).

4.1 Text-Mining Model

To create the skills variables in the *BuscoJobs* data, we employ a text-mining model that combines text preprocessing routines (in Python) from NLP with a rule-based classification method. The rules emanate from the taxonomy and the keywords and expressions that characterize each of the 14 skills subcategories that we predefined in Section 2.[29] Extended through pertinent synonyms, our model consists of 741 unique skills (i.e., keywords and expressions).[30] We define a skills subcategory as present in the vacancy or job spell of applicants whenever we identify at least one of its associated unique skills. We also coded related variables capturing how many unique skills pertain to each observation, as a proxy for skill intensity.

For the creation of the skills variables, we decided to rely on the open text-variables from the *BuscoJobs* data. For the vacancy data, we use the job title and vacancy description, while for the applicants' data, we focus on the applicants' description of each job spell. Compared to other variables in the database, the open-text variables contain the most detailed information on the skills demanded and supplied. These variables are available for almost all vacancies (99.9%) and most of the applicants' job spells (68.5%), in a functional manner as they contain a comparatively low share of missing or meaningless values. In addition, these variables are typically present in similar data sources and thus will allow to replicate our methodology using different databases.

Since we rely on free-text descriptions, we then preprocessed the text variables using NLP routines, with the aim of distilling information that facilitates the efficient mapping between the skills taxonomy and *BuscoJobs* data. The text data preprocessing involved translating the keywords and expressions into Spanish, tokenization, text normalization, and lemmatization of the skills variables and free text in the applicants' and vacancy databases.

We proceeded to create the skills variables, coding indicator variables for each of the 14 subcategories of skills, which take the value of one whenever a relevant keyword or expression was identified in the *BuscoJobs* data. We then refined the variable creation using synonyms through an automated web scraping method that targeted the website www.wordreference.com. Finally, we scrutinized all synonyms manually and excluded those whose meaning would have caused misleading classifications. This was especially true for some synonyms in the manual skills category, which might have mistakenly captured managerial activities. For example, "solucionar" is most relevant in the context of finding solutions but was identified as a synonym of "reparar" ("to repair"). To facilitate replications of our methodology, we provide details on its implementation, as well as the full set of initial keywords and expressions and additional synonyms in Appendix B.

4.2 Performance of the Methodological Approach Based on Assigned Number of Skills Categories

In Table 2, we present the number of vacancies and applicants' job spells, respectively, that we could classify, and the corresponding number of skills assigned.[31] Using only the initial keywords and expressions, we assigned on average 0.87 of our 14 skills subcategories to each applicant-job spell observation (column (1)), and 2.52 to each vacancy (column (3)). The model is thus able to assign skills to 53.0% of the applicants' job spells and 86.5% of vacancies (columns (1) and (3), inverse of the share with 0 skills). Once we additionally consider synonyms, the performance increases noticeably (columns (2) and (4)). The model is then able to assign an average of 1.44 skills subcategories to each applicant's job spell and 3.86 skills subcategories to each vacancy. The use of synonyms significantly increases the number of applicants' job spells that can be assigned a

Table 2. Descriptive Statistics for the Number of Assigned Skills Subcategories, Applicants' Employment Spells and Vacancies, 2010–2020.

	Applicants' Empl. Spells		Vacancy Data	
	(1)	(2)	(3)	(4)
Average number of assigned skills subcategories	0.87	1.44	2.52	3.86
(Standard deviation)	(1.11)	(1.63)	(1.79)	(2.17)
Share with 0 skills	0.470	0.364	0.135	0.057
Share with 1 skill	0.323	0.269	0.193	0.104
Share with 2 skills	0.125	0.154	0.202	0.133
Share with 3 skills	0.049	0.099	0.183	0.158
Share with 4 skills	0.021	0.055	0.143	0.162
Share with 5 skills	0.009	0.030	0.081	0.142
Share with 6 skills	0.003	0.016	0.042	0.114
Share with 7 skills	0.001	0.008	0.015	0.071
Share with 8+ skills	0.000	0.006	0.006	0.059
N	820,788	820,788	86,966	86,966

Notes: Authors' calculations based on *BuscoJobs* database. We defined dummy variables for each of the 14 skills subcategories that are set equal to one whenever at least one of the relevant unique skills is present in an observation, and equal to zero otherwise. See Section 2 for the definition of the 14 skills subcategories. Columns (1) and (3) refer to the number of assigned skills subcategories when using the initial keywords and expressions, whereas columns (2) and (4) pertain to those assigned when also considering synonyms. For the applicants' data, the level of observation is at the applicant-job spell level. The vacancy data refer to individual postings.

skill to 63.6%. While we lack a direct benchmark for job spells due to the limited availability of supply-side data on job boards and job aggregators, we believe 64% allows for a meaningful classification of skills, supporting credible analyses. This is especially pertinent given the substantial heterogeneity in the quality of information within the free-text descriptions used for coding the skills variables, particularly, the self-reported text from applicants, which does not follow any standardized format. In the case of vacancies, using synonyms even increases the classified share of observations to 94.3%. Overall, our results leave us optimistic that our methodology can be replicated in similar data sources.

Table 2 therefore allows drawing two central conclusions. First, the text-mining model performs significantly better when using a combination of keywords and synonyms to capture skills. This implies that the use of synonyms is a decisive step for implementing a skills taxonomy in online labor intermediation data. Second, the vacancy data tend to be richer than the applicants' job spells in terms of the information provided that can be used to capture skills. We note, however, that the statistics on applicants refer to a given employment spell. Once we aggregate skills over workers' employment biographies, as done in Appendix Table A2, the average number of skills subcategories increases to 2.74 (standard deviation of 2.71) per person when we use the method that is based on keywords and their synonyms. Also, the share of persons without any assigned skills subcategory is only 25.0%, compared to 36.4% of all employment spells that cannot be classified.

The primary reason why some observations cannot be classified is the brevity of the open text. The unclassified employment spells have only 5.5 words on average (i.e., with many not containing any meaningful description), whereas the number of words significantly increases for employment spells with one or several mapped skills subcategories (Table 3). This may in part be related to the fact that some applicants include descriptions that are largely self-explanatory. For example, there are observations that only include the description "biology teacher," presumably because the tasks and skills of a biology teacher are considered as being common knowledge.

Fig. 8 further highlights the importance of including synonyms in the variable-coding process. By definition, the number of classified observations increases for each of the 14 subcategories. Interestingly, this effect is not uniform across categories. In the case of the applicants' data, synonyms make a substantial difference for capturing cognitive skills (narrow sense), social skills, people management skills, and finger-dexterity skills. For the vacancy data, synonyms are additionally decisive for significantly increasing the number of identified character skills. Moreover, cognitive skills (narrow sense), customer service skills, people management skills, social skills, and finger-dexterity skills are the five categories that appear most often in the applicants' data (i.e., when accounting for synonyms). These subcategories, likewise, play the greatest role for the vacancy data, although there are some differences in the ordering. In addition, character skills, computer skills, and software skills feature prominently in the vacancy data, but this trend is less obvious in the applicants' data.

Table 3. Correlation Between the Number of Identified Skills Subcategories and the Number of Words Available in the Text Descriptions, 2010–2020.

	Applicant Empl. Spells		Vacancy Data	
Number of Assigned Skills Subcategories	Number of Words (Mean) (1)	Words per Skills Category (2)	Number of Words (Mean) (3)	Words per Skills Category (4)
0	5.5		20.2	
1	9.8	9.8	30.0	30.0
2	18.3	9.1	41.3	20.7
3	30.7	10.2	54.8	18.3
4	46.3	11.6	67.1	16.8
5	65.8	13.2	79.8	16.0
6	90.4	15.1	95.8	16.0
7	120.4	17.2	114.5	16.4
8	157.5	19.7	150.3	18.8
9	217.9	24.2	187.3	20.8
10	467.0	46.7	311.6	31.2
11	333.0	30.3	143.0	13.0
12	–	–	524.0	43.7
Correlation coefficient (Number of words and assigned categories)	0.632		0.632	

Notes: Authors' calculations based on *BuscoJobs* database. Columns (1) and (3) refer to the mean number of words needed to identify the given number of skills subcategories for the applicants' job spells and vacancies, respectively. See again Section 2 above for the definition of the skills subcategories. To facilitate interpretation, columns (2) and (4) divide the respective mean number of words by the number of assigned skills subcategories. The results are based on the initial keywords and expressions, while neglecting the synonyms. The conclusions do not change when including also the synonyms. Note also that the text length refers to preprocessed text, where stop words, etc., have already been removed, such that the text length is shorter than in the original version.

For future work, complementing our text-mining model with a prediction model could have potential. However, a prediction model requires that a proportion of the data are classified using an external source of categorization (i.e., one that is not linked to our taxonomy or its implementation). In this way, the already classified part of the data can serve as a benchmark to train the prediction model against which the results obtained from our coding of the variables can be compared to. For example, this benchmark can be a classification carried out by a group of experts, based on clearly defined criteria and a review process of the classification results. Additionally, this benchmark must be long and varied enough to "teach" the computer and train the model properly, thus making it a demanding exercise. The lack of such a benchmark for skills variables currently prevents us from employing a prediction model. In the future,

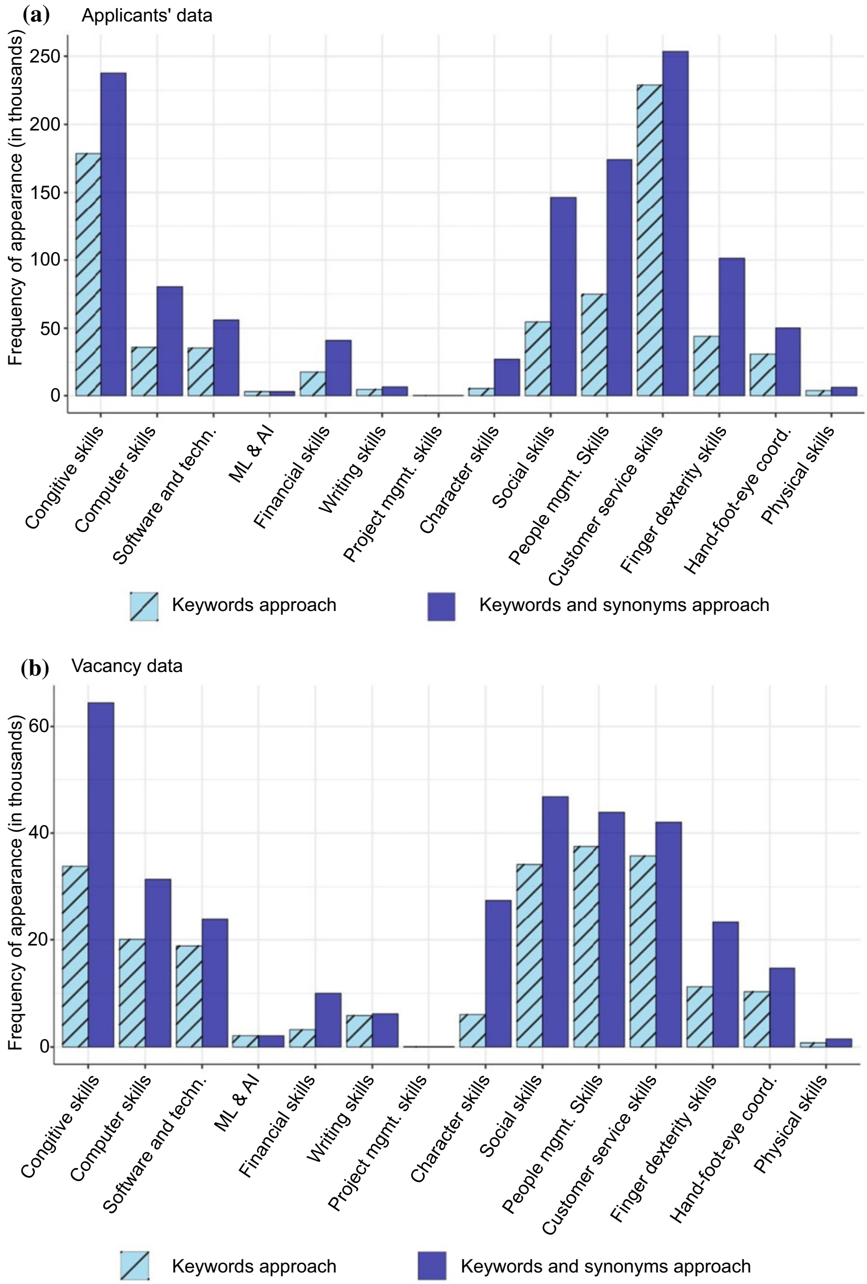

Fig. 8. Skills Distribution for Applicants' and Vacancy Data, Comparing the Initial Keyword Approach and the Extended Approach Relying on Keywords and Synonyms, All Years.

Notes: Authors' elaboration based on *BuscoJobs* database. The figure displays the frequency with which the 14 skills subcategories appear in the applicants' data (panel a) and the vacancy data (panel b), comparing the approach that relies on initial keywords and expressions with the approach that also exploits synonyms.

we will explore whether it might be possible to use a benchmark classification from O-NET Uruguay, which is currently in its pilot phase. We already followed a similar approach for coding two-digit ISCO-08 occupations for both the vacancy and job applicants' data, which is detailed in Appendix B.

