SKILLS FOR JOBS 2022

Mapping skill requirements in occupations based on job postings data

OECD
BETTER POLICIES FOR BETTER LIVES

Contents

Introduction

This note presents the new methodology used to update the Skills for Jobs database in 2022. The Skills for Jobs indicators are constructed in two steps. First, surpluses and shortages at the level of occupations are computed for each country. Second, these occupational surpluses/shortages are transformed into skill surpluses/shortages using information on skill requirements in each occupation.

Until now, this second step used data from O*NET, a database created in 1998 by the U.S. Department of Labor building on its predecessor the Dictionary of Occupational Titles (DOT), and updated on a regular basis. O*NET contains a wealth of information on occupations, including knowledge, skills, and abilities needed to work in each of the almost 1 000 occupations. Most of this information is collected from job incumbents and occupational experts through surveys. Knowledge, skill, and ability requirements of occupations are measured both in terms of importance and level. The former indicates whether the particular skill or ability is important to perform the job. The latter indicates the level of mastery or proficiency in that skill or ability needed for the job. The previous version of the Skills for Jobs database used importance and level requirements (and more specifically a multiplication of the two indexes) to infer knowledge, skill and ability requirements by occupation. The latest version of the database contains 33 knowledge types, 35 skills and 52 abilities. O*NET defines skills as "developed capacities that facilitate learning or the more rapid acquisition of knowledge", abilities as "enduring attributes of the individual that influence performance", and knowledge as "organized sets of principles and facts applying in general domains". The O*NET database has been used extensively in labour market research. For example, Deming (2017) uses the database to measure the extent to which occupations use non-routine analytical tasks, service tasks, and social skills.

As O*NET is developed by the Bureau of Labor Statistics in the United States, it is geared towards the occupational content of jobs in the U.S. labour market. Yet, skill content of occupations might differ across countries. Furthermore, updates of O*NET happen regularly but only for approximately 10% of occupations each year while the skill content of each occupation is likely to evolve more rapidly over time, notably as the result of megatrend such as automation, globalisation, and ageing. It is thus crucial to use more timely data on the skill content of occupations, and data which do not only reflect the U.S. context. Lastly, the O*NET-based version of the Skills for Jobs Database could not identify imbalances in digital skills.

A relatively new dataset commercialised by Emsi Burning Glass Technologies, hereafter referred to as Emsi Burning Glass (EBG) data, was identified as a good candidate to update the way Skills for Jobs indicators are constructed. EBG is particularly rich, timely, and highly granular: it contains data on skills required by employers in job vacancies posted online since 2012 in several countries. Recognising its relevance and quality, these data have been used extensively by academic researchers and policy makers to study labour market dynamics and the evolution of skill demand across occupations (Deming and Kahn, 2018; Hershbein and Kahn, 2018; Modestino, Shoag and Ballance, 2020; Deming and Noray, 2020). While this note argues the value of using EBG data over O*NET to construct the Skills for Jobs 2022 indicators, the new source of data and methodological approach imply that the 2022 indicators are not directly comparable to their previous vintage.

This note briefly describes the Emsi Burning Glass dataset and its value added over more traditional data sources, as well as the way the skill information has been pre-processed to be used in the Skills for Jobs 2022 Database. It then describes the methodology developed to replace the occupations-skill mapping from O*NET by a similar mapping based on skill information contained in the Emsi Burning Glass data. It concludes with a discussion of the results.

Data description

General description of Emsi Burning Glass data used for this study

The Emsi Burning Glass dataset used for this study includes information on more than 200 million job ads gathered online since 2012 across six English-speaking countries: Australia, Canada, New Zealand, Singapore, the United Kingdom, and the United States. For the United States, earlier data for 2007, 2010 and 2011 are also available, but are not used in this work.

The Emsi Burning Glass dataset used for the Skills for Jobs database 2022 contains standardised information retrieved from job postings using more than 45 000 online sources. It includes job characteristics such as detailed industry and occupation codes, location, posting date, name of the employer, and requirements in terms of education, professional experience, and skills. This considerable level of detail allows analysing changes in job requirements within -- rather than only across -- occupations, sectors, and locations. Furthermore, since Emsi Burning Glass data are updated in real or quasi-real time, they also allow for an earlier detection of emerging trends than previously possible.

A number of existing studies describe the representativeness of Emsi Burning Glass data. Carnevale, Jayasundera and Repnikov (2014) show that, for the United States in 2006-2013, the aggregate number of job postings in Emsi Burning Glass strongly correlates with the number of job openings reported in the Bureau of Labor Statistics Job Openings and Labor Turnover Survey (JOLTS). If Emsi Burning Glass data over-represent openings for high-skilled jobs, this feature is constant over time (Deming and Kahn, 2018). Carnevale, Jayasundera and Repnikov (2014) estimate that, in the U.S. files, state, city, occupation title, major occupation group, skills, and education are correctly reported for at least 80% of observations with non-missing information, while accuracy is lower for minor occupation groups, and industry codes. Recent OECD work (Cammeraat and Squicciarini, 2021) has also analysed the representativeness of Emsi Burning Glass data against official employment data at the occupational level, showing that for the period 2010-2018, for the majority of countries, Emsi Burning Glass data is of sufficiently good quality to conduct policy analyses. Representativeness concerns exist for Canada and New Zealand, but issues emerge mostly for the years prior to 2015 and representativeness has improved since. Araki et al. (2022) further benchmark the dataset's representativeness on hiring data sourced from country-specific labour force surveys, and find a high degree of overall and occupation-level representativeness for 15 OECD countries and Singapore.

