



EN RLMM #9/2022

Frankfurt, 14 December 2022

[www.regionallabourmarketmonitoring.net](http://www.regionallabourmarketmonitoring.net)

## THE END OF AN EVENTFUL AND PAINFUL YEAR

Dear network members,

2022 marks an important turning point in our cooperation within the European Network on Regional Labor Market Monitoring. For the first time in our almost 20-year history, our friendly and open international cooperation has been pushed to its limits. It is still very painful for us that, due to the ongoing war of aggression against Ukraine, we were forced to pronounce a temporary suspension of our members in Russian organisations or organisations in Russia. One of our guiding ideas, that a peaceful and cooperative world can be shaped through professional and personal exchange, has cracked, disappointment and anger are also part of it.

However, this year we were also able to experience a face-to-face meeting with excellent hosts, namely EURISPES in Sardinia. Our exchange, our personal encounters have done us good, even if not all network members have been able to be there again. We look forward to the next meeting in Potsdam in 2023. Then there will certainly be more of us again on site in exchange with each other.

We have also broadened our horizons again this year. We had an interesting exchange with a delegation from the ASEAN countries. Thank you to GIZ, which once again built the bridge to our network.

And last but not least, we would like to thank the Scientific Committee and its chair Marco Ricceri in particular for making many courageous decisions, making the network visible at European level and working on its conceptual further development.

We thank everyone for their tireless and friendly commitment. May we be able to continue this in the new year 2023, in the hope that 2023 will be more peaceful, that we can lift our temporary suspension and that we can continue to cultivate cross-border professional exchanges with the human and personal side. Our thoughts and solidarity also go to all colleagues in Ukraine.

Please find attached some last topics for this year, that might be interesting for you.

Yours Christa and Jenny

Coordination and Management of the European Network on Regional Labor Market Monitoring

**NEXT ANNUAL MEETING IN POTSDAM 28<sup>th</sup> AND 29<sup>th</sup> SEPTEMBER 2023**



Please take notice that the next Annual Meeting in 2023 will be hosted by the Brandenburg Economic Development Agency (WFBB) in Potsdam (Greater Berlin Area in Germany). We will focus on the topic “**Pathways of Greening Labour Markets. Opportunities and Challenges for Regional and Local Labour Market Observation in Europe and Beyond**”. We feel very honored that the OECD-LEED-Programme will co-host this meeting. The call for papers to prepare a new anthology is attached. Please have a look. If you think it might be of interest for others please spread it. All further information on the anthology as well as the Annual Meeting in Potsdam we will send to you via this newsletter. You also might be interested in visiting our networks homepage [http://regionallabourmarketmonitoring.net/?page\\_id=3312](http://regionallabourmarketmonitoring.net/?page_id=3312).

#### **REQUEST FROM PROSPEKTIKER AND LANBIDE**

Dear member of the network,

from Lanbide and Prospektiker we are working on a project called "Design of a model for prospecting the employment and training needs of companies" and at this moment we would like to know good practices carried out in the different employment services, both national and regional, in this matter. It would be useful anything related to the subject, from strategies for the detection of needs, to concrete projects that help to obtain information. In this first moment it would be enough for us to have the main ideas of the project, and if later we wanted to delve deeper into the subject we would contact you again.

For us it would be very helpful to know good practices and in advance we want to thank you for the work.

Any questions or needs for information, do not hesitate to contact us.  
Email: [b.pulido@prospektiker.es](mailto:b.pulido@prospektiker.es)

#### **BIG DATA HUB – LOOKING BACK - ONLINE SEMINAR ON 24<sup>th</sup> NOVEMBER**

The third of a series of seminars of the **Big Data Knowledge Hub** took place on **November 24, 2022**. Members of the EN RLMM met online to discuss a project by the CRISP- University of Milano-Bicocca about **Skills mismatch across EU countries using job ads and survey data at the micro-level**. A novel approach using data from PIAAC (Program for the International Assessment of Adult Competences by OECD) and data from online job ads (WIH-OJA by Eurostat and



Cedefop) to measure the skill mismatch in 17 European countries was presented by **Francesco Trentini** (University of Milano-Bicocca, Italy). The methodology reconstructs the labour market matching, combining different types of data sets and using AI to make them comparable. The presentation was followed by an open discussion of European experts. Please find attached the PowerPoint Slides and the report that provides all the details about the methodology and the seminar in case you were not able to join.