4.3 Relevance of Source Types

We now investigate the usefulness of combining a comprehensive array of studies and sources in formulating our initial skills taxonomy (see Section 2). Alternatively, researchers could rely solely on prominent studies using online vacancy data. Table 4 reveals that the majority of identified keywords and expressions indeed stem from such studies (72.3% for the vacancy data and 60.6% for the applicants' data).[32]

However, supplementary keywords and expressions from non-online data sources also hold significant relevance, contributing to 17.7% of identified unique skills in vacancies and 30.6% in applicants' job spells. Additionally, the O-NET Uruguay source adds 10.0% for vacancies and 8.8% for applicants, particularly enhancing our ability to capture manual skills, which remain comparatively more important outside of Europe and the United States. This underscores the effectiveness of our approach in achieving a comprehensive taxonomy by combining a diverse range of seminal sources.

Table 4. Number of Identified Keywords/Expressions in the Vacancy and Applicants' Data, Attributable to Different Types of Sources (Absolute and % for All Years).

	Unique Skills (Keywords/Expressions) Captured (1)	Source Type 1: Online Data (2)	Source Type 2: Non-online Data (3)	Source Type 3: O-NET Uruguay (4)
Vacancies	372,879	269,603 (72.31%)	65,853 (17.66%)	37,411 (10.03%)
Applicants	1,065,305	645,809 (60.62%)	325,416 (30.55%)	94,080 (8.83%)

Notes: Authors' calculations based on *BuscoJobs* database. The table displays the number of keywords/expressions identified across vacancies and applicants' job spells and how these are attributable to different types of sources. For each vacancy or applicants' job spell, we only consider unique skills (i.e., a keyword/expression could appear multiple times but is considered only once per observation). Moreover, we only consider initial keywords/expressions, and neglect synonyms, as these are less straight-forward to attribute to source types. Source types are the following: Type 1 refers to studies based on online-data, namely DK (2018), DN (2020), HK (2012), KBHT (2016). Type 2 refers to non-online-based data, namely ALM (2003), S-O (2006), APST (2020). Type 3 refers to O-NET Uruguay, which we have used as a supplementary source. See Section 2 and Table 1 for more details.

4.4 Significance of Employing Our Approach Instead of US O-NET Data

We also assess how our approach compares to one that would have relied on imputing US O-NET data at the occupational level. For this purpose, we map US O-NET skills categories to our taxonomy. For both the US O-NET and the *BuscoJobs* applicants' data, we then compute scores that capture the relevance of cognitive, socioemotional, and manual skills at the one-digit occupational level, normalized to add to 100.[33] This comparison yields clear differences between the country-specific *BuscoJobs* results and the US-based O-NET results (Fig. 9). Across occupations, manual skills matter comparatively little in the composition of occupations of the United States. In contrast, manual skills play a larger role in the Uruguayan data, consistent with the expectation that manual skills matter more outside of high-income economies.[34] This is especially the case for plant and machine operators and assemblers, elementary occupations, and crafts and related trades workers. Accordingly, these occupations tend to have slightly lower scores in the socioemotional skills category and, especially, the cognitive skills category. Indeed, cognitive skills account for around 47%–56% of the skills composition of all occupations in the US, while in Uruguay, it varies from close to 60% for professionals to around 17% for plant and machine operators.

One consideration in this comparison is that the O-NET data are representative of occupations in the US, while we have noted in Section 3 that the *BuscoJobs* data are not fully representative of the Uruguayan labor market. However, even for occupations with high coverage in *BuscoJobs*, such as clerical support workers, there are discrepancies between the O-NET and *BuscoJobs* results: The Uruguayan data suggest a lesser prominence of cognitive skills and a greater emphasis on manual skills. Furthermore, our findings are in line with existing literature that uses the newly collected O-NET data in Uruguay (available for 22 occupations as part of the pilot phase) to analyze the composition of occupations. Velardez (2021) compares the occupational composition between the US and Uruguay using their respective country-specific O-NET data and finds that the largest differences between occupations pertain to the skills and abilities components. These findings resonate with prior research (Lewandowski et al., 2020, 2022). In summary, our findings and those from the existing literature clearly confirm one of the premises that motivated this study, namely that it is important to employ country-specific data when assessing skills dynamics beyond Europe and the United States.

4.5 Application of Our Approach to Assessing Skills Dynamics

We end this section with an illustration of how our approach can inform the analysis of skills dynamics. As such, Fig. 10 below illustrates the evolution of the supply of and demand for the skills subcategories that appeared more prominently in our analysis (see Fig. 8). The applicant trend presents the proportion of job applications containing keywords or expressions (including synonyms) related to each of the referenced skills subcategories, while the vacancy trend depicts this analogously for job vacancies. In general, we see a consistent gap between the proportion of vacancies demanding a certain skill,

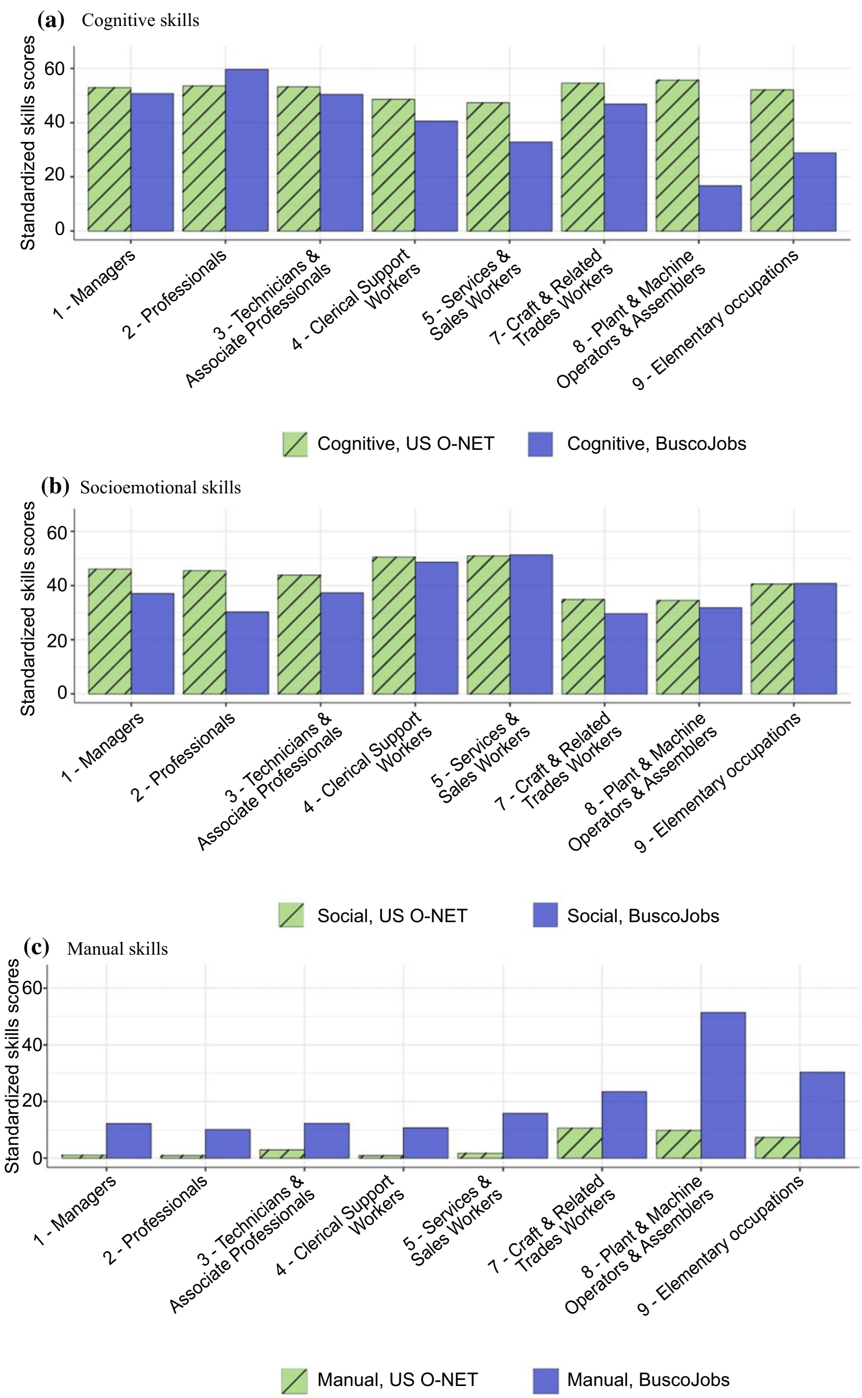

Fig. 9. Relative Importance of Cognitive, Socioemotional, and Manual Skills at the One-Digit Occupational Level, Comparing O-NET Data and *BuscoJobs* Applicant Data (2019).

Notes: Authors' elaboration based on *BuscoJobs* and O-NET databases. We focus on 2019 as this represents the most recent year prior to any distortions induced by the COVID-19 pandemic. For the *BuscoJobs* analysis, we focus on the applicants' data, transformed to an annual panel as before. Across applicants' job spells, we sum up the number of relevant keywords and expressions (including synonyms) identified in the data, per broad skills category and one-digit ISCO-08 occupation; expressed relative to the total number of unique keywords and expressions, including synonyms, that define each broad skills category. We then normalize the resulting scores for the three broad skill levels, such that their sum equals 100. The US O-NET results were obtained by first mapping US O-NET skills to the 14 skills subcategories used in this article (see Appendix Table A3 for the details on this mapping). We rely on the US O-NET database 24.1, where SOC 2010 codes were mapped to ISCO-08 four-digits using the crosswalk of the Bureau of Labour Statistics. The US O-NET importance and level scores were standardized to a scale ranging from zero to 100. The data were then aggregated to ISCO-08 occupations for each skill by taking a simple average of standardized importance and level scores. Finally, a composite score was computed by taking the product between the average standardized importance and level scores.

and the proportion of applicants with the stated ability to be using this skill in their current or during a previous employment spell (with the exception of customer service skills in 2010). This gap widens for all skills subcategories over the reference period, with the most dramatic increase realized by cognitive skills, social skills, and customer service, which rose from 23 to 45pp, 14 to 40pp, and −3 to 23pp, respectively. Thus, while the proportion of vacancies demanding certain skills has grown over time, this increase has not been met by applicants explicitly signaling the possession of these skills in their online CV.

The evolution of these gaps coincides with an increase in the number of words used to describe job vacancies over time and a concomitant decrease and subsequent stagnation in the number of words used by applicants to describe job spells over time. This notable contrast is presented in Fig. 11. The relevance of text length is corroborated by Fig. A2 in Appendix A, which presents a subset of the analysis in Fig. 10 for the period 2014–2017, where the average number of words used in both job vacancies and applications remains constant and the gaps between indicated skills supply and demand do not change significantly. The relationship between the demand for skills and the number of words used in job vacancies could be related to employers' need to reflect their increasing, more holistic demands required from prospective employees through an increase in the length of job postings. Delving into the intricacies of this relationship could be a promising direction for further research.

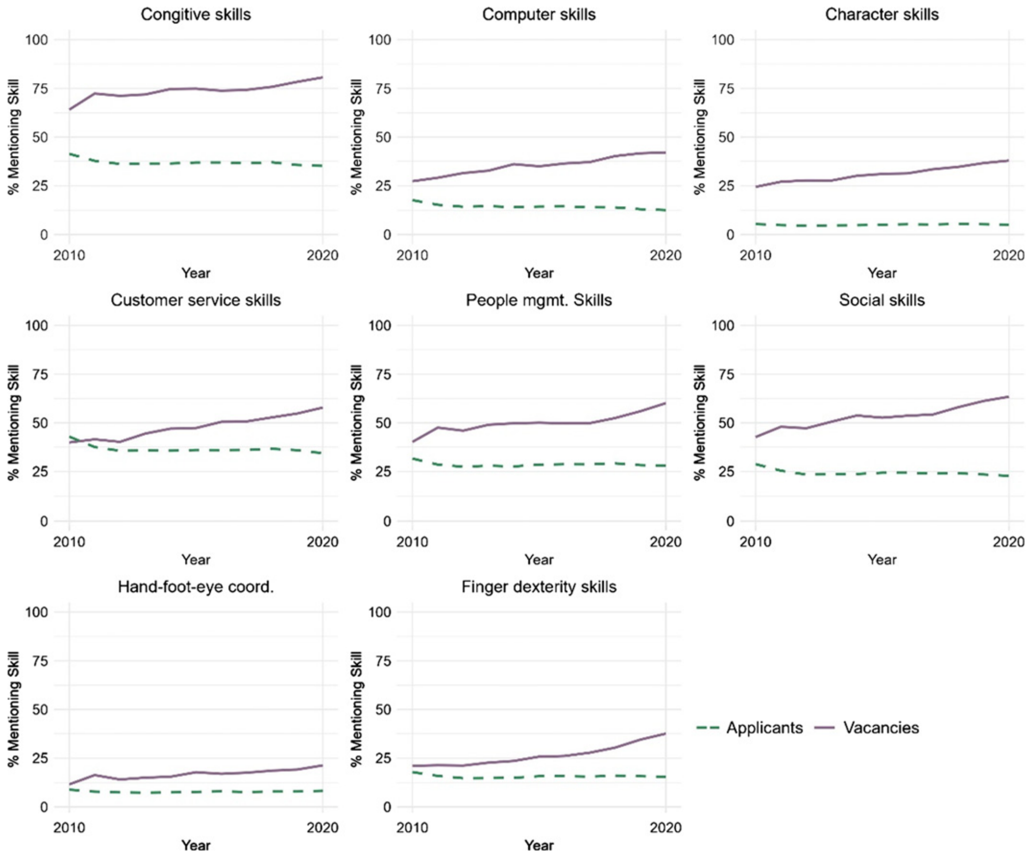

Fig. 10. Share of Applicants and Vacancies Mentioning Prominent Skills per Year, 2010–2020.