Pre-processing of the skill information contained in the Burning Glass dataset

To identify skills required to perform the job, Emsi Burning Glass analyses the text of each job vacancy. This information is processed and standardised, e.g. by removing duplicates, or by treating differences in spelling for the same skill. The resulting list of skill keywords include skills in the sense which is commonly understood (e.g. "Analytical Skills"), but also knowledge (e.g. "Food Safety" or "Environmental Policy") and abilities (e.g. "Detail-Oriented").

The vast majority of job postings contains these skill keywords, i.e. information on skills required to perform the job. The proportion of job ads for which skill requirements are expressed is particularly high in the U.S., where more than 98% of job ads mention at least one skill keyword (except for the year 2018 where this proportion is slightly lower). The percentage of observations with missing skill keywords is the highest in Canada, ranging between 2% and 17% depending on the year considered. In other countries, the share of observations with missing skill information is stable at around 10% across the years.

In total, there are more than 17 000 different unique skill keywords across all years and countries. To conduct meaningful empirical analyses and facilitate the interpretation of the results, these keywords need to be grouped into a lower number of categories. This is precisely the purpose of an earlier work conducted at the OECD (Lassébie et al., 2021). The paper presents a methodology to classify skill keywords found in Emsi Burning Glass data into a pre-existing expert-driven taxonomy of broader skill categories, largely inspired by O*NET (see Table 1 in Annex A). The approach uses a semi-supervised Machine Learning algorithm (called BERT in the Machine Learning literature) that classifies keywords according to their definition. It allows for the classification of the thousands of unique skill keywords contained in the Emsi Burning Glass dataset into 61 detailed categories, themselves organised in 16 broad categories. Compared to a manual classification, the proposed approach organises large amounts of skill information in an analytically tractable form, and with considerable savings in time and human resources.

Methodology

Rationale

The first studies exploiting Emsi Burning Glass skill information (Deming and Kahn, 2018; Hershbein and Kahn, 2018) measured skill requirements as skill frequencies, i.e. the frequency with which certain skill keywords are mentioned in the job postings of a firm or an occupation. However, skills mentioned in job ads may not always accurately reflect the skill content of jobs. First, some skills might be implicit in the job title. If not explicitly mentioned in the text of the ad, this might lead to an under-estimation of the requirements expressed for those skills. Second, some job posts may list skills that are not essential to perform the job. This might lead to an over-estimation of the importance of those skills. This issue might be more important for job ads posted online than offline, as constraints on ad length are less important for the former than for the latter.

To mitigate these potential issues, several researchers proposed an alternative measure of skill requirements, the Relative Comparative Advantage (RCA) of skills within occupations. RCA measures the importance of a skill in an occupation based on whether the skill is more frequently found in the job posts for that occupation compared to other skills. It was adapted by Dawson et al. (2019) for online job ads data from Alabdulkareem et al. (2018) who initially applied the concept of Relative Comparative Advantage, well-known in the trade literature, to skills using O*NET importance values. Dawson, Williams and Rizoiu (2021) then used it to measure skill similarity between occupations and build recommender systems for identifying optimal transition pathways between occupations. Giabelli et al. (2021) computed RCA and normalised RCA figures using skill frequencies in online job ads (see formulas below).

RCA enables researchers to smooth out variations of skill frequencies between occupations that are due to employers' tendency to under- or over-state the importance of certain skills while writing job advertisements[1]. An indicator of RCA thus helps mitigating the issues of implicit and irrelevant skills discussed above, and is here chosen as the preferred indicator to build a mapping of skills in occupations.

Formulas and intuitions

More specifically, in the present work, the frequency of skill category *s* in occupation j, $sf_{s,j}$, is defined as:

$$sf_{s,j} = \frac{Number\ of\ job\ posts\ requiring\ skill\ s\ in\ occupation\ j}{Number\ of\ job\ posts\ in\ occupation\ j}$$

Where *s* is not the skill keyword but the skill category as in Lassébie et al., (2021), to maximise the tractability of the ensuing indicators[2], and where missing skill frequencies are set to 0. To smooth out short-term fluctuations and highlight longer-term trends, 2-year moving averages of skill frequencies are computed. To further reduce undesirable volatility observed in specific years, countries and

1 The underlying hypotheses are that 1) implicit skills are mentioned by some employers and are particular to an occupation (or not mentioned in other occupations), and that 2) irrelevant skills are not mentioned in a majority of job ads. See the next subsection for further details.

2 These are therefore frequencies of the skill categories in job postings in a given occupation. Two job postings containing, respectively, one and two skill keywords falling under the same skill category therefore have the same frequency as defined here above.

occupations, the figures are pooled across the six countries considered for this study. Skill frequencies for broad categories are created as an unweighted average of the skill frequencies of the detailed categories[3].

