Research Brief

July 2024

Revealing New Skills Trends in Emerging Economies: The Power of Online Data and NLP techniques
Verónica Escudero and Franziska Riepl

Key points

- This brief presents an innovative approach that leverages big data of online job boards and natural language processing (NLP) techniques to extract information on transferable skills from online vacancy and applicants’ data, offering insights into skills dynamics in low- and middle-income economies.
- Harnessing insights from online data opens up vast opportunities for economic development and sustainable growth.
- Findings reveal notable skills patterns, highlighting the importance of foundational skills, including fundamental literacies and the associated cognitive skills, as well as higher-level socio-emotional skills.
- This approach empowers policymakers and researchers to deepen their understanding of skills dynamics in previously understudied contexts to inform strategies for employment and skills development initiatives.

Introduction

In a rapidly evolving job landscape marked by technological advancements and the transition to a more sustainable economy, acquiring and adapting to the right skills has become imperative. Possessing these skills can increase the resilience to global transformations and shocks, while their absence increases the risk of unfavourable labour market outcomes, leaving individuals vulnerable to poverty and exclusion. Recognising this urgency, institutions and governments have elevated skills development and lifelong learning to policy priorities (ILO 2023; OECD 2023; UNESCO UIL 2020; European Commission 2020). However, knowledge on skills dynamics outside these regions remains limited due to data constraints. Existing skills classifications and taxonomies are not easily transferable to diverse country-specific contexts. While efforts like the PIAAC and STEP skills measurement surveys (see OECD 2019 and World Bank 2014, respectively) provide some insight, they are limited in scope and coverage.

This brief presents an innovative solution to these challenges by introducing a conceptual framework and a...
methodology to leverage big data, originally developed in Escudero, Liepmann, and Podjanin (Forthcoming) and further refined in ongoing work at the ILO Research Department in preparation of the forthcoming World Employment and Social Outlook (WESO) Report on Lifelong Learning and Skills Dynamics. The conceptual framework categorizes tasks performed on the job into skills categories and subcategories, which coupled with a natural language processing (NLP) methodology allows to extract skills information from unstructured online vacancy data.

The taxonomy captures transferable skills that can be applied across various occupations, rather than technical occupation-specific skills. It comprises 15 unique skills subcategories across cognitive, socioemotional, and manual categories, which are tailored to low- and middle-income economies and adaptable to country-specific contexts. It captures all sectors and occupations, including manual labour, and can also be applied to applicant data to study skills supply and mismatch as well as the relationship between skills and job transitions. The skills subcategories are designed to be comprehensive enough for high-level analyses while encompassing a broad set of skills across various occupations.

This methodology aims to shed light on skills dynamics in previously understudied economies, as job board data is now available across numerous countries and years. These insights will empower governments, businesses, and individuals to target skills development efforts more effectively, fostering resilient economies and promoting decent work for all.

The remainder of this brief outlines the key elements of the methodology and demonstrates its implementation with data from Uruguay and South Africa. The taxonomy and accompanying methodology form the basis for future work on skills within the ILO, with further analyses based on data from Brazil and the Russian Federation already under way.

### Methodology: Taxonomy and implementation

#### Taxonomy

The underlying building block of the approach is a skills taxonomy encompassing job-specific tasks and personal attributes, categorised into cognitive, socioemotional and manual skills. Within each broad category, there are 15 more nuanced skills subcategories (14 in the original concept) (Table 1). Drawing from labour economics (particularly Deming and Kahn 2018) and psychology literature (e.g., Almlund et al. 2011), the taxonomy expands existing efforts to accommodate individual country-contexts, including low- and middle-income countries, and is designed to be applicable to online data. Notably, the taxonomy includes three types of manual skills (i.e., finger dexterity, hand-foot-eye coordination, and physical skills) that are often overlooked in analyses focused on high-income countries. The taxonomy focuses on transferable skills that are applicable across occupations, rather than on occupation-specific skills that are thus non-transferable to other occupations.

For each skill subcategory, a set of keywords is selected based on existing taxonomies and seminal studies (Almlund et al. 2011; Autor, Levy, and Murnane 2003; Atalay et al. 2020; Deming and Noray 2020; Kureková et al. 2016; Heckman and Kautz 2012; Hershbein and Kahn 2018). This list of keywords is supplemented with terms relevant to emerging countries, based on a country-level skills classification (O-NET Uruguay), and further extended with synonyms of the original words obtained from an online synonym generator (www.wordreference.com). The final dictionary for Uruguay contains 669 keywords and expressions and that of South Africa 1,686, given the larger set of synonyms available in English.