Notes: Authors' elaboration based on *BuscoJobs* database. The dataset was subset to include only applicants who had applied to at least one vacancy, and vacancies that had been applied to by at least one applicant. The applicant data were further subset to only include the years in which an applicant had applied to a job post to ensure its relevance in comparison to the vacancy data. In any given year in which an applicant applied to a job, we considered the skills exhibited during their employment in that year, in addition to the skills exhibited in previous job spells occurring after their first job application. All years in which applicants were unemployed were not considered.

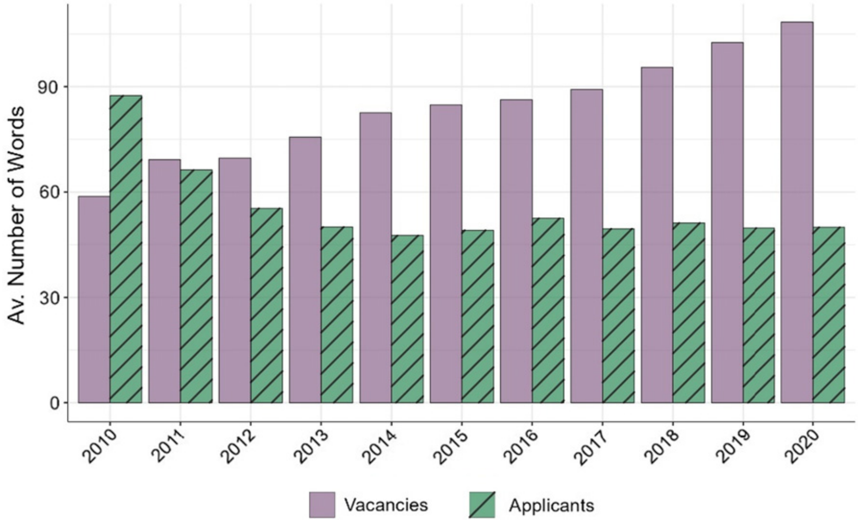

Fig. 11. Average Number of Words in Vacancies and Applicants' Job Spells per Year (2010–2020).

Notes: Authors' elaboration based on *BuscoJobs* database. The same subsets of applicants and vacancies were used as in the previous figure. The number of words in each vacancy and application refer to the number contained in the cleaned job description or application.

5. CONCLUSIONS

Many countries outside of Europe and the United States currently lack longitudinal data on skills, despite the importance of the topic for policymakers and academic debates. In this article, we assess whether data from online job vacancies and applicants' profiles, which are increasingly becoming available, can be a suitable resource for studying skills dynamics.

Drawing from the literature from the social sciences, in particular from labor economics and psychology, we derive a comprehensive yet succinct skills taxonomy. This taxonomy is designed to be adaptable to individual country-contexts and applicable to online data. It comprises the three broad categories of cognitive, socioemotional, and manual skills, along with 14 more detailed subcategories, which are defined in terms of keywords and expressions. In comparison to existing seminal skills taxonomies for online data, our taxonomy goes further by including manual skills and expanding the conceptual underpinnings surrounding socio-emotional skills. This broader approach allows us to encompass a more comprehensive spectrum of workers, which represents a critical consideration for any labor market, but is particularly significant outside of the United States and Europe.

Based on a text-mining model that combines text preprocessing routines from NLP with a rule-based skill classification method, we then develop a methodology that serves as a tool for researchers and practitioners to classify skills variables in raw text data. We implement this methodology, exploiting data from the Uruguayan job board *BuscoJobs*. We are able to classify skills requirements and skills that applicants possess for 94% of job vacancies and 64% of applicants' employment spells. This goes beyond our initial expectations, especially when considering that the implementation is based on free-text descriptions that do not necessarily follow a standardized format.

We conclude that data from online job vacancies and applicants' profiles are a promising source for analyzing skills dynamics, even in countries where job boards and job aggregators have not traditionally been prevalent. This is a relevant finding, since these data capture country-specific developments, are available in numerous countries, and offer granular and longitudinal information often for both labor demand and supply.

However, such data may not fully represent a country's entire labor force, which might require weighting techniques and/or an analytical focus on specific labor market segments. The *BuscoJobs* data are no exception to this trend. Yet, these data effectively capture job vacancies and job seekers across a range of educational levels, including lower and especially intermediate educational levels in addition to highly qualified labor. Moreover, representativeness biases in the data appear not to fluctuate substantially across time.

Most importantly, *BuscoJobs* and similar sources of data enable the study of skills dynamics in countries where such analysis would otherwise be challenging due to the limited availability of alternative data sources. Indeed, our findings reveal considerable differences in the skills composition of occupations when compared to what we would have obtained had we relied on O-NET data from the United States. We also provide a first illustration of a possible analysis by assessing the evolving trends in the supply and demand for specific skills. The analysis illustrates a consistent gap between the skills sought by job vacancies and those explicitly indicated by applicants, which increases over time. This might be related to employers' more complex demands, reflected by an increase in the length of job postings, emphasizing the need for further research to explore this relationship in depth.

Going forward, this conceptual and methodological effort, the first carried out outside Europe and the United States, opens new doors for future research on skills dynamics. Such initiative can delve more deeply into empirical questions related to the role of skills in fostering transitions to better jobs and enhancing the resilience of both firms and individuals when these are confronted with fundamental labor market transformations. Moreover, there is an opportunity to explore the skills composition of occupations and the dynamics of skills within occupations at the national level. This research can also shed light on skills mismatches between labor demand and supply, the role of skills in job application behavior and the impact of shocks and regulations, as well as test for the widespread use of other classifications of skills inspired from high-income countries.

ACKNOWLEDGMENTS

We extend our heartfelt gratitude to those who have contributed to this work. Fidel Bennett, a co-author of a previous version, played a pivotal role in the initial development of our text-mining approach. Valuable contributions were also made by Willian Adamczyk and Evgeny Gushchin. We are grateful to Claudia Abdallah, Johannes Brehm, Mariano Calderón, Angela Doku, Joana Duran-Franch, and Henry Stemmler for their outstanding research assistance throughout various stages of this article's production. Our sincere thanks go to Marcos Aguiar, Diego Alanis, Jorge Eguren, and Lucas Ng for providing data and addressing our enquiries. We are equally grateful for the invaluable insights provided by Janine Berg, David Deming, Gonzalo Graña, Cornelius Gregg, Piotr Lewandowski, Clemente Pignatti, Olga Strietska-Ilina, Michael Stops, Bolormaa Tumurchudur Klok, and Fernando Vargas. The responsibility for opinions expressed in this article rests solely with its authors, and publication does not constitute an endorsement by the International Labour Office of the opinions expressed in it. The authors have no declarations of interest to disclose.

NOTES

1. Further improvements to our methodology might be possible by training a prediction model. For example, we discuss the potential of using as a benchmark the skills classification of the O-NET Uruguay, which is currently in its pilot phase.

2. Additional studies that exploit country-level survey data on skills include Almeida et al. (2020), Ballon and Dávalos (2020), Bidisha et al. (2021), Bustelo et al. (2019), Davies and van Seventer (2020), Khurana and Mahajan (2020), Marouani et al. (2020), Maurizio and Monsalvo (2021), Valerio et al. (2016), and Yusuf and Halim (2021), among others.

3. In addition, the STEP data, which tend to cover less wealthy countries, include urban populations only.

4. To illustrate this, one may think of science and engineering associate professionals (such as aircraft pilots), machine operators, sales workers, or cleaners and helpers. These occupations differ in average qualification levels, but each occupation entails a combination of skills from at least two of the broader categories of cognitive, socioemotional, and manual skills. The same is true for most other occupations.

5. A noteworthy contribution is the framework for core skills developed in ILO (2021), which identifies core skills that improve workers' resilience vis-à-vis transformative changes in contemporary labor markets, with a view to guiding practitioners on the integration of these core skills in national education and training policies.

6. See Ministerio de Trabajo y Seguridad Social (2020) and Velardez (2021) for details on the O-NET Project Uruguay, which is in its pilot phase and has implemented a survey following the US O-NET model. It so far characterizes 22 selected occupations in Uruguay. The objective (as in the case of US O-NET) is to provide a complete and very detailed characterization of the requirements and attributes of workers within each occupation. This source is thus not tasked with creating skills categorizations that can be used for research purposes, but with providing a close to exhaustive list of skills observed in each occupation.

7. For example, "teamwork" and "collaboration" refer to the same skill, but their respective use has changed across time (Deming & Noray, 2020). Our taxonomy should encompass all keywords in situations of this kind.

8. By grounding our taxonomy in the academic literature, we aim for a taxonomy that is suitable for research. In contrast, two noteworthy initiatives have classified skills in job vacancy data with a more direct policy angle. The CEDEFOP-OVATE analysis categorizes job vacancy data according to the European ESCO scheme, focusing on countries

from the European Union and the United Kingdom (Cedefop, 2019). Stops et al. (2020) classify German vacancy data based on categories from the German BERUFENET. These approaches thus exploit pre-existing classification schemes for the European countries analyzed, which depending on the context, might require further systematic aggregation before they can be used for research purposes.

9. In the task-based model, Autor et al. (2003) define tasks as units of a discrete work activity that map to workers' skills, that is, their ability to perform a certain task (Acemoglu & Autor, 2011).

10. For example, "character skills" have been found to be highly correlated with "cognitive skills (narrow sense)." Still, the words used to characterize each set of skills are specific to each subcategory and there is no word that repeats in both subcategories.

11. See, for example: Acemoglu and Autor (2011), Atalay et al. (2020), Frey and Osborne (2017), Hardy et al. (2018), Keister and Lewandowski (2017), Spitz-Oener (2006), and Autor and Dorn (2013).

12. See, for example: Atalay et al. (2020), Arntz et al. (2016), Beaudry et al. (2016), Deming and Kahn (2018), Hardy et al. (2018), Hershbein and Kahn (2018), Spitz-Oener (2006), and Modestino et al. (2020).

13. "Personality traits" are defined by Roberts (2009) in the psychology literature as "the relatively enduring patterns of thoughts, feelings, and behaviours that reflect the tendency to respond in certain ways under certain circumstances" (in Almlund et al., 2011, p. 8). As this terminology could convey a sense of immutability − even if that was not the intent of the psychology literature − some strands of the literature prefer to avoid it (see, for example, Heckman et al., 2019; Kautz et al., 2014).

14. See for example, Almlund et al. (2011), Kautz et al. (2014), Heckman and Kautz (2012), Mischel and Shoda (1995), and Mischel and Shoda (2008). Behavioral economists are the exception (see, for example Thaler et al., 2008), who believe instead that situations have specific constraints or incentives, which determine behavior almost entirely (Almlund et al., 2011; Kautz et al., 2014).

15. Brunello and Schlotter (2011) summarize the arguments of the social psychology literature, including Cherniss et al. (1998), Boyatzis (2008), and Goleman (2000).

16. As illustrated by Heckman and Kautz (2012, p. 454), "including the measures of personality in a regression with cognitive measures explains additional variance."

17. See Almlund et al. (2011), Borghans et al. (2008), and Heckman and Kautz (2012) for a review of the relevant literature from psychology and economics. Studies in occupational psychology moreover emphasize the relationship between conscientiousness and other moderator variables, like motivation, ability, and work engagement (see Bakker et al. (2012) and references therein).

18. These categories might need to be assessed critically as far as workers' self-descriptions are concerned. Some workers may understand the importance of signaling certain character skills rather than these being a part of their personality. In addition, some traits are expected more often of women than men, and undervalued in terms of monetary returns (Grugulis & Vincent, 2009). Such gender stereotypes are confirmed in a study from China, which represents one of the few such analyses outside of high-income countries (Glewwe et al., 2022).

19. The job board was first launched in 2007 in Uruguay, then expanded to 20 Latin American countries and Spain (BuscoJobs, 2021), and is currently present in 33 countries globally (BuscoJobs Internacional, 2021). In addition to its wide coverage in Latin America, *BuscoJobs* exists in four African countries (Ghana, Kenya, Nigeria, South Africa), six countries in Asia and the Pacific (Australia, India, Indonesia, Malaysia, New Zealand, the Philippines), three European countries (Spain, Portugal, Italy), and the United States. It uses the names "BuscoJobs" (for Spanish-speaking countries), "Findojobs" (for English-speaking countries), and "Cercojobs" (in Italy).

20. For this, the *BuscoJobs* data were matched to the administrative database of enterprises produced by Uruguay's statistical institute (Instituto Nacional de Estadísticas, INE) with information until 2017, using unique firm identifiers. From 2018 forward, the

matching to the ISIC Revision 4 was done manually using the economic activity reported by firms when registering in *BuscoJobs*.

21. The job board allows users also to take a soft-skills test, which is provided by the enterprise d'Anchiano (D'Anchiano, 2021) and generates a list of soft skills possessed by each user. Furthermore, technical skills can be reported on their general profile and in association with each work experience. These variables are different from the ones we generate and are not considered in our methodology to implement the skills taxonomy, given the large number of missing values.

22. Throughout, we focus on 2019 when citing general labor market indicators, as 2019 represents the most recent year prior to any distortions induced by the COVID-19 pandemic.

23. Note that our findings do not change in any major way when we instead focus on the period of 2011–2019, that is, when we drop the initial year in which fewer people registered and when we consider only the pre-COVID-19 period.