The Relative Comparative Advantage (RCA) of a skill *s* in an occupation *j* is then defined as:

$$rca(j,s) = \frac{\overbrace{sf_{s,j} / \sum_{s' \epsilon S} sf_{s',j}}^{\text{Importance of a skill relative to other skills in the occupation}}}{\underbrace{\sum_{j' \epsilon J} sf_{s,j'} / \sum_{j' \epsilon J, s' \epsilon S} sf_{s',j'}}_{\text{Importance of the skill in the occupation, relative to that of the same skill in all other occupations}}} \qquad (1)$$

where $sf_{s,j}$ is the skill frequency, $\sum_{s' \epsilon S} sf_{s',j}$ is the sum of skill frequencies across all skills for occupation *j*, $\sum_{j' \epsilon J} sf_{s,j'}$ is the sum of skill frequencies of sill *s* across all occupations, and $\sum_{j' \epsilon J, s' \epsilon S} sf_{s',j'}$ is the sum of skill frequencies over all skills and all occupations.

The numerator assesses the relative importance of a skill category in an occupation. The weight of some job ads that require irrelevant skills is decreased, since their share in the total number of job ads is low. The denominator corresponds to the relative importance of the skill category in all other occupations. The denominator highlights what skill categories are particular to an occupation and decreases the importance of categories that are universally important, or more important for other occupations. For example, the skill frequency of the occupation "Labourers in Mining, Construction, Manufacturing and Transport" that require "Digital Skills" will be weighted by the total number of job ads mentioning "Digital Skills", which includes the job ads of occupations that rely more heavily on Digital Skills (e.g. "Science and Engineering Professionals"). Therefore, the ratio can be interpreted as a measure of relative importance of the skill category in an occupation, compared to the importance of the same skill category in all other occupations.

As can be inferred from the formula above, very low skill frequencies result in extremely high RCAs. To avoid this undesirable property, the RCA of some low-frequency skill categories is set to 0. In addition, two skill categories, Visual Abilities, and Equipment Selection, are excluded from the analysis, because they show very low frequencies in all years, countries, and occupations. If included, these two categories would make the denominator in the formula below close to 0, artificially inflating the RCA for those skill categories. As a result, 56 detailed skill categories are used in the Skills for Jobs database 2022.

The resulting indicator ranges between 0 and + ∞. When the RCA is greater than 1, the skill category is more important for this occupation than for all other occupations, indicating a relative comparative advantage. However, it is important to note that, while RCA values of different skill categories can be compared within occupations, values are not comparable across occupations because the span of RCA values is not bounded and changes from occupation to occupation. As a solution, Giabelli et al. (2021) compute a normalised RCA, by dividing the RCA by the maximum RCA value obtained across skill categories for a given occupation. By doing so, the most important skill category for each occupation has a normalised RCA equal to 1:

$$rca_{norm}(j,s) = \frac{rca(j,s)}{max_k \, rca(j,s_k)} \qquad (2)$$

where $max_k \, rca(j, s_k)$ is the maximum RCA in a given occupation across the different skill categories. Normalised RCAs thus range between 0 and 1 and can be compared both within and across occupations.

3 This ensures, by construction of the RCA, that when all detailed categories within a single broad category are in shortage (resp. surplus) for a given occupation, then also the broad skill category is flagged as in shortage (resp. surplus).