---

1 Credit is due to Willian Adamczyk and Isaure Delaporte of the ILO Research Department, and Trang Luu of the University of Geneva, for their contribution to the data analysis and preparation of the figures. We are grateful for comments and feedback provided by our ILO colleagues Simon Böhmer (Research Department) and Pedro Moreno da Fonseca (Skills Branch), and an anonymous referee of the science-policy brief series for the 2024 UN Science, Technology and Innovation Forum.

2 Note that the larger number of synonyms does not substantially affect the number of matches as they only contribute to a small share of matches overall—12.6 per cent in South Africa and 20.6 per cent in Uruguay. For details, see Adamczyk et al. (Forthcoming).
For South Africa, data is sourced from the job aggregator Adzuna, which collects postings from various websites and standardizes them on its platform. Adzuna covers about 85 per cent of the South African online labour market, providing a comprehensive dataset comprising 5,262,557 observations spanning from 2016 to 2021.

Using online data offers significant advantages over traditional sources: it is widely available across countries, providing high-frequency updates and high level of granularity. Moreover, it captures a diverse set of jobs, including informal labour that is often omitted from official statistics. Additionally, it offers rich insights into skills requirements, wages, non-wage amenities and, in some cases, hiring and application behaviour (Escudero, Liepmann, and Podjanin Forthcoming; Fabo and Kureková 2022).

Nonetheless, it is crucial to acknowledge the limitations of such type of data. One significant drawback is representativeness. Since these data are not derived from random sampling, caution must be exercised in generalising findings to the broader working-age population or all firms (Escudero, Liepmann, and Podjanin Forthcoming). Online job data tend to be skewed towards higher-skilled occupations, with limited representation from rural areas, small firms, low-skilled jobs and informal employment sectors (Hershbein and Kahn 2018; Fabo and Kureková 2022). Despite these limitations, the data from Uruguay and South Africa, while underrepresenting lower-skilled occupations (ISCO codes 6 to 9) compared to census data, still provide sufficient representation to draw meaningful conclusions for these groups (Escudero, Liepmann, and Podjanin Forthcoming).

Implementation

The taxonomy is adapted to the relevant language and context through literal translation and by adding context-specific words where needed. It is then applied to online vacancy data from Uruguay and South Africa using NLP techniques. First, open-text descriptions of job postings are pre-processed to fit the structured format of the skills dictionary. A skill subcategory is identified as present if at least one keyword or expression from the dictionary appears in the vacancy posting. For applicant data (available for Uruguay), the employment histories uploaded by applicants are analysed using the same NLP procedure applied to job postings, leveraging the
ILO Brief

Revealing New Skills Trends in Emerging Economies: The Power of Online Data and NLP techniques

descriptions of all job spells reported before an application. While applicants’ descriptions of their past jobs tend to be shorter than vacancy postings for similar jobs, and thus capture less skills on average, the taxonomy nonetheless succeeds in capturing a significant number of skills (Escudero, Liepmann, and Podjanin Forthcoming; Adamczyk et al. Forthcoming).4

This implementation demonstrates how advancements in NLP have enhanced access to unstructured data, facilitating exploration of labour and development questions that were previously unattainable.

Findings and possible applications

After providing a general overview of the methodology, we will now illustrate how it can be used to study specific labour market-related questions.

In countries where job board data include information on vacancies and applicants’ job histories, such as Uruguay, this methodology can be used to analyse trends in the skills supplied by jobseekers and those demanded by firms. The skills patterns observed in job postings and applicants’ CVs for Uruguay reveal notable similarities and some key differences (Figure 1). Vacancy postings generally require more skills than applicants mention in the description of their job spells, except for customer service skills, which are more frequently mentioned by applicants. Core cognitive, customer service and social skills are most frequently mentioned in both vacancy and applicants’ data. Conversely, machine learning and AI (ML&AI), project and process management, and physical as well as finger dexterity skills are less prevalent for both applicants and vacancies.

These patterns reflect labour market characteristics, but also the nature of the data. Cognitive and social skills are widely applicable and crucial in white-collar clerical and service occupations, which are dominant in online data. Customer service skills are important for many jobs, not only for entry-level service positions, but also for client-oriented managerial and professional roles. Meanwhile, specialized skills such as project management and ML&AI are confined to certain jobs, such as business managers and software engineers, that require extensive education and additional complementary skills. The small presence of physical skills is related to the typical underrepresentation of elementary occupations in online data, particularly within the agricultural sector, where jobs are primarily advertised via personal networks or offline postings.