24. The authors report a correlation coefficient of 0.71 between the two distributions. Note, however, that their reference group are jobseekers, and not the national employment distribution, as in our case.

25. For example, Deming and Noray (2020) focus on college graduates, while Deming and Kahn (2018) focus on professional jobs, noting that these are particularly well represented in online data. These authors match job vacancies from BurningGlass with survey data sources.

26. Conclusions would be the same if we were to refer to the period between 2011 and 2019 instead.

27. This comparison should ideally be complemented by other sources that capture vacancies directly, such as the Job Openings and Labor Turnover Survey (JOLTS) from the United States (Hershbein & Kahn, 2018). Unfortunately, a similar survey is not conducted by the Uruguayan National Statistical Office (INE Uruguay, 2021).

28. Note, however, that in some years, the variable has low coverage in the *BuscoJobs* data and that this is associated with annual fluctuations. For more recent years, the data coverage for this variable increases and the Montevideo area is clearly overrepresented.

29. "Keywords" refers to one-word concepts, whereas "expressions" refers to concepts with more than one word.

30. For the replication of our method in other contexts, the number of unique skills will differ from one language to another, since the synonym selection, but to some extent also the terminology used to characterize tasks, is language-specific.

31. Note that these statistics refer to those observations for which it was feasible to code skills variables, since they contained non-missing text information. For the vacancies, this was almost always the case, while for the applicants' job spells, we had to drop around 30% of employment spells that lacked actual text. This is partly a reflection of the fact that some users end up not actively using the platform, whereas those who do search and apply for jobs tend to provide richer information. As in Section 3, we focus on applicants aged 15 and older and restrict our attention to employment spells and vacancies located in Uruguay.

32. By construction, we expect source type 1 to be the most relevant. This source type was our starting point for devising the taxonomy in Section 2, given the similarity in underlying data. The other two source types were instead used as complementary sources, where we included concepts that source type 1 had not yet captured.

33. See the explanatory notes of Fig. 9 for methodological details.

34. As explained in Section 4.1, we made sure that this is not an artefact of the synonyms included, by manually excluding those synonyms that would have mistakenly assigned manual skills to managerial and related activities.

REFERENCES

Acemoglu, D., & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In O. Ashenfelter & D. Card (Eds.), *The handbook of labor economics* (Vol. 4, pp. 1043–1171). Elsevier.

Acevedo, P., Cruces, G., Paul, G., & Martinez, S. (2020). How vocational education made women better off but left men behind. *Labour Economics*, *65*, 101824.

Adhvaryu, A., Kala, N., & Nyshadham, A. (2023). Returns to on-the-job soft skills training. *Journal of Political Economy*, *131*(8), 2165–2208.

Almeida, R., Corseuil, C., & Poole, J. (2017). *The impact of digital technologies on routine tasks. Do labor policies matter?* Policy Research Working Paper 8187. World Bank.

Almeida, R., Fernandez, A., & Viollaz, M. (2020). Software adoption, employment composition, and the skill content of occupations in Chilean firms. *Journal of Development Studies*, *56*(1), 169–185.

Almlund, M., Lee Duckworth, A., Heckman, J., & Kautz, T. (2011). Chapter 1 – Personality psychology and economics. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (Vol. 4, pp. 1–181). Elsevier.

American Psychological Association. (2020). *APA dictionary of psychology*. APA Dictionary of Psychology. https://dictionary.apa.org/openness-to-experience

Arntz, M., Gregory, T., & Ulrich, Z. (2016). *The risk of automation for jobs in OECD countries: A comparative analysis*. OECD Social, Employment and Migration Working Paper. OECD.

Atalay, E., Phongthiengtham, P., Sotelo, S., & Tannenbaum, D. (2020). The evolution of work in the United States. *American Economic Journal: Applied Economics*, *12*(2), 1–34.

Autor, D. (2014). *Polanyi's paradox and the shape of employment growth*. Working Paper 20485. National Bureau of Economic Research.

Autor, D., & Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the US labor market. *The American Economic Review*, *103*(5), 1553–1597.

Autor, D., Levy, F., & Murnane, R. (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics*, *118*(4), 1279–1333.

Bakker, A., Demeranti, E., & ten Brummelhuis, L. L. (2012). Work engagement, performance, and active learning: The role of conscientiousness. *Journal of Vocational Behavior*, *80*, 555–564.

Ballon, P., & Dávalos, J. (2020). *Inequality and the changing nature of work in Peru*. WIDER Working Paper 2020/168. UNU-WIDER.

Barbarasa, E., Barrett, J., & Goldin, N. (2017). *Skills gap or signaling gap?: Insights from LinkedIn in emerging markets of Brazil, India, Indonesia, and South Africa*. Report. Solutions for Youth Employment and LinkedIn. https://www.voced.edu.au/content/ngv%3A81942

Beaudry, P., Green, D. A., & Sand, B. M. (2016). The great reversal in the demand for skill and cognitive tasks. *Journal of Labor Economics*, *34*(S1), S199–S247.

Bhorat, H., Ossthuizen, M., Lilenstein, K., & Thornton, A. (2018). *The rise of the "missing middle" in an emerging economy: The case of South Africa*. Mimeo.

Bidisha, S. H., Mahmood, T., & Rahman, M. A. (2021). *Earnings inequality and the changing nature of work: Evidence from labour force survey data of Bangladesh*. WIDER Working Paper 2021/7. UNU-WIDER.

Borghans, L., Lee Duckworth, A., Heckman, J. J., & Ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of Human Resources*, *43*(4), 972–1059.

Borghans, L., Ter Weel, B., & Weinberg, B. (2014). People skills and the labor-market outcomes of underrepresented groups. *Industrial and Labor Relations Review*, *67*(2), 287–334.

Bowles, S., Gintis, H., & Osborne, M. (2001). The determinants of earnings: A behavioral approach. *Journal of Economic Literature*, *39*(4), 1137–1176.

Boyatzis, R. E. (2008). Competencies in the 21st century. *The Journal of Management Development*, *27*(1), 5–12.

Brunello, G., & Schlotter, M. (2011). *Non cognitive skills and personality traits: Labour market relevance and their development in education & training systems*. Discussion Paper 5743. IZA – Institute of Labor Economics.

BuscoJobs. (2021). BuscoJobs. https://www.buscojobs.com.uy/paginas/quienes-somosn

BuscoJobs Internacional. (2021). BuscoJobs Internacional. https://www.buscojobs.com/

Bustelo, M., Flabbi, L., & Viollaz, M. (2019). *The gender labor gap in the digital economy*. IDB Working Paper 01056. Inter-American Development Bank.

Campos, F., Frese, M., Goldstein, M., Iacovone, L., Johnson, H. C., McKenzie, D., & Mensmann, M. (2017). Teaching personal initiative beats traditional training in boosting small business in west Africa. *Science*, *357*(6357), 1287–1290.

Carbonero, F., Davies, J., Ernst, E., Fossen, F., Samaan, D., & Sorgner, A. (2023). The impact of artificial intelligence on labor markets in developing countries: A new method with an illustration for Lao PDR and Viet Nam. *Journal of Evolutionary Economics*, *33*(3), 707–736.

Carneiro, P., & Heckman, J. J. (2005). Human capital policy. In J. J. Heckman & A. B. Krueger (Eds.), *Inequality in America: What role for human capital policies?* (Revised ed.). The MIT Press.

Caunedo, J., Keller, E., & Shin, Y. (2023). Technology and the task content of jobs across the development spectrum. *The World Bank Economic Review*, *37*(3), 479–493.

Cedefop. (2015). *Work-based learning in continuing vocational education and training: Policies and practices in Europe*. Research Paper 49. CEDEFOP: European Centre for the Development of Vocational Training. https://www.cedefop.europa.eu/en/publications/5549

Cedefop. (2019). *Online job vacancies and skills analysis: A Cedefop Pan-European approach*. Publications Office of the European Union.

Cedefop. (2021). *Perspectives on policy and practice: Tapping into the potential of big data for skills policy*. https://data.europa.eu/doi/10.2801/25160

Cherniss, C., Goleman, D., Emmerling, R., Cowan, K., & Adler, M. (1998). *Bringing emotional intelligence to the workplace*. The Consortium for Research on Emotional Intelligence in Organizations.

D'Anchiano. (2021). D'Anchiano. https://danchiano.com/

Davies, R. H., & van Seventer, D. (2020). *Labour market polarization in South Africa: A decomposition analysis*. WIDER Working Paper 2020/17. UNU-WIDER.

Deming, D., & Kahn, L. (2018). Skill requirements across firms and labor markets: Evidence from job postings for professionals. *Journal of Labor Economics*, *36*(S1), S337–S369.

Deming, D., & Noray, K. (2020). Earnings dynamics, changing job skills, and STEM careers. *Quarterly Journal of Economics*, *135*(8), 1965–2005.

Di Capua, L., Queijo, V., & Rucci, G. (2020). *Demanda de Trabajo En Uruguay: Un Análisis de Vacantes on Line*. Inter-American Development Bank. https://doi.org/10.18235/0002511

Equipos Consultores. (2020). *Uruguay: Análisis de Oferta y Demanda de Empleo a Partir de Bases de Datos a 4 Meses de La Pandemia COVID-19*. Equipos Consultores.

European Commission. (2013). *Work-based learning in Europe: Practices and policy pointers*. European Commission. https://www.wbl-toolkit.eu/index.php?id=16

Evans, D., Mason, C., Chen, H., & Reeson, A. (2023). An algorithm for predicting job vacancies using online job postings in Australia. *Humanities and Social Sciences Communications*, *10*(1), 1–9.

Fabo, B., & Kureková, L. M. (2022). *Methodological issues related to the use of online labour market data*. ILO Working Paper 68. International Labour Organization.

Frey, C., & Osborne, M. (2017). The future of employment: How susceptible are jobs to computerization? *Technological Forecasting and Social Change*, *114*, 254–280.

Gentzkow, M., Shapiro, J. M., & Taddy, M. (2019). Measuring group differences in high-dimensional choices: Method and application to congressional speech. *Econometrica*, *87*(4), 1307–1340.

Glewwe, P., Yang, S., & Zou, X. (2022). Labor market outcomes, cognitive skills, and noncognitive skills in rural China. *Journal of Economic Behavior & Organization*, *193*(January), 294–311.

Goleman, D. (2000). Leadership that gets results. *Harvard Business Review*, (March–April), 2–17.

Goos, M., Manning, A., & Salomons, A. (2014). Explaining job polarization: Routine-biased technological change and offshoring. *The American Economic Review*, *104*(8), 2509–2526.

Green, F., Ashton, D., & Felstead, A. (2001). Estimating the determinants of supply of computing, problem-solving, communication, social, and teamworking skills. *Oxford Economic Papers*, *53*(3), 406–433.

Groh, M., Krishnan, N., McKenzie, D., & Vishwanath, T. (2016). The impact of soft skills training on female youth employment: Evidence from a randomized experiment in Jordan. *IZA Journal of Labor & Development*, *5*(1), 9.

Grugulis, I., & Vincent, S. (2009). Whose skill is it anyway? "Soft" skills and polarization. *Work, Employment & Society*, *23*(4), 597–615.

Hardy, W., Lewandowski, P., Park, A., & Yang, D. (2018). *The global distribution of routine and non-routine work*. Institute for Structural Research Working Paper 5 (p. 2018). https://conference.iza.org/conference_files/WoLabConf_2018/lewandowski_p9882.pdf

Heckman, J., Jagelka, T., & Kautz, T. D. (2019). *Some contributions of economics to the study of personality*. Working Paper 26459. National Bureau of Economic Research.

Heckman, J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, *19*(4), 451–464.

Heckman, J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, *24*(3), 411–482.

Hershbein, B., & Kahn, L. B. (2018). Do recessions accelerate routine-biased technological change? Evidence from vacancy postings. *The American Economic Review*, *108*(7), 1737–1772.

Humphries, J. E., & Kosse, F. (2017). On the interpretation of non-cognitive skills – What is being measured and why it matters. *Journal of Economic Behavior & Organization*, *136*(April), 174–185.

ILO. (2014). *Las Relaciones Laborales En Uruguay: 10 Años de Cambios*. Montevideo.

ILO. (2017). *ILO Toolkit for quality apprenticeships – Vol. 1: Guide for policy makers*. International Labour Organization (ILO). http://www.ilo.org/skills/pubs/WCMS_607466/lang–en/index.htm

ILO. (2020). *The feasibility of using big data in anticipating and matching skills needs*. International Labour Office. https://www.ilo.org/wcmsp5/groups/public/—ed_emp/—emp_ent/documents/publication/wcms_759330.pdf

ILO. (2021). *Global framework on core skills for life and work in the 21st century*. ILO.

ILO. (2023). *ILOSTAT*. https://ilostat.ilo.org/

INE Uruguay. (2021). *Instituto Nacional de Estadística de Uruguay* (Vol. 2021). Instituto Nacional de Estadística de Uruguay. https://www.ine.gub.uy/

Jijkoun, V. (2016, September 2). How to detect non-exact duplicates in job postings. *Textkernel* (blog). https://www.textkernel.com/newsroom/online-job-postings-have-many-duplicates-but-how-can-you-detect-them-if-they-are-not-exact-copies-of-each-other/

Kautz, T., Heckman, J., Diris, R., Ter Weel, B., & Borghans, L. (2014). Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success. In *Commissioned report through the project on education and social progress*. OECD.

Keister, R., & Lewandowski, P. (2017). A routine transition in the digital era? The rise of routine work in Central and Eastern Europe. *Transfer: European Review of Labour and Research*, *23*(3), 263–279.