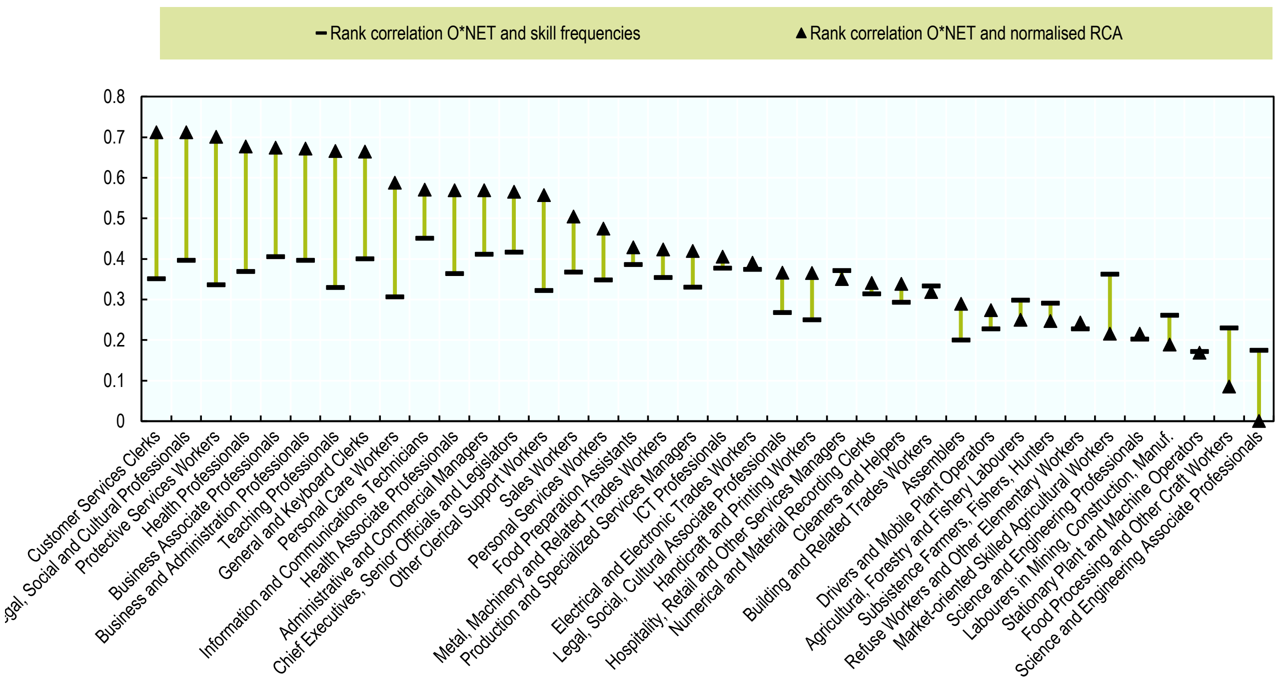
Discussion of results

*Comparison of skills mappings obtained from Emsi Burning Glass and from O*NET*

In the OECD Skills for Jobs database 2022, skill requirements by importance sourced from O*NET is substituted with similar information extracted from Emsi Burning Glass. Yet, and as highlighted above, the two datasets do not contain the exact same information. In particular, information in Emsi Burning Glass is more up-to-date and granular than O*NET, and based on several English-speaking countries rather than on the U.S. alone. Furthermore, O*NET reflects the skill content of occupations in the entire economy, while Emsi Burning Glass data capture a snapshot of the online vacancy market, which may not be representative of all existing jobs. Moreover, there is some imprecision in mapping the O*NET occupations according to the US SOC classification system (SOC) to the international one (ISCO) used for the EGB data. For these reasons, the two datasets are not expected to deliver the exact same occupations-skills mapping.

Figure 1 shows the correlation between the two mappings. More specifically, given that only the relative values from O*NET – i.e. the ranking of skills importance by occupation – matter for the construction of the Skills for Jobs database, the figure shows the rank correlation (Spearman correlation statistics) between O*NET importance values and the normalised RCA measures computed using Emsi Burning Glass, by occupation (ISCO-08, 2 digits). The correlation between Emsi Burning Glass skill frequencies (sf_sj) and O*NET importance values is also reported for comparison. The figure shows that the normalised RCA-based rankings are more strongly correlated with O*NET-based rankings in most occupations, while the skill frequency's rank correlates only weakly with O*NET. Yet, and as expected, there are a number of occupations for which the correlation is low. This may be because these are low-skilled occupations for which job vacancies are less likely to be posted online (e.g. "Assemblers"), or because these occupations rely more heavily on "Digital Skills", which are not included in O*NET.

Figure 1. Spearman rank correlation for the U.S. in 2019



Note: Skills in an occupation can be ranked by "importance", based on expert judgement (in O*NET), or on information on skills as reported in the job ads for the occupation (skills frequency vs RCA). This figure correlates the ranking of skills when O*NET is used with the ranking when skill frequencies are used (horizontal marker), or when the normalised RCA is used (triangle marker). Correlations are by 2-digit ISCO-08 occupations and range between 0 and 1. Only skill categories which originate from O*NET appear in the correlation.
Source: OECD calculations on U.S. O*NET and Emsi Burning Glass data.

To further check whether it is reasonable to associate normalised RCAs with skill importance, Table A B.1 (in Annex B) shows, for each occupation, the five skill categories with the highest normalised RCA values, with sensible and reassuring overall results. For instance, management and negotiation skills are assessed as important for "Legislators, senior officials and managers", and the same applies to specialised knowledge fields for "Professionals" and "Associate Professionals and Technicians". Clerical support workers mostly require administrative skills such as "Clerical", "Office Tools And Collaboration Software", and "Reading Comprehension". "Service and sales workers" mostly need skills to deal with customers: "Customer and personal service", "Sales And Marketing" but also "Public Safety And Security". For "Skilled agricultural, forestry and fishery workers", skills "Biology", "Building and Construction", "Installation and Maintenance" are especially important.

Final skill imbalances

Defining a mapping of skills in occupations is but one step in the process of measuring skill shortages at the country level. As reported in Formula (3) below, calculating an index of skill imbalance (shortage or surplus) for country c and a given skill category s requires three fundamental pieces of information: i) the importance of that skill category in each occupation j based on the normalised RCA (which relies here on cross-country pooled data), ii) a measure of the size of the occupational imbalance in the country (i.e. whether occupation j is in shortage or surplus), iii) and the relative size of the occupation in the country's total employment. Further details on this methodology, and in particular on (ii), can be found in (OECD, 2017).

The mapping (i) given by the normalised RCA ranges between 0 and 1, the occupational imbalance index (ii) ranges between -2.5 (surplus) and 2.5 (shortage), and employment shares (iii) range between 0 and 1. By construction, each country displays occupations in shortage as well as in surplus (OECD, 2017).