Figure 1. Skills identified in vacancies and applicants’ job histories in Uruguay, 2010-2023.

Notes: Skills are expressed as the share of all vacancies and job spells, respectively.
Data source: BuscoJobs Uruguay.

Therefore, when analysing online job board data, it is crucial to consider data coverage: dynamics within well-represented sectors and occupations can be studied in great detail, but caution is needed regarding statements that generalize the findings to the entire economy. Reassuringly, similar analyses for Brazil, the Russian Federation and South Africa confirm that the overall ordering of skills frequencies is remarkably stable across countries.

These findings serve as a robust foundation for further analyses on skills dynamics, mismatches between required and available skills in the workforce, skills complementarities, and the effect of skills on labour market outcomes. These analyses can be conducted at both the aggregate level, as demonstrated in the examples below, or for specific subgroups such as by occupation, firm size,

---

4 For Uruguay, 64 per cent of job spells and 94 per cent of vacancies can be assigned at least one skill, with an average of 1.44 skills per job advert and 3.86 skills per vacancy. To increase the comparability and quality of the data, we only use applicants with at least one skill in their job history and vacancies which require at least one skill for the analyses.
or by applicants’ age and gender. To illustrate the insights derived from such detailed analyses, Figure 2 presents the skills demand in Uruguay disaggregated by occupation. While some trends observed in the aggregate analysis (Figure 1) are consistent across all subgroups – such as the prominence of customer service and character skills – there are also occupation-specific insights to be gleaned. Notably, the demand for both core and sophisticated cognitive skills varies significantly; for example, while 36 per cent of professional positions require sophisticated cognitive skills, this figure drops to just 7 per cent for plant and machine operators and elementary jobs.

**Figure 2. Skills composition by ISCO 1-digit occupation in Uruguay, 2010-23.**

<table>
<thead>
<tr>
<th>Skill categories</th>
<th>1 - Managers</th>
<th>2 - Professionals</th>
<th>3 - Technicians and Associate Professionals</th>
<th>4 - Clerical Support Workers</th>
<th>5 - Services and Sales Workers</th>
<th>6 - Food and Catering Workers</th>
<th>7 - Craft and related trade workers</th>
<th>8 - Plant and Machine Operators and Assemblers</th>
<th>9 - Elementary Occupations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core cognitive skills</td>
<td>68%</td>
<td>52%</td>
<td>51%</td>
<td>48%</td>
<td>29%</td>
<td>32%</td>
<td>17%</td>
<td>19%</td>
<td></td>
</tr>
<tr>
<td>Sophisticated cognitive skills</td>
<td>30%</td>
<td>36%</td>
<td>29%</td>
<td>20%</td>
<td>12%</td>
<td>19%</td>
<td>7%</td>
<td>7%</td>
<td></td>
</tr>
<tr>
<td>General computer skills</td>
<td>38%</td>
<td>45%</td>
<td>47%</td>
<td>49%</td>
<td>24%</td>
<td>43%</td>
<td>11%</td>
<td>18%</td>
<td></td>
</tr>
<tr>
<td>Software (specific) skills &amp; technical support</td>
<td>35%</td>
<td>36%</td>
<td>28%</td>
<td>15%</td>
<td>19%</td>
<td>18%</td>
<td>10%</td>
<td>12%</td>
<td></td>
</tr>
<tr>
<td>Machine learning &amp; AI skills</td>
<td>1%</td>
<td>2%</td>
<td>2%</td>
<td>0%</td>
<td>1%</td>
<td>3%</td>
<td>0%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>Financial skills</td>
<td>36%</td>
<td>17%</td>
<td>15%</td>
<td>13%</td>
<td>13%</td>
<td>5%</td>
<td>5%</td>
<td>5%</td>
<td></td>
</tr>
<tr>
<td>Writing skills</td>
<td>18%</td>
<td>9%</td>
<td>8%</td>
<td>7%</td>
<td>8%</td>
<td>7%</td>
<td>5%</td>
<td>6%</td>
<td></td>
</tr>
<tr>
<td>Project &amp; process management skills</td>
<td>7%</td>
<td>3%</td>
<td>3%</td>
<td>3%</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td></td>
</tr>
<tr>
<td>Character skills</td>
<td>36%</td>
<td>27%</td>
<td>33%</td>
<td>35%</td>
<td>35%</td>
<td>28%</td>
<td>24%</td>
<td>32%</td>
<td></td>
</tr>
<tr>
<td>Social skills</td>
<td>63%</td>
<td>50%</td>
<td>52%</td>
<td>45%</td>
<td>50%</td>
<td>49%</td>
<td>30%</td>
<td>37%</td>
<td></td>
</tr>
<tr>
<td>People management skills</td>
<td>46%</td>
<td>14%</td>
<td>16%</td>
<td>6%</td>
<td>10%</td>
<td>10%</td>
<td>10%</td>
<td>5%</td>
<td></td>
</tr>
<tr>
<td>Customer service skills</td>
<td>61%</td>
<td>43%</td>
<td>54%</td>
<td>57%</td>
<td>77%</td>
<td>41%</td>
<td>40%</td>
<td>44%</td>
<td></td>
</tr>
<tr>
<td>Finger-dexterity skills</td>
<td>5%</td>
<td>4%</td>
<td>5%</td>
<td>3%</td>
<td>4%</td>
<td>14%</td>
<td>6%</td>
<td>8%</td>
<td></td>
</tr>
<tr>
<td>Hand-foot-eye coordination skills</td>
<td>12%</td>
<td>10%</td>
<td>14%</td>
<td>12%</td>
<td>16%</td>
<td>28%</td>
<td>57%</td>
<td>31%</td>
<td></td>
</tr>
<tr>
<td>Physical skills</td>
<td>8%</td>
<td>3%</td>
<td>4%</td>
<td>4%</td>
<td>3%</td>
<td>5%</td>
<td>21%</td>
<td>12%</td>
<td></td>
</tr>
</tbody>
</table>