Khurana, S., & Mahajan, K. (2020). *Evolution of wage inequality in India (1983–2017): The role of occupational task content*. WIDER Working Paper 2020/167. UNU-WIDER.

Kis, V., & Catriona Windisch, H. (2018). *Making skills transparent: Recognising vocational skills acquired through workbased learning*. OECD Education Working Papers 180. OECD.

Kureková, L. M., Beblavý, M., Haita, C., & Thum, A.-E. (2016). Employers' skill preferences across Europe: Between cognitive and non-cognitive skills. *Journal of Education and Work*, *29*(6), 662–687.

Kureková, L. M., & Žilinčíková, Z. (2018). What is the value of foreign work experience for young return migrants? *International Journal of Manpower*, *39*(1), 71–92.

Lee, J., Shim, M., & Yang, H.-S. (2022). The rise of low-skill service employment: The role of dual-earner households. *Journal of Economic Behavior & Organization*, *202*(October), 255–273.

Lewandowski, P., Park, A., Hardy, W., & Du, Y. (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.

Lewandowski, P., Park, A., & Schotte, S. (2020). *The global distribution of routine and non-routine work*. WIDER Working Paper 75/2020. UNU-WIDER.

Lindqvist, E., & Vestman, R. (2011). The labor market returns to cognitive and noncognitive ability: Evidence from the Swedish enlistment. *American Economic Journal: Applied Economics*, *3*(1), 101–128.

Lise, J., & Postel-Vinay, F. (2020). Multidimensional skills, sorting, and human capital accumulation. *The American Economic Review*, *110*(8), 2328–2376.

Lo Bello, S., Sanchez Puerta, M. L., & Winkler, H. (2019). *From Ghana to America, the skill content of jobs and economic development*. Policy Research Working Paper 8758. World Bank.

Lu, Q. (2015). *The end of polarization? Technological change and employment in the U.S. labor market*. Working Paper w20485. University of Texas at Austin.

Marinescu, I., & Rathelot, R. (2018). Mismatch unemployment and the geography of job search. *American Economic Journal: Macroeconomics*, *10*(3), 42–70.

Marinescu, I., & Skandalis, D. (2021). Unemployment insurance and job search behavior. *Quarterly Journal of Economics*, *136*(2), 887–931.

Marouani, M. A., Le Minh, P., & Marshalian, M. (2020). *Jobs, earnings, and routine-task occupational change in times of revolution: The Tunisian perspective*. WIDER Working Paper 2020/171. UNU-WIDER.

Maurizio, R., & Monsalvo, A. P. (2021). *Changes in occupations and their task content*. WIDER Working Paper 15. WIDER Working Paper. UNU-WIDER.

McCrae, R. R., & Costa, P. T., Jr. (2008). Empirical and theoretical status of the five-factor model of personality traits. In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The Sage handbook of personality theory and assessment: Personality theories and models* (Vol. 1, pp. 273–294). SAGE Publications Ltd.

Ministerio de Trabajo y Seguridad Social. (2020). *Análisis Primario de Resultados de La Primera Ola de Relevamiento Del Perfil de Ocupaciones – O*Net*. Unpublished Report. Montevideo.

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, *102*(2), 246–268.

Mischel, W., & Shoda, Y. (2008). Toward a unified theory of personality: Integrating dispositions and processing dynamics within the cognitive-affective processing system. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 208–241). The Guilford Press.

Modestino, A., Shoag, D., & Ballance, J. (2020). Upskilling: Do employers demand greater skill when workers are plentiful? *The Review of Economics and Statistics*, *102*(4), 793–805.

Office for National Statistics. (2021). *Using Adzuna data to derive an indicator of weekly vacancies*. Experimental Statistics – Office for National Statistics. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/methodologies/usingadzunadatatoderiveanindicatorofweeklyvacanciesexperimentalstatistics

Reijnders, L., & de Vries, G. (2018). Technology, offshoring and the rise of non-routine jobs. *Journal of Development Economics*, *135*, 412–432.

Rodrik, D. (2018). *New technologies, global value chains, and developing economies*. Working Paper 25164. National Bureau of Economic Research.

Roys, N. A., & Taber, C. R. (2019). *Skill prices, occupations, and changes in the wage structure for low skilled men*. Working Paper w26453. National Bureau of Economic Research.

Spitz-Oener, A. (2006). Technical change, job tasks, and rising educational demands: Looking outside the wage structure. *Journal of Labor Economics*, *24*(2), 235–270.

Stops, M., Bächmann, A.-C., Glassner, R., Janser, M., Matthes, B., Metzger, L.-J., Müller, C., Seitz, J., & Hanebrink, A. (2020). *Machbarkeitsstudie Kompetenz-Kompass*. Forschungsbericht 553. Bundesministerium für Arbeit und Soziales.

Thaler, R. H., Mullanthan, S., & Kahnemann, D. (2008). Master class 2008: Putting psychology into behavioral economics (Class 6). Edge Master Class 2008. https://www.edge.org/conversation/richard_h_thaler-daniel_kahneman-sendhil_mullainathan-master-class-2008-putting. Accessed on September 22, 2021.

Valerio, A., Puerta, M. L. S., Tognatta, N., & Monroy-Taborda, S. (2016). *Are there skills payoffs in low- and middle-income countries? Empirical evidence using STEP data*. Policy Research Working Paper 7879. World Bank.

Velardez, M. O. (2021). *Análisis de distancias ocupacionales y familias de ocupaciones en el Uruguay*. Documento de Proyectos LC/TS.2021/36. Desarrollo Económico. CEPAL.

Weinberger, C. (2014). The increasing complementarity between cognitive and social skills. *The Review of Economics and Statistics*, *96*(4), 849–861.

World Bank. (2023). World Bank Databank. https://databank.worldbank.org/home.aspx

Xing, C. (2021). *The changing nature of work and earnings inequality in China*. WIDER Working Paper 2021. UNU-WIDER.

Yusuf, A. A., & Halim, P. R. (2021). *Inequality and structural transformation in the changing nature of work: The case of Indonesia*. WIDER Working Paper 81/2021. UNU-WIDER.

Zapata-Román, G. (2021). *The role of skills and tasks in changing employment trends and income inequality in Chile* WIDER Working Paper 48/2021. UNU-WIDER.

APPENDIX

APPENDIX A: ADDITIONAL TABLES

Table A1. Sectoral Distribution, *BuscoJobs* Vacancy Database Compared to Household Survey Data, 2010–2020.

	Share HS Data (1)	Share BJ Vacancies (2)
A – Agriculture, forestry, and fishing	8.9	NA
B – Mining and quarrying	0.2	0.0
C – Manufacturing	10.2	8.6
D – Electricity, gas, steam, and air conditioning supply	1.7	0.0
E – Water supply; sewerage, waste management, and remediation activities	0.7	0.1
F – Construction	7.5	2.2
G – Wholesale and retail trade; repair of motor vehicles and motorcycles	18.0	22.7
H – Transportation and storage	4.7	2.0
I – Accommodation and food service activities	3.6	2.1
J – Information and communication	2.1	11.6
K – Financial and insurance activities	2.1	2.4
L – Real estate activities	1.0	1.0
M – Professional, scientific, and technical activities	3.9	12.7
N – Administrative and support service activities	4.2	27.6
O – Public administration and defence; compulsory social security	6.4	NA
P – Education	6.6	1.5
Q – Human health and social work activities	7.3	3.3
R – Arts, entertainment, and recreation	1.6	0.9
S – Other service activities	2.7	1.3
T – Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use	6.3	NA
U – Activities of extraterritorial organizations and bodies	0.2	NA
X – Not elsewhere classified	0.1	NA

Notes: Authors' calculations based on the Uruguayan household survey (column 1) and the *BuscoJobs* vacancy database (column 2). We display the one-digit categories of ISIC Revision 4. The household survey analysis focuses on employed individuals only.

Table A2. Descriptive Statistics for the Number of Assigned Skills Subcategories, Applicant Results at the Person-Level, 2010–2020.

	(1)	(2)
Average number of assigned skills subcategories	1.87	2.74
(Standard deviation)	(2.23)	(2.71)
Share with 0 skills	0.337	0.250
Share with 1 skill	0.257	0.199
Share with 2 skills	0.143	0.141
Share with 3 skills	0.081	0.099
Share with 4 skills	0.052	0.070
Share with 5 skills	0.034	0.049
Share with 6 skills	0.024	0.037
Share with 7 skills	0.017	0.028
Share with 8+ skills	0.055	0.127
N	345,420	345,420

Notes: Authors' calculations based on *BuscoJobs* database. We defined dummy variables for each of the 14 skills subcategory that are set equal to one whenever at least one of the relevant unique skills is present in an observation, and equal to zero otherwise. See Section 2 for the definition of the 14 skills subcategories. Column (1) refers to the number of assigned skills subcategories when using the initial keywords and expressions, whereas column (2) pertains to those assigned when also considering synonyms. The applicants' data were aggregated at the person-level.

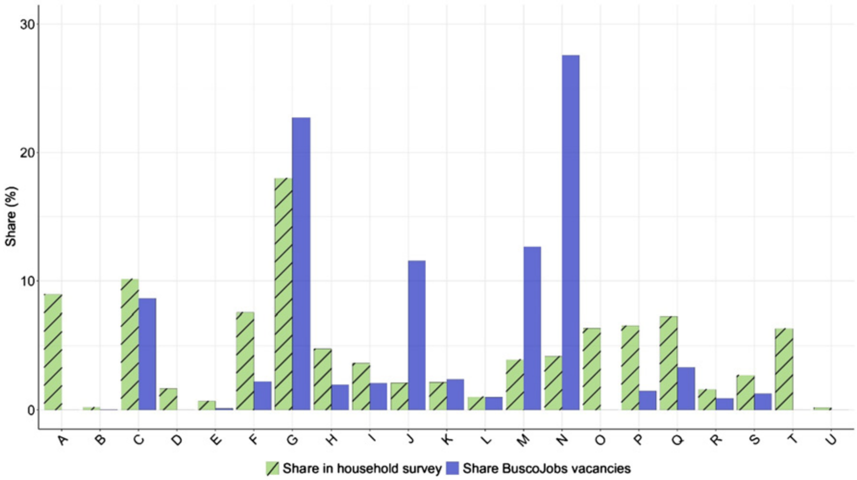

Fig. A1. Sectoral Distribution, *BuscoJobs* Vacancy Database Compared to Household Survey Data, 2010–2020.

Notes: A – Agriculture, forestry, and fishing; B – Mining and quarrying; C – Manufacturing; D – Electricity, gas, steam, and air conditioning supply; E – Water supply; sewerage, waste management, and remediation activities; F – Construction; G – Wholesale and retail trade; repair of motor vehicles and motorcycles; H – Transportation and storage; I – Accommodation and food service activities; J – Information and communication; K – Financial and insurance activities; L – Real estate activities; M – Professional, scientific, and technical activities; N – Administrative and support service activities; O – Public administration and defence; compulsory social security; P – Education; Q – Human health and social work activities; R – Arts, entertainment, and recreation; S – Other service activities; T – Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use; U – Activities of extraterritorial organizations and bodies. Authors' calculations based on the Uruguayan household survey and the *BuscoJobs* vacancy database. We display one-digit categories of ISIC Revision 4. The household survey analysis focuses on employed individuals only.

Table A3. Mapping Between US O-NET Skills and Our Skills Categories.

US O-NET (1)	Assigned Skills Subcategory that Matches Best (2)	Assigned Broad Category of Skills (3)
Monitoring	Cognitive skills (narrow sense)	Cognitive skills
Time management	Character skills	Socioemotional skills
Active learning	Character skills	Socioemotional skills
Complex problem solving	Cognitive skills (narrow sense)	Cognitive skills
Critical thinking	Cognitive skills (narrow sense)	Cognitive skills
Equipment selection	Cognitive skills (narrow sense)	Cognitive skills
Mathematics	Cognitive skills (narrow sense)	Cognitive skills
Operation & control	Cognitive skills (narrow sense)	Cognitive skills
Operations analysis	Cognitive skills (narrow sense)	Cognitive skills
Operations monitoring	Cognitive skills (narrow sense)	Cognitive skills
Programming	Software (specific) skills and technical support	Cognitive skills
Reading comprehension	Cognitive skills (narrow sense)	Cognitive skills
Science	Cognitive skills (narrow sense)	Cognitive skills
Systems analysis	Cognitive skills (narrow sense)	Cognitive skills
Systems evaluation	Cognitive skills (narrow sense)	Cognitive skills
Technology design	Cognitive skills (narrow sense)	Cognitive skills
Troubleshooting	Cognitive skills (narrow sense)	Cognitive skills
Judgment and decision making	Cognitive skills (narrow sense)	Cognitive skills
Management of financial resources	Financial skills	Cognitive skills
Equipment maintenance	Finger dexterity skills	Manual skills
Installation	Finger dexterity skills	Manual skills
Quality control analysis	Hand-foot-eye coordination skills	Manual skills
Repairing	H-foot-eye coordination skills	Manual skills
Instructing	Social skills	Socioemotional skills
Management of personnel resources	People management skills	Socioemotional skills
Management of material resources	Project management skills	Cognitive skills
Active listening	Social skills	Socioemotional skills
Coordination	Project management	Cognitive skills
Learning strategies	Social skills	Socioemotional skills
Negotiation	Social skills	Socioemotional skills
Persuasion	Customer service	Socioemotional skills
Service orientation	Customer service	Socioemotional skills
Social perceptiveness	Social skills	Socioemotional skills
Speaking	Social skills	Socioemotional skills
Writing	Writing skills	Cognitive skills

Notes: Authors' elaboration based on *BuscoJobs* and O-NET data. The table shows our mapping between US O-NET skills (column (1)) and our skills subcategories (column (2)) and broader skills categories (column (3)). We are aware that the mapping between columns (1) and (2) can be associated with certain inaccuracies. To reduce these classification errors, we display results at the higher level of aggregation (column (3)).