As a way of example, one can imagine an economy composed by two occupations (a and b) being respectively in large surplus and small shortage and accounting for respectively 70% and 30% of the country's employment. In this economy there are two skill categories (*x* and *y*), with x displaying a higher RCA in occupation *b* than in *a*, and the vice versa for *y*.

$$\mathit{Skill\ imbalance}_{c,s} = \textstyle\sum_j \Big(\underbrace{\mathit{skill\ mapping}_{s,j}}_{\text{(i) From EBG}} * \underbrace{\mathit{occ\ imbalance\ index}_{c,j}}_{\text{(ii) From Skills for Jobs}} * \underbrace{\frac{\mathbf{occ\ employment\ share}_{c,j}}{\sum_j \mathbf{occ\ employment\ share}_{c,j}}\Big)}_{\text{(iii) From Labour Force Surveys}} \qquad (3)$$

The mapping (i) given by the normalised RCA ranges between 0 and 1, the occupational imbalance index (ii) ranges between -2.5 (surplus) and 2.5 (shortage), and employment shares (iii) range between 0 and 1. By construction, each country displays an approximately equal number of occupations in shortage or surplus (OECD, 2017).

As a way of example, one can imagine an economy composed by two occupations (a and b) being respectively in large surplus and small shortage and accounting for respectively 70% and 30% of the country's employment. In this economy there are two skill categories (x and y), with x displaying a higher RCA in occupation B than in A, and the viceversa for y.

Occupation	Occupation's imbalance	Occupation's employment share	Skill category	Normalised RCA
a	-1.6	0.7	x	0.1
			y	0.8
b	0.4	0.3	x	0.8
			y	0.1

Then the country's imbalance for skill categories x and y is:

$$\begin{aligned}
Skill\ imbalance_{\boldsymbol{x}} &= RCA_{x,a} * occ\ imbalance\ index_a * empl\ share_a + \\
&+ RCA_{x,b} * occ\ imbalance\ index_b * empl\ share_b = \\
&= 0.1 * (-1.6) * 0.7 + 0.8 * 0.4 * 0.3 \\
&= -0.112 + 0.096 = \mathbf{-0.016} \\
Skill\ imbalance_{\boldsymbol{y}} &= RCA_{y,a} * occ\ imbalance\ index_a * empl\ share_a + \\
&+ RCA_{y,b} * occ\ imbalance\ index_b * empl\ share_b = \\
&= 0.8 * (-1.6) * 0.7 + 0.1 * 0.4 * 0.3 \\
&= -0.896 + 0.012 = -\mathbf{0.884} \qquad (4)
\end{aligned}$$

Skill category x is therefore in balance (or mild surplus) in this economy, despite the fact that occupation a is in large surplus and represents a large share of total employment. This is especially due to the fact that skill category x is not very important for occupation a, as flagged by the low normalised RCA (0.1), while it is very important in occupation b (0.8), which is conversely in shortage. The relative importance of skill category y across occupations, conversely, goes in parallel with the relative occupational imbalance, resulting in a large skill surplus at the aggregate level.

As the country-level skill-specific imbalance is the product of three different items referring to different concepts, the exact magnitude of the imbalance has no straight-forward interpretation. One can, however, compare the imbalance for two distinct skill categories, and mention that skill y is in larger surplus than skill x. Similarly, one can place the imbalance for skill x in the distribution of imbalances created by all the skill categories of the O*NET+ classification.

The example also shows that skill categories with a similar average importance across occupations (by construction here 1/2* 0.1+1/2*0.8= 0.45) can display significantly different aggregate imbalances depending on the relative importance of these skills across occupations in shortage or surplus.

In the example, both skill categories are in surplus. Even in a more complex economy with more skills and more occupations, and a balanced number of occupations in shortage or surplus by construction, some countries can display a lot (if not the totality) of skills in surplus (respectively, in shortage), as long as:

- Occupations in shortage (resp., surplus) are small (i.e., low proportion of total employment), or
- Occupations in shortage (resp., surplus) display small shortages (resp., surpluses), or
- For the occupations in shortage (resp., surplus) the skill is not very important (i.e., has a low normalised RCA).

Similarly, the same skill category can be in shortage and in surplus in two different moments of time due to changes in any of the three components described above (skill requirements in occupations, occupational imbalances, employment shares). For example, a skill can switch from being in shortage to being in surplus if:

- The skill becomes more important for occupations that are and were in surplus, or
- The skill has always been important for occupations that grow substantially in surplus (while being in small surplus or even in shortage at the beginning of the period), or
- The importance of the skill for the occupation and the size of the occupational imbalance stay approximately unchanged throughout the period, but the share of employment in surplus occupations increases, or
- Any combination of the three options above.

Lastly, the skill imbalances in the Skills for Jobs database are often expressed not in their absolute value (-0.016 and 0.884 in the example) but after rescaling. The rescaling divides the skill imbalances by the highest skill imbalance available in the dataset (across all countries, years and skill categories). This extra rescaling is used for presentational purposes only, and aims to provide an indication of the relative size of the imbalance in a given country, skill or moment in time.