| No. of vacancies                  | 1,458       | 40,423            | 22,448                                      | 30,696                        | 29,330                        | 5,288                           | 1,950                               | 7,713                                        |                               |

Notes: The ISCO category 6 has been excluded from the analysis due to the low number of vacancies observed in the data. Data source: BuscoJobs Uruguay.

Next, we illustrate how the data and methodology can be used to investigate skills dynamics. Figure 3 presents an analysis of labour market tightness, which could aid in identifying priorities for skills development policies by pinpointing areas experiencing skill shortages. The data reveals a U-shaped trend, with tightness initially declining alongside a deterioration of economic conditions and rise in unemployment in Uruguay from 2015 onwards, and then increasing again as firms post more vacancies after the COVID-19 pandemic.

Notably, tightness exceeds a value of one for character and writing skills, indicating that there are more vacancies requiring those skills than applicants supplying them. For other skills, there is an adequate pool of applicants. Still, tightness has increased in the most recent years for all skills analysed. It is imperative to recognize that the availability of applicants with specific skills does not guarantee ease in filling vacancies. Skills are often complementary, as illustrated below, and firms seek candidates with a blend of transferable skills (as captured here), job- and firm-specific skills, and personal attributes that align with the role. It is noteworthy that our online data suggests heightened tightness for more general skills, which are highly sought after by employers but not easily measured and, consequently, underreported by applicants.
While many workers may possess character skills like trustworthiness, punctuality, and organization, they do not necessarily showcase these abilities when applying for a job. This underscores the importance of understanding job search and recruiting behaviours to enhance labour market matching procedures.

Most countries do not have job-board data that includes information on both vacancies and the applicants searching for jobs. Vacancy data is the most common type of information available. While this data captures only the demand side, it can be very useful, especially when the dataset is sufficiently large, as with job vacancy aggregators. This data can provide valuable insights into labour market dynamics and skill demands, identify emerging trends in required skills and inform targeted skills development programs.

Figure 4 presents an example for South Africa, illustrating how exploring the relationship between the skills required by firms and the wages advertised in vacancies can offer insights into salary distributions and the value of different skill sets in the job market. The spider web charts show the average skills profile of jobs in the top and bottom 20 per cent of the salary distribution for South Africa for all occupations (panel A) and for crafts and related trade workers – ISCO 7 (panel B). We observe in panel A that high-paying job ads require a broader range of skills, particularly social, core cognitive and financial skills, while low-paying job ads prioritize coordination and customer service skills. The prevalence of general abilities, such as cognitive core and social skills, in both high- and low-paying jobs highlights their importance as foundational competencies that complement more specialized skills like software or ML&AI.