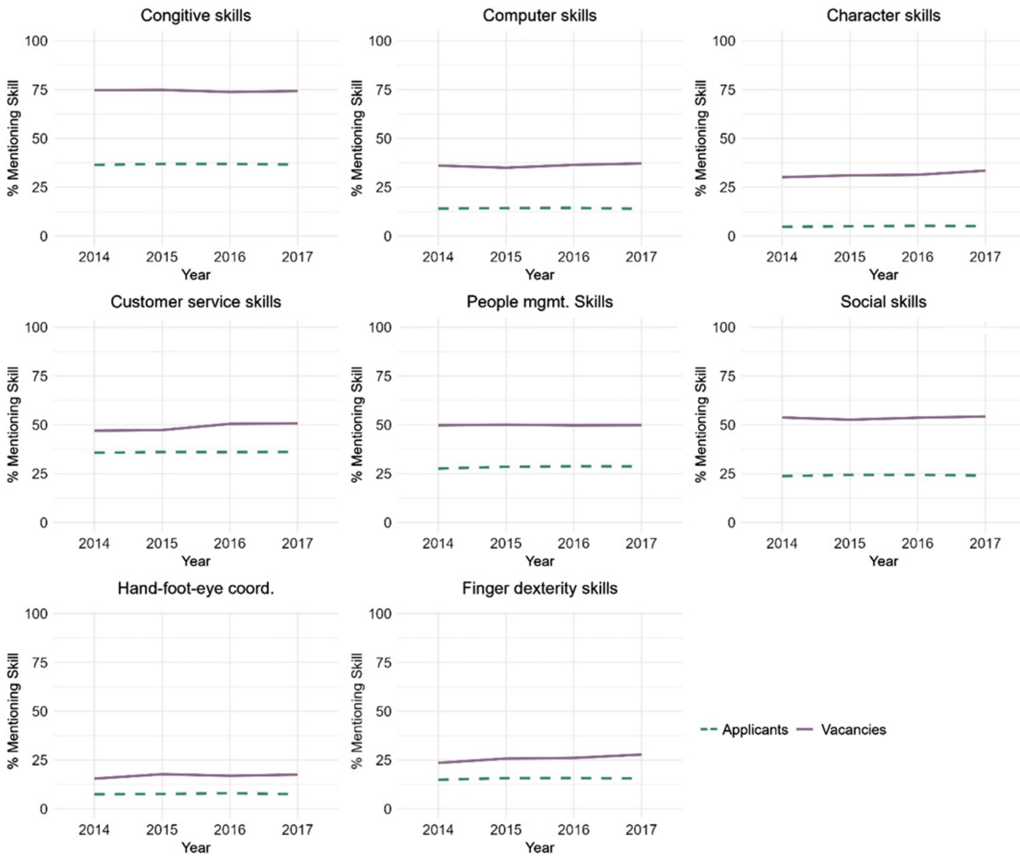

Fig. A2. Proportion of Applicants and Vacancies Including Skills per Year (2014–2017).

Notes: Authors' elaboration based on *BuscoJobs* database. The dataset was subset to include only applicants who had applied to at least one vacancy, and vacancies that had been applied to by at least one applicant. The applicant data were further subset to only include the years in which an applicant had applied to a job post to ensure its relevance in comparison to the vacancy data. In any given year in which an applicant applied to a job, we considered the skills exhibited during their employment in that year, in addition to the skills exhibited in previous job spells occurring after their first job application. All years in which applicants were unemployed were not considered.

APPENDIX B: DETAILED METHODOLOGY USED TO CREATE SKILLS AND OCCUPATIONS VARIABLES

This appendix first presents the methodology used to create skills and occupations variables. Section B.1 provides a comprehensive description of the text-mining methodology used to implement the skills taxonomy exploiting the open-text variables available in the *BuscoJobs* data. In the context of the vacancy data, these variables furnished information pertaining to job titles and vacancy descriptions, while for applicants' data, they provided descriptions of each job spell. Each of these variables required the open-text data to be structured in a manner that could be processed and interpreted by computer algorithms; as well as classified according to our pre-defined taxonomy. Section B.2 outlines the creation of ISCO-08 occupational variables. It starts by detailing a process of text processing similar to the one detailed in Section B.1. It then explains how we combined this process with a predictive model that employed *BuscoJobs'* existing classification as a benchmark.

B.1 Methodology to Create Skills Variables

The methodology used to create skills variables consists of a text-mining model/approach. Since we rely on free-text descriptions, we preprocessed the text variables using NLP routines, with the aim of distilling information that facilitates

Table B1. Dictionary of Initial Keywords and Expressions, per Skills Subcategory (in Spanish).

Skills Subcategory	Initial Keywords/Expressions
HABILIDADES COGNITIVAS (SENTIDO ESTRICTO)	Resolver problemas, investigacion, analisis, pensamiento critico, matematica, estadistica, matematica, adaptabilidad, direccion, control, planificacion, analisis datos, ingenieria datos, modelamiento datos, visualizacion datos, mineria datos, ciencia datos, analisis predictivo, modelos predictivos, analizar, disenar, reglas diseno, evaluacion, interpretacion, calculo, contabilidad, corregir, medicion, procesamiento informacion, toma decisiones, generacion ideas, memoria
HABILIDADES COMPUTACIONALES (GENERALES)	Computadora, hojas calculo, programa, software, excel, powerpoint, internet, word, outlook, office, windows
HABILIDADES COMPUTACIONALES (ESPECÍFICAS)	Lenguaje programacion, programacion, java, sql, python, instalacion de computadoras, reparacion de computadoras, mantenimiento computadoras, desarrollo web, diseno web
HABILIDADES DE APRENDIZAJE MAQUINAL E INTELIGENCIA ARTIFICIAL	Inteligencia artificial, artificial intelligence, aprendizaje maquinal, machine learning, arboles de decision, apache hadoop, redes bayesianas, automatizacion, redes neuronales, support vector

Table B1. *(Continued)*

Skills Subcategory	Initial Keywords/Expressions
	machines, svm, tensorflow, mapreduce, splunk, convolutional neural network, analisis cluster
HABILIDADES FINANCIERAS	Presupuesto, contabilidad, finanzas, costos
HABILIDADES DE ESCRITURA	Escribir, editar, reportes, propuestas
HABILIDADES DE ADMINISTRACIÓN DE PROYECTOS	Administracion proyectos
HABILIDADES DE CARÁCTER	Organizado, detallista, multitarea, puntual, energico, iniciativa propia, motivado, competente, diligente, esforzado, confiable, puntual, resistente estres, creativo, independiente
HABILIDADES SOCIALES	Comunicacion, trabajo equipo, colaboracion, negociacion, presentacion, equipo, persuasion, escucha, flexibilidad, empatia, asertividad, consejo, entretener, lobby, ensenar, interaccion, habilidades verbales
HABILIDADES DE GESTION DE PERSONAL	Supervision, liderazgo, gestion, mentoria, staff, supervision equipo, desarrollo equipo, gestion desempeno, gestion personas
HABILIDADES DE SERVICIO AL CLIENTE	Cliente, venta, paciente, persuadir, vender, publicitar, vender, comprar, pagar, servicio cliente
HABILIDADES DE DESTREZA CON LOS DEDOS	Recoleccion, clasificacion, ensamblaje, mezclar ingredientes, hornear, costura, corte, maquina tabulacion, empaque productos agricola, controlar maquinas, controlar aparatos, controlar artefactos, equipar, operar, movimientos repetitivos
HABILIDADES DE COORDINACION OJO-MANO-PIE	Atender ganado, atender animales, conducir transporte pasajeros, conducir transporte carga, pilotar aviones, podar arboles, gimnasia, deporte equilibrio, acomodar, reparar, renovar, restaurar, servir, limpiar, reaccionar tiempo, manipulacion fina
HABILIDADES FÍSICAS	Resistencia, caminar, correr, cargar peso

Notes: This table is the Spanish analogue of Table 1 in Section 2. Authors' elaboration.

the efficient mapping between the skills taxonomy and *BuscoJobs* data. We then combined this approach with a rule-based skill classification method, where the rules emanate from the predefined taxonomy that we have described in the main text (Section 2). This involved the following steps:

- *Translation of keywords and expressions:* As explained in Section 2, our skills taxonomy initially defines 14 subcategories through keywords/expressions that we identified from the existing literature on skills dynamics. Since our online data are in Spanish, we first translated the keywords/expressions from English. Appendix Table B1 above provides this translation.
- *Tokenization:* We started the text preprocessing by this pivotal initial step, which involves the systematic dissection of textual data. In this process, the

open text of the relevant variables (vacancies and applicants' job spells) was deconstructed into discrete units known as "tokens." These tokens correspond to individual words or subword elements broken down by blank spaces, including punctuation marks or segments of words, which were later removed. By transforming raw, unstructured text into a structured format, tokenization allowed for the following computational analysis.

- *Text normalization:* As is typical for data analyses of this kind (see, e.g., Gentzkow et al., 2019), we normalized the text through text-mining techniques, mainly using the Natural Language Toolkit (NLTK) library in Python. This step was similarly performed for the relevant variables in the *BuscoJobs* data and for the list of keywords/expressions in the skills taxonomy. It included: (1) lowercasing of capital letters; (2) "unidecoding" to simplify characters and delete accents that are used in Spanish (for example, changing "á" to "a"); (3) eliminating all characters that are not actual text (for example, "%," "$," "&," "NaN" or "Null" values, which are the common representation of missing data in Python); (4) eliminating stop words, which are commonly used prepositions and conjunctions that do not provide useful information by themselves (for example, the Spanish equivalents of "the," "a," "of," and "in"); and (5) eliminating other concepts that do not add value, such as names of months, days, cities, and countries and single letters or words with less than three letters.[1] Exceptions, however, were the three-letter words "sql" and "web" that identify unique skills.
- *Extended keywords and expressions in the skills taxonomy:* In order to capture conjugations of words related to the same concept (e.g., when referring to "collaboration," this could be done through the noun or through the verb "collaborate"), we relied on "lemmatization," which replaces variations of words with a common lemma. This is equivalent to extending the set of words, so they capture their various forms of expression (i.e., as a noun or verb, feminine or masculine forms, singular and plural forms, etc.). To give one example, the Spanish words "estadística," "estadístico," "estadísticas," "estadísticos" are all replaced by the form "estadistic" (i.e., "statistics"). The lemmas can be either keywords or words that are part of an expression.
- *N-gram creation in the skills taxonomy:* N-grams involve creating text segments by combining sequences of one or more words (tokens). These segments can be 2-grams, 3-grams, and n-grams (where "n" denotes the number of words included), and they are typically represented as separated elements in a list. To give an example, we had initially included in our skills taxonomy the expression "predictive analytics" as one of the terms characterizing "cognitive skills (narrow sense)." We manually identify and consider other (Spanish) versions of the same concept, such as "Análisis Predictivo de Datos" or – after applying the text-normalization described before – "analisis predictivo datos." Once we carry out the n-gram creation, a total of seven elements are added to our list: the one-grams "analisis," "predictivo," and "datos"; the 2-grams "analisis

[1]The choice of these stop words also stems from the NLTK library.

predictivo," "analisis datos," and "predictivo datos"; and the original 3-gram "analisis predictivo datos." In this way, we give our classification algorithms greater power to capture context within the free-text descriptions of vacancies and applicants' job spells. This process is carried out using the NLTK library but requires a manual revision of the results. For example, for "programming language" (included in the skills subcategory of "software (specific) skills and technical support"), we kept the token "programming" but deleted the token "language," as the latter does not pertain to software skills and would have induced an error in the classification. The number of keywords/expressions in our taxonomy was still small enough to allow for such manual processing within a reasonable amount of time.
- *N-gram creation in the vacancy and applicants' data:* We similarly transformed the relevant *BuscoJobs* variables, using a process based on the NLTK library, to distil n-grams combinations from the data. We imposed two restrictions on this automated process: (1) We identified combinations of tokens, 2-grams, and 3-grams, but neglected combinations of four and more words. We imposed this restriction as our taxonomy includes normalized expressions with a maximum length of three words (such as the above example of "analisis predictivo datos"). (2) We kept the order of words as they appear in the original text description, as otherwise the task would not be manageable for a regular server to process.