The Speaker of the ENRLMM Big Data Working Group is available for any further information. Eugenia Atin, Tl. +34 688 809 708, E-mail: [e.atin@prospektiker.es](mailto:e.atin@prospektiker.es)

#### **BIG DATA HUB – ONLINE SEMINAR ON 23rd FEBRUARY – 13:00-14:30**

The Big Data Working Group from the EN RLMM European Network is pleased to invite you to participate in the fourth online seminar of the series "**Seminars of the Big Data Knowledge Hub**". This fourth session will be dedicated to two different projects that our colleagues from Switzerland are conducting. The seminar is scheduled for **Thursday February 23, 2022, 1.00 - 2.30 pm** (CET) and has the following title:

**"Use of big data tools to support labour market (re)integration in the Swiss context"**. The Program for this fourth seminar includes a presentation by **Professor Dominik Hangartner** (Immigration Policy Lab, ETH Zürich, Switzerland) who will present the project **Improving refugee integration through data-driven algorithmic assignment** and a presentation by **Martin Gasser** (SECO, Switzerland) who will present the project **Implementation of Risk Assessment Tools in the Public Employment Service- State of Work in Switzerland**.

An expert will comment on both experiences: **Michel van Smoorenburg** (UWV, The Netherlands). There will also be an open debate with registered participants, with the possibility of being able to express perceptions, and exchange questions and answers.

The online seminar is held in English. The Microsoft TEAMS Platform will be used. To participate you need to register. To register, simply open the link below and write the basic information: <https://teams.microsoft.com/join/6xVauF8dnkST2Hepy6UAEw,-8xAfjVAIUOTJKHO3vYNyQ,yutJ3-hvbk6fRB5vsSuU9Q,iBzANEh4FEOke2fdEFVQ2g,rAJE9cyVPke3dpHAOte0mA,XO-KNJVGgU6MsXGZFmClyA?mode=read&tenantId=b85a15eb-1d5f-449e-93d8-77a9cba50013>



The Speaker of the ENRLMM Big Data Working Group is available for any further information both in reference to the event and to the registration and access methods. Eugenia Atin, Tl. +34 688 809 708, E-mail: [e.atin@prospektiker.es](mailto:e.atin@prospektiker.es)

#### **ETF SKILLS LAB NETWORK – CALL FOR JOINT RESEARCH PROPOSELS**

Our colleagues from the ETF Skills Lab Network are looking out for joint projects/proposels. Please find more information here <file:///C:/Users/Asus/Downloads/ETF%20Skills%20Lab%20Network%20of%20Experts%20-%20Call%20for%20joint%20research%20proposals.pdf>

#### **EUROPEAN COMMISSION – CALL**

This call from the EC on “Strengthen economic fairness and resilience of active labour market policies and address high unemployment” might be of interest for some of you. Please let us know and we will connect all interested persons/organisations. Please find more information here <https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/horizon-cl2-2024-transformations-01-02;callCode=null;freeTextSearchKeyword=labour;matchWholeText=true;typeCodes=0,1,2,8;statusCodes=31094501,31094502;programmePeriod=null;programCcm2Id=null;programDivisionCode=null;focusAreaCode=31087048;destinationGroup=null;missionGroup=null;geographicalZonesCode=null;programmeDivisionProspect=null;startDateLte=null;startDateGte=null;crossCuttingPriorityCode=null;cpvCode=null;performanceOfDelivery=null;sortQuery=sortStatus;orderBy=asc;onlyTenders=false;topicListKey=topicSearchTablePageState>

Have a wonderful time until the end of the year

Marco, Christa, Ida, Jenny & the EN RLMM Team

#### References

##### **Prof. Marco Ricceri**

Head of the Scientific Committee of the EN RLMM  
EURISPES – The Secretary General  
Via Cagliari 14, 00198 Roma I Italy  
Phone: +39 338 9766007  
E-mail: [eurispes.intl-dept@libero.it](mailto:eurispes.intl-dept@libero.it)  
Homepage: [www.eurispes.eu](http://www.eurispes.eu)