The granularity of online job board data can be leveraged to make specific recommendations for various occupations based on this type of analysis. For example, panel B reports radar charts for crafts and related trades workers (ISCO group 7) in South Africa, revealing that those job ads advertising higher wages on average require higher people management and financial skills, likely because they are associated with positions where crafts workers have managerial duties. To provide upward career opportunities for crafts workers, governments could cooperate with workers’ and employers’ associations active in this sector to offer management and financial training to interested workers. Similar insights can be derived for other occupational or demographic groups, and the recommendations can dynamically respond to changing labour market conditions as new data becomes available.
Revealing New Skills Trends in Emerging Economies: The Power of Online Data and NLP techniques

Figure 4. Radar chart of skills by the top and bottom quintile of posted wages in South Africa, 2016-2021.

Panel A: Full dataset
Panel B: Crafts and related trades workers (ISCO 7)

Notes: The figures show the average skills profile of jobs in the top (purple) and bottom (orange) 20 per cent of the salary distribution for South Africa. The further away from the centre a marker is, the more important is a given skill. Wages are measured as the minimum salaries advertised in the vacancies.

Data source: Adzuna, South Africa.

Figure 5. Skills networks for the top and bottom quintile of posted wages in South Africa, 2016-2021

Panel A: Top quintile of posted wages
Panel B: Bottom quintile of posted wages

Notes: The thicker the line between skills, the more interconnected they are. This means that job adverts require both skills concurrently. Wage quintiles are calculated based on the minimum salaries advertised in the vacancies.

Data source: Adzuna, South Africa.
Finally, Figure 5 explores the systematic interconnections between skills, represented by the density of lines between skills. The figure reveals that there are strong links especially between core cognitive and social skills, irrespective of the wage level advertised. However, there are also significant differences depending on the wage level of the job ad. Panel A, representing high-paying job ads, shows stronger connections among more technical capabilities like financial and writing skills, whereas panel B, representing low-paying job ads, shows additional connections between customer service and character skills.

In a more formal analysis, we assess the contribution of required skills to the wages posted in job ads, finding that technical skills such as software skills, ML&AI, financial and management skills are more strongly associated with higher wages. However, these skills are not required in isolation. Between 75 per cent and 90 per cent of the vacancies requiring these higher-paying skills also require social skills and core cognitive skills, respectively. Customer service and character skills are also frequently demanded in these adverts. This highlights the importance of a well-rounded skill set, combining technical expertise with essential social and cognitive abilities, to achieve higher wage levels.

**Policy implications**

The empirical findings based on the taxonomy and NLP methodology developed by Escudero, Liepmann, and Podjanin (Forthcoming), and expanded in the preparation of the next WESO Report on Lifelong Learning and Skills Dynamics for 2026 (Adamczyk et al. Forthcoming), illustrate the potential analyses possible with this approach and provide valuable insights with significant policy implications. The analysis underscores the persistent importance of social and character skills, alongside core cognitive skills, which are often undervalued despite their critical role in fostering collaboration, adaptability, and integration into modern workplaces. Policymakers should prioritize the development of foundational skills and higher-level socio-emotional skills to enhance workforce readiness and promote career advancement. Recognizing their relevance across various job levels and occupations, establishing a robust foundation in these broadly applicable skills enable individuals to better adapt to evolving job requirements, with additional technical skills building upon these core competencies.

The freely available taxonomy presented here provides a valuable resource for policymakers and researchers worldwide, with the flexibility to be adapted to local contexts through translation and the inclusion of country-specific keywords and expressions. This adaptability is made possible by advancements in NLP that have revolutionized the accessibility of unstructured information, enabling the exploration of previously unattainable labour and development questions.

Further analyses regarding complementarities between different skills as well as the skill content of occupations, combined with insights on individuals’ labour market trajectories, can enable the identification of individual pathways for skill development and career progression. The granularity of online job board data allows for specific recommendations for various occupations and demographic groups, helping to tailor training programs, and reskilling and upskilling initiatives to the unique needs of different segments of the workforce. Moreover, such targeted approaches can dynamically respond to changing labour market conditions as new data becomes available.

To conclude, harnessing the insights derived from online data, together with a comprehensive skill taxonomy geared towards low- and middle-income economies, offers vast opportunities to deepen our understanding of labour market and skills dynamics. Through the analysis of online job board data, policymakers, researchers, and workers’ and enterprises’ organizations can now gain deeper insights into skills dynamics and their implications for labour market outcomes and, more broadly, for economic development.
References