With the text organized, we proceeded to create the skills variables:

- *Implementing the skills taxonomy in the data:* For the creation of the skills variables, we coded indicator variables for each of the 14 subcategories of skills, which take the value of one whenever a relevant token or n-gram was identified in the *BuscoJobs* data. We also coded related variables capturing how many times a relevant keyword/expression appeared in the data, as a proxy for skill intensity. Here, we did not count repetitions of the same keyword/expression, but considered each keyword/expression only once per observation.
- *Refined variable creation using synonyms:* We further expanded the initial list of keywords by also accounting for their synonyms. For this, we used an automated web scraping method that targeted the website www.wordreference.com and recorded, for each initial keyword from our taxonomy, direct, or first-order, synonyms.[2] Once we had identified these additional keywords, we again performed the steps above with the extended keyword list (mainly the "lemmatization" process) and recoded the skills variables.
- *Manual correction of synonyms:* We scrutinized all synonyms manually and excluded a few that were duplicates or whose meaning would have caused misleading classifications. This was especially true for some synonyms in the

[2]This automated search for keywords only pertains to one-word keywords. In contrast, the method cannot capture more complex semantic concepts (website last accessed on 1 December 2021).

manual skills category, which might have mistakenly captured managerial activities. For example, "solucionar" is most relevant in the context of finding solutions but was identified as a synonym of "reparar" ("to repair"). For a similar reason, we manually changed "controlar" ("to control") to "controlar máquinas," "controlar aparatos," and "controlar artefactos" ("control machines" etc.). We also added relevant synonyms that we identified based on our work with the *BuscoJobs* data and previous knowledge, including accountant software programs that frequently appear in vacancy texts and are relevant for capturing financial skills. Meanwhile, software names or technical words did not have synonyms attached. To facilitate replications of our methodology, we provide the full set of initial keywords and additional synonyms in Appendix Table B2.

B.2 Methodology to Create ISCO-08 Occupations Variables[3]

In order to categorize the ISCO-08 occupation variables, we followed an analogous approach using the same open-text variables that were used to classify skills, and subsequently preprocessed the data using an NLP model. For the classification, we utilized the four key variables that contained the most valuable

Table B2. Dictionary of Synonyms for Keywords (in Spanish).

Subcategory	Keyword	Synonyms
Cognitive Skills		
S01 – Cognitive skills (narrow sense)	Resolver	Solucionar, aclarar, averiguar, descifrar, solventar
	Investigacion	Exploracion, indagacion, averiguacion, busqueda, encuesta, pesquisa, sondeo
	Analisis	Estudio, examen, observacion, comparacion, particion, separacion, distincion
	Matematico	Exacto, cabal, preciso, justo, riguroso, automatico
	Estadistico	Catastral, censual, demografico, descriptivo
	Adaptabilidad	Ductilidad, elasticidad
	Direccion	Gobierno, mando, jefatura, administracion, directivo, gerencia
	Control	Inspeccion, observacion, examen, comprobacion, registro
	Planificacion	Proyecto

[3]We gratefully acknowledge that the occupational variable creation was largely coordinated by Fidel Bennett, with contributions from Sergio Herrera and Javiera Lobos in the context of the working paper preparation.

Table B2. (*Continued*)

Subcategory	Keyword	Synonyms
	Dato	
	Ingenio	Genio, inteligencia, listeza, talento, perspicacia, capacidad, seso, lucidez, razon
	Modelamiento	
	Visualizacion	
	Ciencia	Sabiduria, sapiencia, conocimiento, erudicion
	Predictivo	
	Analizar	Examinar, estudiar, observar, averiguar, comparar, considerar, descomponer, detallar, distinguir, individualizar, separar
	Disenar, diseno	Proyectar, trazar, esbozar, esquematizar, abocetar, delinear, plantear
	Evaluacion	Valoracion, tasacion, peritaje, estimacion, apreciacion
	Interpretacion	Comentario, explicacion, analisis, apreciacion, lectura, glosa, definicion, conclusion, deduccion, entendimiento, exegesis
	Calculo	Computo
	Contabilidad	Administracion, tesoreria, caja
	Corregir	Enmendar, subsanar, reformar, rehacer, modificar, retocar, perfeccionar
	Medicion	Medida, evaluacion, calculo, sondeo
	Procesamiento	Proceso
	Informacion	
	Decision	Determinacion, resolucion
	Generacion	
	Idea	Representacion, sensacion, percepcion, imaginacion, ilusion, pensamiento, juicio, comprension, conocimiento, concepto, nocion, reflexion, designio, arquetipo, modelo
	Memoria	Recuerdo, evocacion, retentivo, rememoracion, mencion, conmemoracion
S02 – Computer (general) skills	Computador	
	Programa	Exposicion, plan, planteamiento, proyecto, sistema, linea, conducto, programacion, esquema, borrador, boceto, bosquejo, anuncio, aviso
	Software	
	Internet	
	Office	
	*	Powerpoint, word, outlook, excel, windows
S03 – Software (specific) skills and technical support	Programacion	Programa
	Desarrollo	Ordenador, calculadora, procesador, electronico
	Computadora	
	*	Java, python
S04 – Machine Learning and Artificial Intelligence	Automatizacion	

Table B2. *(Continued)*

Subcategory	Keyword	Synonyms
	Cluster	
	Intelligence	
	Maquinal	
	Machine	
	Learning	
	*	Apache, hadoop, redes, bayesianas, redes, neuronales, support, vector, machin, svm, tensorflow, mapreduce, splunk, convolutional, neural, network
S05 – Financial skills	Presupuesto	Calculo, computo, estimacion, evaluacion, partida, fondo, coste, determinacion
	Finanzas	Negocio, economia, dinero, inversion, hacienda, capital
	Costos	Coste, precio, importe, gasto, tarifa
	*	Softland, erpsap, xubio, wave, cloudbooks, nubox, bloomberg, anfix, web, sql
S06 – Writing skills	Escribir	Transcribir, manuscribir, copiar, anotar, firmar, rubricar, autografiar, trazar, caligrafiar, mecanografiar, taquigrafiar
	Editar	Publicar, imprimir, difundir, reproducir, reimprimir
	Reporter	Contener, refrenar, frenar, aplacar, apaciguar, calmar, sosegar
	Propuesta	Proposicion
S07 – Project management	Administracion proyecto	
Socioemotional Skills		
S08 – Character Skills (conscientiousness, emotional stability, and openness to experience)	Organizado	Organico, estructurado, sistematizado, planeado, ideado
	Detallista	
	Multitarea	
	Punctual	Regular, exacto, preciso, formal, metodico, escrupuloso, diligente, rapido
	Energico	Activo, decidido, resuelto, firme, eficaz, eficiente, emprendedor, dinamico, intenso, poderoso, tenaz, vigoroso, fuerte, concluyente, autoritario
	Iniciativa	Decision, dinamismo, imaginacion, idea, adelanto, advenimiento, delantera, iniciacion, proyecto
	Motivado	Originar, causar, promover, producir
	Competente	Capacitado, cualificado, apto, idoneo, entendido, experto, diestro, capaz, especialista, eficiente, eficaz, habil, preparado
	Diligente	Rapido, activo, agil, presto, resuelto, solicito, vivo, inquieto, expeditivo, listo
	Esforzado	Animoso, atrevido, bizarro, valiente, luchador, ardoroso, brioso, afanoso
	Confiable	

Table B2. *(Continued)*

Subcategory	Keyword	Synonyms
	Resistente	
	Estres	
	Creativo	
	Independiente	Individualista, autosuficiente, liberado, emancipado, libre, autogobernado, autonomo, autonomico, alejado, aislado, neutral, autarquico, imparcial
S09 – Social skills (including agreeableness and extraversion)	Comunicacion	Comunicado, mensaje, oficio, nota, misiva, escrito, telegrama, circular, aviso, saludo, notificacion
	Equipo	Conjunto, agrupacion, grupo, personal, cuadrilla, brigado, pandilla, camarillo
	Colaboracion	Cooperacion, asistencia, auxilio, ayuda, contribucion
	Negociacion	Convenio, pacto, tratar, concierto, tratado
	Presentacion	Mostrar, manifestacion, exhibicion, exposicion, aparicion
	Persuasion	Argumentacion, convencimiento, atraccion, seduccion, incitacion, sugestion
	Escuchar	Atender, percibir, enterar
	Flexibilidad	Ductilidad, elasticidad, maleabilidad, cimbreo, plasticidad
	Empatizar	
	Asertividad	
	Consejo	Recomendacion, sugerenciar, advertencia, aviso, exhortacion, asesoramiento, indicacion, invitacion, observacion, opinion, parecer
	Entretener	Distraer, divertir, agradar, amenizar, animar, recrear, alegrar, deleitar, aliviar
	Lobby	
	Ensenar	Instruir, adiestrar, educar, criar, adoctrinar, ilustrar, alfabetizar, catequizar, iniciar, explicar, aleccionar, preparar
	Interaccion	
S10 – People management skills	Supervision	Inspeccion, control, revision, verificacion, vigilancia
	Liderazgo	
	Gestion	Tramite, diligencia, papeleo, mandato, encargo, mision, cometido
	Mentorio	
	Staff	
	Desempeno	Desembargo, rescate, recuperacion, descargo
	Persona	Individuo, sujeto, semejante
S11 – Customer service skills	Cliente	Parroquiano, asiduo, comprador, consumidor, usuario
	Venta	Enajenacion, transaccion, cesion, oferta, reventar, negocio, adjudicacion, saldo, comercio, despacho, exportacion
	Paciente	

Table B2. *(Continued)*

Subcategory	Keyword	Synonyms
		Tolerante, sosegado, calmoso, tranquilo, estoico, resignado, sufrido, enfermo, flematico, manso
	Persuadir	Convencer, inducir, mover, seducir, fascinar, impresionar, atraer, inclinar, incitar, arrastrar, impulsar
	Vender	Traspasar, enajenar, expender, despachar, subastar, saldar, liquidar, exportar
	Publicitar	
	Comprar	Adquirir, obtener, mercar, comerciar, traficar, negociar, chalanear, comerciar
	Pagar	Abonar, remunerar, sufragar, apoquinar, retribuir, reembolsar, cotizar, desembolsar, compensar, recompensar, gratificar, costear, reintegrar, cancelar, liquidar
	Servicio	Encargo, prestacion, asistencia, actuacion, destino, funcion, mision, oficio, ocupacion, favor, ayuda, auxilio
Manual Skills		
S12 – Finger-dexterity skills	Recoleccion	Cosecha, siega, vendimia, acopio, acumulacion
	Clasificacion	Ordenacion, separacion, distribucion
	Ensamblaje	Ensambladura
	Mezclar	Revolver, agitar, aunar, diluir, barajar, enredar
	Ingrediente	Componente, remedio
	Hornear	Gratinar, tostar, dorar, asar, brasear, calentar, cocer, preparar
	Costura	Cosido, zurcido, calado, embaste, encaje, hilar, pespunte, vainica, bordado, dobladillo, cadeneto, sutura
	Corte	Tajo, cortadura, incision, hendidura, herido, amputacion, tajadurar, cisura, tijeretado
	Tabulacion	
	Empaque	
	Producto	Articulo, fruto, manufactura, genero, elaboracion, resultado, obra
	Agricola	
	Equipar	Abastecer, proveer, dotar, aprovisionar, surtir, suministrar, vestir
	Operar	Actuar, ejecutar, obrar, elaborar, ejercitar, manipular, efectuar
S13 – Hand-foot-eye coordination skills	Ganado	Ganaderia, reses, animal, rebano, manada, hato, vacado, yeguada
	Conducir	Transportar, acarrear, trasladar, canalizar, encauzar
	Transporte	Acarreo, traslado, porte, traslacion, carga, mudanzar, pasaje, transito, transbordo
	Pasajero	
	Carga	Fardo, bulto, embalaje, lastre
	Pilotar	Navegar

Table B2. *(Continued)*

Subcategory	Keyword	Synonyms
	Avion	
	Podar	Cortar, talar, limpiar, desmochar, cercenar, escamondar, mondar
	Gimnasia	Ejercicio, atletismo, deporte, entrenamiento, acrobacia, ejercitacion
	Deporte	Ejercicio, gimnasia
	Equilibrio	
	Acomodar	
	Reparar	Recomponer, restaurar, arreglar, remendar
	Renovar	Restaurar, reconstruir, sustituir
	Restaurar	Reparar, recomponer, renovar
	Server	
	Limpiar	Asear, adecentar, acicalar, higienizar, desinfectar, lavar, fregar, barrer, banar, duchar, enjuagar, humedecer, mojar, rociar, quitar, deshollinar, lustrar, abrillantar, pulir, frotar
	Reaccionar	
	Manipulacion	Fabricacion
S14 – Physical skills	Resistencia	Aguante, vigor, vitalidad, fuerza, energia, fortaleza, entereza, potencia
	Caminar	Andar, pasear, trotar, vagar, trasladar, deambular, transitar
	Correr	Trotar, galopar
	Cargar	Embarcar, abarrotar, lastrar, colmar, estibar, transportar, acarrear

Notes: See the text for details on how the synonyms were obtained. The fields marked by a * were manually added to capture software programs for which there are no synonyms. Moreover, we manually went through all synonyms and took out those that capture other meanings than the initial keyword and underlying concept. For example, a synonym of "to transport" ("transporte") is "conducir," which can capture a supervisory activity where someone is leading a team. This explains why some cells are empty and others have fewer synonyms than would be identified when no manual correction is done. Authors' elaboration.

open-text information: job title (vacancies only), job description (vacancies and applicants), level of education (vacancies and applicants), and hierarchical level (vacancies and applicants).

The preprocessing of the variables followed the same stages described above, namely translation; tokenization; text normalization; and stemming. The application in the two datasets is slightly different, given that they differ in the available data. For the vacancy dataset, we applied the preprocessing also on the *job title* variable, and for both the vacancy and applicants' dataset, the preprocessing was applied on the *job description* variable.

Following the preprocessing, we proceeded with the classification in three steps:

(1) a classification based on a rule-based model and NLP, similar to that carried out in our classification of skills;
(2) a classification based on a predictive model to classify ISCO-occupations at one-digit level, using an existing classification carried out by *BuscoJobs*, as a benchmark;
(3) a second machine learning classification to create ISCO-occupations at a two-digit level, based on the one-digit predictions.