**Dr. Christa Larsen**

Co-ordinator of the EN RLMM, Managing Director of IWAK  
Institute for Economics, Labour and Culture (IWAK) Centre of Goethe University Frankfurt  
Campus Westend IKB-Gebäude | 1. Floor | Eschersheimer Landstraße 121  
60322 Frankfurt am Main | Germany  
Phone: +49-(0)69-798 25474  
E-mail: [c.larsen@em.uni-frankfurt.de](mailto:c.larsen@em.uni-frankfurt.de)  
Web: [www.iwak-frankfurt.de](http://www.iwak-frankfurt.de)

**Dr. Jenny Kipper**

Manager of the EN RLMM in 2023  
E-mail: [jenny@jennykipper.de](mailto:jenny@jennykipper.de)



**Topic & Call Anthology 2023**

**Pathways of Greening Labour Markets. Opportunities and Challenges for Regional and Local Labour Market Observation in Europe and Beyond**

The reduction of greenhouse gas emissions is one of the main long-term policy goals of the European Union and individual European states. This is impacting on a broad range of political activities and policies as it is defined in particular by the environmental taxonomy regulation of the EU.

This green transformation will affect a lot of areas of the labour market, requiring new skills and qualifications and impacting the nature and geographic location of employer demand. Considering the importance of this new ‘industrial revolution’, the topic for the next anthology of papers, published annually by the EN RLMM, is „Pathways of Greening Labour Markets. Opportunities and Challenges for Regional and Local Labour Market Observation in Europe and beyond “.

Therefore, network members and associated colleagues, and/or organisations, are asked to submit papers on this year’s topic. A number of papers will be presented at the Annual Conference, held for the first time in collaboration with the OECD LEED Programme on the 28<sup>th</sup> – 29<sup>th</sup> September, 2023 in Potsdam, Germany. This will be the 18<sup>th</sup> in a series of successful conferences on Regional and Local Labour Market Monitoring, hosted by the Brandenburg Economic Development Agency (WFBB) and the Network.

Europe’s green transition will impact on nearly all occupations, skills and competencies. While some sectors will face decreasing demand for their products and services, others will experience quite the opposite. Inevitably decarbonisation will lead to job losses in some sectors and companies, while it will simultaneously produce new jobs and opportunities in others. These new risks and emerging opportunities, costs and benefits, will not necessarily materialise in the same regions. Enterprises as well as public and non-profit-sectors, will have to adapt to new demands and develop new products and services. This will require a labour force with new skills and competencies.

Regional and local Labour Market Observatories will need to understand and interpret these transitions and new complexities. They will need to develop new concepts and tools for measuring these changes and ensure their impacts and potentials are understood by policymakers.

Although this greening will influence the economy as a whole, it may be necessary to define specific targets within particular labour markets. Interested researchers and practitioners are invited to discuss the advantages and disadvantages of different approaches to labour market observation, and their applications, within their papers. Topics and ideas such as the following could be covered:

- What are the major challenges and opportunities impacting the green transition of regional and local labour markets?
- How are our labour markets transforming – in practice/real world scenarios? What are new emerging trends? In how far does the European Green Deal impacting the labour market in your region?
- How do you measure the greening of regional and local labour markets? What are the green skills sets and which of them are identified as priority in decarbonisation of the economy? And how to consider unintended effects e.g. within or across occupations, sectors or regions? Do you apply already existing classifications for sectors (branches), occupations and skills or do you explore new criteria or in how far are you related to the EU taxonomy? What are the specific schemes which are to be monitored (e.g. guidance, training, apprenticeship, reskilling



programmes, jobcenters, sectoral or occupational policies)? What kind of data do you use and what are your results?

- Do you have information about the readiness of the labour force for green skills?
- Could new technologies be enablers for the transition? Is AI relevant in this framework and are there other factors taking over the function of enablers? How do you tackle the problem of complexity and other transition factors such as regional or local factors e.g. infrastructure including digital/artificial intelligence, VET-structures and relevant policies? How do labour market actors including policy makers see their roles in assisting green transition and are there examples on how to promote green innovations in regional and local labour markets? How to understand the interplay between green and digital agendas having effects on reshaping local and regional demand and supply?
- How are policy makers informed by observatories, how to improve the usefulness of local and regional labour market information in times of systematic transitions? How do you measure the socio-economic impact of the green transition within local labour markets, which is crucial for ensuring a just transition?