Limitations

The methodological section above discussed several shortcomings usually associated with EBG data that are mitigated by the use of normalised RCA measures. Other issues, conversely, cannot be solved with the use of normalised RCA measures and should be kept in mind when interpreting the final results.

A number of existing studies have investigated the representativeness of EBG data along several dimensions and found satisfying results (Cammeraat and Squicciarini, 2021; Carnevale, Jayasundera and Repnikov, 2014; Deming, 2017). Yet, vacancies for a specific occupation may not be representative of all existing jobs in that occupation, and their skill requirements might be biased towards some types of skills. For example, employers may be seeking candidates with up-to-date knowledge and skills related to new technologies whereas most incumbent workers may not have these skills. If this is the case, the mapping constructed using EBG data will be more representative of new jobs created than of the whole labour market, and hence, in a sense, more future-oriented.. If this the case, the RCA would be biased in an ex-ante unknown direction. It is important to acknowledge, however, that there exist no empirical test for such selection on unobservable characteristics to date.

Furthermore, there is a concern that the increasing number of skills mentioned in job postings over time, and improving data scraping technologies with which the advertisement texts are stored into a dataset, can bias the mapping of skill requirements in occupations if using RCAs. For this to be the case, however, two conditions should simultaneously apply: (i) that what increases is not just the number of skill keywords in an occupation, but the number of skill categories (i.e. broader concepts), and (ii) that the newly-added skill categories are very frequent across postings and receive high normalised RCAs. Both conditions are possible but unlikely, and in particular (ii), since a whole new skill category emerging for an occupation is very unlikely to become ubiquitous in a short amount of time.

Lastly, online job postings data in the Skills for Jobs database 2022 are used to measure skill requirements in occupations, as opposed to actual skill imbalances directly. This methodological approach is consistent with the previous vintage of the Skills for Jobs database, but implies that differences in skill-level imbalances across occupations are averaged away in the aggregate. Alternative approaches can be explored in future work, whereby previously-unavailable granular data on skill-level demand and supply can be used instead.

References

Alabdulkareem, A. et al. (2018), "Unpacking the polarization of workplace skills", Science Advances, Vol. 4/7, https://doi.org/10.1126/sciadv.aao6030.

Araki, S. et al. (2022), "Monopsony and concentration in labour markets", in OECD Employment Outlook 2022, OECD Publishing.

Cammeraat, E. and M. Squicciarini (2021), Assessing the properties of Burning Glass Technologies' data to inform use in policy-relevant analysis, OECD Publishing.

Carnevale, A., T. Jayasundera and D. Repnikov (2014), Understanding online job ads data.

Dawson, N. et al. (2019), "Adaptively selecting occupations to detect skill shortages from online job ads", 2019 IEEE International Conference on Big Data (Big Data), https://doi.org/10.1109/bigdata47090.2019.9005967.

Deming, D. (2017), "The Growing Importance of Social Skills in the Labor Market*", The Quarterly Journal of Economics, Vol. 132/4, pp. 1593-1640, https://doi.org/10.1093/qje/qjx022.

Deming, D. and L. Kahn (2018), "Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals", Journal of Labor Economics, Vol. 36/S1, pp. S337-S369, https://doi.org/10.1086/694106.

Deming, D. and K. Noray (2020), "Earnings Dynamics, Changing Job Skills, and STEM Careers*", The Quarterly Journal of Economics, Vol. 135/4, pp. 1965-2005, https://doi.org/10.1093/qje/qjaa021.

Fu, S. (ed.) (2021), "Skill-driven recommendations for job transition pathways", PLOS ONE, Vol. 16/8, p. e0254722, https://doi.org/10.1371/journal.pone.0254722.

Giabelli, A. et al. (2021), "Skills2Job: A recommender system that encodes job offer embeddings on graph databases", Applied Soft Computing, Vol. 101, p. 107049, https://doi.org/10.1016/j.asoc.2020.107049.

Hershbein, B. and L. Kahn (2018), "Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings", American Economic Review, Vol. 108/7, pp. 1737-1772, https://doi.org/10.1257/aer.20161570.

Lassébie, J. et al. (2021), "Speaking the same language: A machine learning approach to classify skills in Burning Glass Technologies data", OECD Social, Employment and Migration Working Papers, No. 263, OECD Publishing, Paris, https://dx.doi.org/10.1787/adb03746-en.

Modestino, A., D. Shoag and J. Ballance (2020), "Upskilling: Do Employers Demand Greater Skill When Workers Are Plentiful?", The Review of Economics and Statistics, Vol. 102/4, pp. 793-805, https://doi.org/10.1162/rest_a_00835.