Step (1) – Rule-based model and NLP

As part of process (1), we created a dictionary of keywords and expressions that appear with high frequency in the data, using the free text of the subset of both the vacancies and applicants' job-spells variables, to which the corresponding ISCO-08 code was already assigned by BuscoJobs. The keywords and expressions used in this dictionary originate from the official ISCO-08 occupations classification.[4] In specific situations, this required manual classification. This output provided us with a set of rules with clear indications on how to classify an ISCO code. Then, this dictionary was paired with the preprocessed texts achieving the classification based on the text matches that is available in Appendix Table B3.

In order to improve the classification performance, we used two auxiliary variables to clarify ambiguities from the classification: first, we used education level to differentiate between ISCO categories 2 and 3. Here, a professional or higher-education level would be classified as 2, and any other educational level as 3. Second, we used hierarchical level to identify jobs at Director or Managerial level (ISCO 1). When the rule-based model did not assign a nonambiguous classification, it used the hierarchical variable for confirmation.

Step (2) and (3) – Machine learning classification

After performing step (1), we used the approximately 5,000 observations already labelled at the four-digit level by BuscoJobs in each of the databases. For this process, we first applied a machine learning algorithm to classify observations that were not classified by step (1), or that diverged substantially from those labelled by BuscoJobs. Second, after performing several tests, we decided to combine the already labelled observations with the output from step (1) to train a predictive model to assign ISCO-08 occupations to job titles and combinations of additional open-text variables. This step (2) predicted the ISCO classification at the one-digit ISCO level.

Then, we moved to step (3) that used the input of step (2) in another machine learning process to classify observations at a two-digit level. In the case of the vacancy data, the variables used for the prediction were: job title, job description, CIIU, education level, sex, the rule-based model's predicted ISCO group, and first-stage ML predicted ISCO group. For the applicants' data, job title and CIIU were not available.

[4] https://www.ilo.org/public/spanish/bureau/stat/isco/isco08/index.htm

Table B3. Dictionary of Keywords for ISCO Classification (in Spanish).

ISCO Code	Keywords
1	Gerente zonal, gerentes zonales
2	Jefe mantenimiento proyectos, jefe planta, jefe produccion, jefe proveedores, responsable general, responsable it, responsable logistica, responsable seguridad
3	Head of production, tecnico, operario
7	Operario
11	Director ejecutivo, director fundación, gerente, gerente general
12	Accoung manager, account manager, commercial manager, director administrativo, director comercial, director financiero, director financiero, director linea negocio, director marketing, director ventas, finance manager, finance manager, gerencia comercial, gerencia general, gerente administracion, gerente administracion finanzas, gerente administrativo financiero, gerente comercial, gerente compras, gerente cuentas wholesale, gerente financiero, gerente marketing, gerente recursos humanos, gerente ventas, plant manager, sales manager, sales manager
13	Art director, director call center, director campus, director colegio, director comunicación, director consultoria, director educaciñon, director pre escolar, director prescolar, director primaria, director proyecto educacional, director seccion secundaria, director secundaria, director seguridad electronica, directr proyectos civiles, encargado regursos humanos, gerencia proyecto, gerencia tecnica despacho, gerenia recursos humanos, gerente it, gerente logistica, gerente operaciones, gerente produccion, gerente producto, gerente proyecto, gerente proyectos, human resources manager, operations manager, product manager, supply chain manager
14	Encargado pizzeria, gerente hotel, gerente local, head of marketing, manager backoofice financiero, manager collections, manager gastronomico, safety manager, sourcing manager, store manager
21	Agrimensor multinacional, analista, analista estadistica, analista matematico, arquitecta arquitecto, arquitecto ingeniero, coordinador compras, director nacional, director obra, director proyecto, director regional, disenador grafico, encargado expedicion, ing agronomo, ing civil, ing electrico, ing electrico potencia, ingeniero, ingeniero civil, ingeniero rubro ventilacion, jefe alimentos, jefe obra, jefe proyecto, jefes obra, proyecto arquitectura, supervisor obra
22	Dermatologo, enfermero, medico veterinario, odontologo
23	Directora departamento de ingles, docencia primaria, maestro kinder, maestro primaria, primary teacher, profesor comunicación, profesor expresion corporal, profesor expresion musical
24	Administracion, administracion finanzas, administrador, administrador comunicaciones, administrador ventas, administrativas atencion, analista ambiental, analista business intelligence, analista cobranzas, analista comercial, analista control gestion, analista creditos, analista planeamiento comercial, analista ventas, community manager, contador, contador publico, coordinador fuerza ventas, encargado comercio exterior, encargado logistica, encargado marketing ventas, encargado operativo comercial, gerente cuentas, jefe comercial, jefe compras, jefe contaduria, jefe logistica, jefe marketing, jefe negocios, jefe operaciones, jefe recursos humanos, jefe rrhh, jefe tienda, jefe ventas, key account manager, representante ventas, responsable area contable, responsable comercial, responsable gestion desarrollo humano, responsable gestion humana, responsable local comercial, responsable marketing, responsable rrhh, responsable sucursal, responsable tienda, riesgo operativo, subgerente tienda, supervisor tiendas, supervisora tiendas, tecnicos administracion
25	Administracion operacion servidores, administrador base datos, administrador bases datos, administrador infraestructura it, administrador linux, administrador linux experiencia, administrador redes, administrador servidores, administrador sistemas, administrador tecnologias informacion, analista funcional, analista genexus, analista

Table B3. *(Continued)*

ISCO Code	Keywords
	programador, analista programador c, analista programador delphi, analista programador experiencia, analista programador funcional, analista programador it, analista programador jpos, analista sistemas, analista sistemas junior, arquitectura it, arquitectura software, departamento it, desarrollador, desarrollador genexus, desarrollador java, desarrollador net, desarrollador php, desarrollador web, desarrolladores net, desarrolladores web, desarrolladores web senors, desarrollo microsoft sharepoint, desarrollo software java, disenador web, especialista bi, informatica, informatico, mesa ayuda tecnologia, net developer, net junior developer, net senior developer, programador, programador java, programador junior, programador php, programador web, responsable comunicación, seguridad informacion, soporte administracion desarrollo, soporte aplicaciones junior, tester, web developer
26	Abogada, abogado, bi junior developer, consultor economista, director camaras, economistas, estudiante ciencias economicas, lic trabajo social, licenciado psicologia
31	Administrador produccion, auxiliar operativo aereo, operario aviacion, tec agropecuario, tecnico soporte
32	Auxiliar enfermeria, cuidadores auxiliar enfermeria, idoneo farmacia
33	Admin ventas, administrativa, administrativa contadora, analista administracion finanzas, analista comercio exterior, analista compras, analista contable, analista contable, analista contable auditor, analista contable cajas, analista contable corporativo, analista financiero, asistente comercial, asistente comercio exterior, asistente ejecutiva gerente, asistente gerente, auxiliar comercio exterior, auxiliar compras, comercio exterior, encargado local, project manager, secreataria direccion, secretaria academica, secretaria administracion tramites, secretaria administrativa bilingue, secretaria administrativa calificada, secretaria administrativa comercial, secretaria administrativa contable, secretaria administrativa directorio, secretaria administrativa gerencia, secretaria administrativa gestoria, secretaria administrativa inmobiliaria, secretaria administrativa junior, secretaria administrativa recepcion, secretaria administrativa recepcionista, secretaria administrativa suplencia, secretaria administrativa temporaria, secretaria administrativa tramites, secretaria administrativa vendedora, secretaria administrativa ventas, secretaria administrativo contable, secretaria adminstrativa, secretaria aleman ingles, secretaria asistente administrativa, secretaria asistente contable, secretaria asistente gerencia, secretaria ejecutiva, secretaria ejecutiva administrativa, secretaria ejecutiva bilingue, secretaria ejecutiva recepcionista, secretaria ejecutiva suplencia, secretaria gerencia, secretaria gerencia general, secretario administrativo
34	Chef, chef ejecutivo
35	Administrador contenidos, analista arquitectura software, tecnico informatica, tecnico it, tecnico it junior, tecnico redes informatica, tecnico seguridad it
41	Administrativo oficina, asistente oficina tecnica, auxiliar administrativa, auxiliar administrativo, estudiantes carreras derecho, estudiantes deg deg, recepcion secretaria, recepcion secretariado, secretaria, secretaria administracion, secretaria administrativa, secretaria asistente, secretaria asistente comercial, secretaria asistente direccion, secretaria atencion telefonica, secretaria bilingue, secretaria bilingue direccion, secretaria bilingue suplencia, secretaria blingue, secretaria comercial, secretaria cubrir suplencia, secretaria direccion, secretaria direccion general, secretaria director, secretaria dpto comercial, secretaria estudio contable, secretaria estudio juridico, secretaria gestoria, secretaria junior, secretaria part time, secretaria recepcionista, secretaria telefonista, secretaria telefonista administrativa, secretaria telefonista experiencia, secretaria telefonista suplencia, telefonista, telefonista bilingue, telefonista bilingue suplencia, telefonista call center, telefonista central reservas, telefonista cubrir guardias, telefonista despachador, telefonista facturacion, telefonista recepcionista, telemarketer agente cobranzas, telemarketers
42	Administrativas inventarios, administrativo contable, administrativos contable cajeros, analista auditoria, analista costos, analista costos presupuesto, analista impuestos, asistente administrativo contable, asistente contable, auxiliar administrativa contable,

Table B3. *(Continued)*

ISCO Code	Keywords
	auxiliar administrativo contable, auxiliar contable, auxiliar contable administrativo, auxiliar contable calificado, auxiliar deposito, auxiliar inventario, encargado deposito, estudiantes contador publico, personal inventario, responsable inventario, responsable inventario armas
43	Administrativo comercial, asistente administrativo, asistente marketing, asistente recursos humanos, auxiliar biblioteca, auxiliar recursos humanos, auxiliar rrhh, tecnico prevencionista
51	Asesora belleza, ayudante cocina cocinero, camarero alimentos bebidas, camareros, cocina, cocinero, cocinero ayudante cocina, cocinero cocinera, cocinero colonia sacramento, cocinero experiencia, cocinero experiencia reposteria, cocinero hotel montevideo, cocinero jefe cocina, cocinero parrillero, cocinero planchero, cocinero repostero, consejera belleza, encargado mantenimiento, mozo repostero, oficial mantenimiento, peluquera, peluquera ayudante peluqueria, peluquera brushinista, peluquera completa, peluquera montevideo, peluquera zona pocitos, peluqueras salon belleza, peluquero, peluquero experiencia, salon belleza, sous chef, sous cheff, tecnico mantenimiento
52	Agente venta, agente venta telefonica, agente ventas, asistente ventas, auxiliar ventas, cajera, cajera local shopping, cajero, ejecutivo comercial, ejecutivo cuentas, ejecutivo ventas, merchandiser, supervisor ventas, telemarketer, vendedor, vendedor cadete, vendedor call center, vendedor experiencia, vendedor experiencia optica, vendedor externo, vendedor farmacia, vendedor idoneo farmacia, vendedor independiente, vendedor inmobiliario, vendedor interior, vendedor interno, vendedor junior, vendedor local comercial, vendedor local shopping, vendedor local tecnologia, vendedor mostrador, vendedor plaza, vendedor salon, vendedor tecnico, vendedor telefonico, vendedora, vendedores, vendedores telefonicos
53	Adscripto, adscripto orientador, adscriptos, auxiliar enfermeria cuidadora, auxiliar enfermeria cuidadores, cuidadora adulto mayor, cuidadoras, ninera
54	Guardia, guardia armado, guardia motorizado, guardia remesero armado, guardia seguridad arma, guardia seguridad armado, guardias, guardias seguridad, guardias seguridad arma
71	Albanil, armador carpintero, carpintero, oficial albanil, oficial carpinteria metalica, oficiales albaniles
72	Herrero, mecanico, mecanico automotriz, oficial herrero, oficial soldador, soldador, tecnico infraestructura
74	Electricista, electricista instalador, electromecanico, oficial electricista, tecnico electronica, tecnico instalador
75	Costurera, maestro panadero, modista, panadero, panadero confitero, panaderos, pastelero, pastelero pastelera
81	Operario armado maquinas, operario maquina, operario maquinista, operario mecanico, operario metalurgica, operario metalurgico
83	Chofer, chofer repartidor, merchandiser moto
91	Auxiliar limpieza, auxiliar mantenimiento, bachero lavandin, mercaderista reponedor, mucama, mucama importante hotel, peon limpieza, personal limpieza
94	Ayudante cocina
96	Acompanante chofer, auxiliar deposito cadete, cadete administrativo, cadete cobrador, cadete cobrador moto, cobrador cadete, guardia seguridad
8/9	Operario deposito

Notes: The table presents the final dictionary used for classification, which includes the original ISCO-08 classification and those classifications defined manually to make improvements given the existing data. Authors' elaboration.

Ultimately, we tested three classification models: Random Forest, Support Vector Machine, and Gradient Boosting. A comprehensive strategy using training (two thirds of the data) and test samples (one third) allow us to test hyperparameters and various cross-validations for each combination. Based on this, we chose Gradient Boosting to code the one- and two-digit ISCO-08 occupations in the vacancy data, and Random Forest for the applicants' data.

The strategy resulted in the successful classification of the 87,030 observations in the vacancy dataset at one-digit level, and 94.8% at the two-digit level. Among the 1,230,588 observations in the applicants' dataset, only 70% have a text description. All of them were classified at one-digit level and 97.8% at the two-digit level.