Impacts from decarbonisation will not be felt consistently across regions. Those whose economies include activities related to mining, fossil fuel production, refineries etc., could lose considerable numbers of jobs, including many well-paid technical ones. Such regions will have to manage these transformations or structural changes, and Labour Market Observatories will be central to understanding the labour market-related consequences. It may, in practice, be easier to measure and understand these major changes than to nuance the impact across the economy as a whole.

As regional structural changes have always been part of general economic development, and many European regions have experienced such processes already in the 1970s and 1980s. Papers would be welcomed which seek to utilise existing understanding and recognised approaches which could be transferable to monitoring the regional structural changes caused by decarbonisation. Common factors to look at include de-industrialisation, unemployment, sectoral shifts and decreasing tax revenues. Regional and local Labour Market Observatories should inform the management of such processes and therefore engage in effective and responsive labour market monitoring.

Papers focussing on a single region or locality addressing structural change and what we can learn from that - which can be transferred to other localities – particularly those likely to be most impacted, are welcome. Practical examples of how observatories are monitoring an already transforming economy and/or labour market are also welcome.

For consistency, papers should seek to include answers to the following questions:

- Which region/locality is affected by the structural change? What is causing the transformation and what are the impacts on the labour market? What are the transferable lessons to help us understand the transition into a decarbonised economy?
- Is there already any institutionalised labour market monitoring in the region? Which quantitative and qualitative indicators are applied for monitoring effects on regional or local labour markets? What insights do they deliver?

As a special notion of the green transition could be also considered that it is structured by tools of an environmental taxonomy which are rigid, general and focus on specific classifications, with indicators, etc. At the EU level such taxonomy is already entered into force with the tool of the Regulation (EU exclusive competence) and Directives (shared competences between EU and states). This taxonomy shows industries, for instance, what kind of products they can produce or not, what are the green products, etc. Some questions which could be covered in papers:

- Are the regional and local Labour Market Observatories ready to open their visions and could they collect data according to the taxonomy as well as to consider the impact of such taxonomy in their regions or localities. This could be of relevance as it is still an open question



if and in how far companies and regions could react well to the requirements embedded in the taxonomy in terms of modifying products, services and skills sets. What will happen to companies which are not in the condition to react on the new requirements of the taxonomy and what are the consequences from this for the regional economy and labour market?

- In how far do these framework conditions have also impact on the services of public employment services (PES)? Are there sufficient knowledge and competences available to organise programmes and trainings in regions and localities?

If you would like to submit a paper, please send the preliminary title of the contribution and a short abstract to Jenny Kipper ([jenny@jennykipper.de](mailto:jenny@jennykipper.de)) and Christa Larsen ([c.larsen@em.uni-frankfurt.de](mailto:c.larsen@em.uni-frankfurt.de)) by 31<sup>th</sup> January 2023. The full papers should be submitted to Jenny latest by 15<sup>th</sup> March 2023. If you have any further questions, please contact Jenny. We look forward to your contributions.





**CRISP**

Interuniversity  
Research Centre  
for Public Services



## NOTE

### **Seminars of the Big Data Knowledge Hub** **Skills mismatch across EU countries** **using job ads and survey data at the micro-level**

*Notes of the online seminar promoted by the Big Data Knowledge Hub of the European Network on Regional Labour Market Monitoring (ENRLMM). November 24, 2022*

The third of the Seminars of the Big Data Knowledge Hub took place on November 24, 2022. The aim of these series of seminars is to offer an opportunity to deepen the Network's knowledge on how to use Big Data for labour market research and consulting by presenting practical cases and demonstrations.

The project "Skills mismatch across EU countries using job ads and survey data at the micro-level" was presented by **Francesco Trentini** (University of Milano-Bicocca, Italy) and accompanied by **Lorenzo Malandri** (University of Milano-Bicocca, Italy). The open discussion counted with the participation of: **Claudia Plaimauer** (3s Unternehmensberatung GmbH), **Joanna Napierala** (Cedefop), **Michel van Smoorenburg** (UWV), **Aleksandra Webb** (University of the West of Scotland, UK) and **Christa Larsen** (IWAK Goethe University Frankfurt am Main, Germany).