Annex A. The O*NET+ taxonomy

Table A.1. Skill categories used in the Skills for Jobs database

Broad category	Category label	Source
Attitudes	Adaptability/resilience	ESCO
	Motivation/commitment	ESCO
	Self-management/rigor	ESCO
	Values	ESCO
Arts and Humanities	Fine Arts	O*NET
	History and Archaeology	O*NET
	Philosophy and Theology	O*NET
Business Processes	Clerical	O*NET
	Sales and Marketing	O*NET
	Customer and Personal Service	O*NET merged
Production and Technology	Telecommunications	O*NET
	Building and Construction	O*NET
	Engineering, Mechanics and Technology	O*NET
	Design	O*NET
	Food Production	O*NET
	Production and Processing	O*NET
	Transportation	O*NET
	Quality Control Analysis	O*NET
	Installation and Maintenance	O*NET merged
Medicine	Medicine and Dentistry	O*NET
	Psychology, Therapy, Counselling	O*NET merged
Law and Public Safety	Law and Government	O*NET
	Public Safety and Security	O*NET
Science	Biology	O*NET
	Chemistry	O*NET
	Geography	O*NET
	Physics	O*NET
	Sociology and Anthropology	O*NET
Physical Skills	Psychomotor Abilities	O*NET
	Auditory and Speech Abilities	O*NET
	Physical Abilities	O*NET merged
Cognitive Skills	Originality	O*NET
	Quantitative Abilities	O*NET
	Reasoning and Problem-solving	O*NET merged
	Learning	O*NET merged
Communication	Active Listening	O*NET
	Reading Comprehension	O*NET
	Speaking	O*NET
	Writing	O*NET
	Communications and Media	O*NET
Digital	Office Tools and Collaboration Software	New category, based on ESCO
	Digital Content Creation	ESCO
	Digital Data Processing	ESCO
	ICT Safety, Networks and Servers	New category, based on ESCO
	Computer Programming	O*NET
	Web Development and Cloud Technologies	New category
Resource Management	Time Management	O*NET
	Management of Material Resources	O*NET
	Management of Financial Resources	O*NET merged
	Management of Personnel Resources	O*NET merged
	Administration and Management	O*NET
Social Skills	Coordination	O*NET
	Persuasion and Negotiation	O*NET
	Social Perceptiveness	O*NET
	Judgment and Decision Making	O*NET merged
Training and Education	Training and Education	O*NET

Note: The mention "O*NET merged" in the last column indicates when the category is the result of merging two or more O*NET original categories. The list only contains the O*NET+ categories from Lassébie et al. (2021) that are used in the Skills for Jobs database.
Source: OECD elaborations on O*NET and ESCO hierarchies.

Annex B. The most important skill categories by occupation

Table B.1. Five skill categories with the highest normalised RCA value, by occupation

Occupation	Category 1	RCA	Category 2	RCA	Category 3	RCA	Category 4	RCA	Category 5	RCA
11 Chief executives, senior officials and legislators	Persuasion And Negotiation	1	Reading Comprehension	0.81	Administration And Management	0.77	Management Of Financial Resources	0.58	Law And Government	0.58
12 Administrative and commercial managers	Persuasion And Negotiation	1	Sales And Marketing	0.64	Administration And Management	0.58	Management Of Financial Resources	0.56	Judgment And Decision Making	0.56
13 Production and specialized services managers	Social Perceptiveness	1	Sociology And Anthropology	0.95	Psychology, Therapy, Counselling	0.40	Judgment And Decision Making	0.35	Medicine And Dentistry	0.33
14 Hospitality, retail and other services managers	Philosophy And Theology	1	Food Production	0.95	Administration And Management	0.85	Management Of Material Resources	0.63	Management Of Financial Resources	0.49
21 Science and engineering professionals	Geography	1	Digital Content Creation	0.82	Engineering, Mechanics And Technology	0.68	Chemistry	0.67	Design	0.51
22 Health professionals	Psychology, Therapy, Counselling	1	Medicine And Dentistry	0.61	Judgment And Decision Making	0.34	Learning	0.24	Training And Education	0.19
23 Teaching professionals	Learning	1	Training And Education	0.65	Values	0.36	History And Archaeology	0.25	Fine Arts	0.23
24 Business and administration professionals	Persuasion And Negotiation	1	Sales And Marketing	0.76	Management Of Financial Resources	0.65	Auditory And Speech Abilities	0.58	Originality	0.49
25 Information and communications technology professionals	Computer Programming	1	Web Development And Cloud Technologies	0.74	Ict Safety, Networks And Servers	0.58	Digital Data Processing	0.50	Adaptability/Resilience	0.34
26 Legal, social and cultural professionals	Psychology, Therapy, Counselling	1	Sociology And Anthropology	0.97	Reading Comprehension	0.92	Law And Government	0.60	Geography	0.52
31 Science and engineering associate professionals	Digital Content Creation	1	Chemistry	0.77	Quality Control Analysis	0.61	Production And Processing	0.60	Engineering, Mechanics And Technology	0.57
32 Health associate professionals	Chemistry	1	Medicine And Dentistry	0.96	Psychology, Therapy, Counselling	0.60	Reading Comprehension	0.52	Biology	0.50
33 Business and administration associate professionals	Philosophy And Theology	1	Persuasion And Negotiation	0.53	Management Of Financial Resources	0.36	Sales And Marketing	0.30	Auditory And Speech Abilities	0.26
34 Legal, social, cultural and related associate professionals	Food Production	1	Law And Government	0.48	Originality	0.47	Adaptability/Resilience	0.44	History And Archaeology	0.37
35 Information and communications technicians	Ict Safety, Networks And Servers	1	Computer Programming	0.71	Telecommunications	0.68	Web Development And Cloud Technologies	0.38	Office Tools And Collaboration Software	0.32
41 General and keyboard clerks	Clerical	1	Digital Data Processing	0.49	Office Tools And Collaboration Software	0.39	History And Archaeology	0.33	Writing	0.32
42 Customer services clerks	Active Listening	1	Customer And Personal Service	0.86	Clerical	0.60	Reading Comprehension	0.56	Social Perceptiveness	0.50
43 Numerical and material recording clerks	Transportation	1	Management Of Material Resources	0.94	Geography	0.93	Quantitative Abilities	0.81	Reading Comprehension	0.79
44 Other clerical support workers	Clerical	1	Judgment And Decision Making	0.69	Quantitative Abilities	0.69	Management Of Personnel Resources	0.51	Telecommunications	0.42

51 Personal service workers	Motivation/Commitment	1	Food Production	0.97	Customer And Personal Service	0.91	Installation And Maintenance	0.61	Physical Abilities	0.61
52 Sales workers	Management Of Material Resources	1	Sales And Marketing	0.78	Active Listening	0.61	Customer And Personal Service	0.50	Design	0.44
53 Personal care workers	Training And Education	1	Sociology And Anthropology	0.88	Medicine And Dentistry	0.76	Psychology, Therapy, Counselling	0.73	Learning	0.51
54 Protective services workers	Public Safety And Security	1	Ict Safety, Networks And Servers	0.44	Auditory And Speech Abilities	0.43	Learning	0.39	Law And Government	0.36
61 Market-oriented skilled agricultural workers	Psychomotor Abilities	1	Biology	0.41	Values	0.33	Public Safety And Security	0.30	Physical Abilities	0.21
62 Market-oriented skilled forestry, fishery and hunting workers	Biology	1	Food Production	0.68	Physical Abilities	0.56	Production And Processing	0.33	Installation And Maintenance	0.24
63 Subsistence farmers, fishers, hunters and gatherers	Building And Construction	1	Chemistry	0.28	Installation And Maintenance	0.27	Engineering, Mechanics And Technology	0.22	History And Archaeology	0.20
71 Building and related trades workers, excluding electricians	Production And Processing	1	Building And Construction	0.97	Engineering, Mechanics And Technology	0.76	Installation And Maintenance	0.63	Design	0.60
72 Metal, machinery and related trades workers	Digital Content Creation	1	Psychomotor Abilities	0.87	Physics	0.69	Clerical	0.64	Physical Abilities	0.62
74 Electrical and electronic trades workers	Telecommunications	1	Engineering, Mechanics And Technology	0.97	Installation And Maintenance	0.69	Quality Control Analysis	0.49	Physics	0.48
81 Stationary plant and machine operators	Quality Control Analysis	1	Production And Processing	0.82	Physics	0.71	Engineering, Mechanics And Technology	0.61	Chemistry	0.57
82 Assemblers	Design	1	Building And Construction	0.83	Psychomotor Abilities	0.58	Quantitative Abilities	0.56	Chemistry	0.50
83 Drivers and mobile plant operators	Physics	1	Psychomotor Abilities	0.69	Engineering, Mechanics And Technology	0.42	Building And Construction	0.38	Design	0.31
91 Cleaners and helpers	Transportation	1	Values	0.35	Physical Abilities	0.23	Installation And Maintenance	0.22	Public Safety And Security	0.21
92 Agricultural, forestry and fishery labourers	Installation And Maintenance	1	Customer And Personal Service	0.49	Active Listening	0.46	Motivation/Commitment	0.43	Philosophy And Theology	0.42
93 Labourers in mining, construction, manufacturing and transport	Values	1	Biology	0.54	Psychomotor Abilities	0.53	Physical Abilities	0.20	Building And Construction	0.20
94 Food preparation assistants	Transportation	1	Quantitative Abilities	0.90	Physical Abilities	0.68	Management Of Material Resources	0.57	Psychomotor Abilities	0.55
95 Street and related sales and service workers	Food Production	1	Management Of Material Resources	0.22	Motivation/Commitment	0.20	Active Listening	0.19	Coordination	0.15
96 Refuse workers and other elementary workers	Building And Construction	1	Engineering, Mechanics And Technology	0.82	Installation And Maintenance	0.73	Adaptability/Resilience	0.70	Quality Control Analysis	0.55

Note: Occupations refer to the 2-digit ISCO2008 nomenclature.
Source: OECD calculations on U.S. O*NET and Emsi Burning Glass data.

Contacts:

Glenda Quintini glenda.quintini@oecd.org
Luca Marcolin luca.marcolin@oecd.org
Julie Lassebie julie.lassebie@oecd.org

Design and editorial support:
Monica Meza-Essid

www.oecdskillsforjobsdatabase.org

BETTER POLICIES FOR BETTER LIVES

NEW SKILLS AT WORK

JPMorgan Chase & Co.