#### **Introduction**

**Eugenia Atin** (Speaker of the Big Data Working Group of the ENRLMM) after the initial greetings and thanks to the participants, contextualises this session in the work being done by the Big Data Working Group of the ENRLMM (European Network on Regional Labour Market Monitoring).

**Christa Larsen** (Founder and Coordinator of the ENRLMM), then presents the European Network on Regional Labour Market Monitoring (ENRLMM) <http://regionallabourmarketmonitoring.net/>, that exists since 2007, and focuses on regional and local monitoring, especially on the demand side (the needs of companies). In 2016, Mario Mezzanzanica from the University of Milano Bicocca introduced a new way to explore data extracting information from the internet, with the focus on online job advertisements. Over the years the approach is growing and maturing,



The Big Data Working Group was also born then, in 2016, to help the network members in the application of big data techniques in their monitoring projects. And then, 2 years ago, the Group designed the Big Data Knowledge Hub <https://bigdatahub.uvt.ro/> which is a collaborative platform for mutual exchange and learning. The Knowledge Hub is the place where all the members of the EN RLMM can look for guidance when aiming to use big data in their labour market monitoring projects. It is an easy accessible source of information on the techniques used by other reference labour market observatories for a particular topic or challenge. The "Seminars of the Big Data Knowledge Hub" are an initiative within this Hub, to make it more interactive.

This third session was dedicated to a project that our partners from Italy are working on. Francesco Trentini from the CRISP research centre of the University of Milano Bicocca presented and described the process used to identify skill mismatches using job ads and survey data at the micro-level.

### **Presentation by Francesco Trentini**

Francesco Trentini first gives an overview of what he is going to speak about:

- Skill mismatch. Definition and used measures
- Novel approach using
  - PIAAC – Program for the International Assessment of Adult Competences (OECD)
  - Online job ads from WIH-OJA (Cedefop and Eurostat)
  - Bridging them with ESCO Skills pillar using AI
- Application of the mapping to measure skill mismatch in 17 European countries

The research has been developed in the framework of the EU funded H2020 [Pillars – Pathways towards inclusive labour markets](#).

We start with the definition of Skill mismatch which broadly is the deviation of workers' competencies from those required by employers to perform a job. So, we have both sides, the supply (the competences of the workers) and what is demanded by employers.

How is it possible to identify skill mismatch? It is complex in terms of the availability of good data.

Data on the supply (the skills of workers) usually comes from the highest level of education attained. On the demand side, there are surveys that are directed to employers, however, usually they lack the timely, frequent and detailed requirements.



In general, we can identify different approaches to measurement:

- Normative approaches: it is possible to ask experts to state the needed level of education or qualification to perform a job
- Objective approaches (or statistical): use the distribution of realised matches and measure the distance from, e.g., the median
- Subjective approaches: workers are asked to self-assess the coherence between their education/qualification and the task that perform on the job

In general, all the variety of approaches have important limitations: low granularity and high heterogeneity within education classes used to identify competencies on the supply side.

So, here we present a new approach where we have identified two sources of data:

1. On the supply side, the PIAAC questionnaire which consists of 2 parts:
  - a. Survey of Adult Skills: direct assessment of skills where participants are asked to complete certain tasks (e.g. use the browser, search for job online) and their performance is measured.
  - b. The [Background questionnaire](#) of individual characteristics: skill use in different domains: at home or at work, this is the part where we have focused on
2. On the demand side, the information on job ads provided by the web intelligence hub (WIH-OJA), the online job ad system, by Eurostat and Cedefop.

There is clearly one structural point to be solved: the two data sources are very valuable, but they express skills using different taxonomies. So, how can we bridge the two, avoiding aggregation and subjective alignment PIAAC question with ESCO skills?

First, we need to understand how the data looks like in both data sets.

### **Supply:**

PIAAC is the Program for the International Assessment of Adult Competences (OECD)

In the background questionnaire, there are different modules that cover different domains in the use of skills. It is a self-declared intensity (frequency or extent) of skill use in different domains. So, a person is asked about how he or she depicts himself. It applies at work, and it covers Literacy, Numeracy and ICT at work and everyday life. The skills are implicitly referenced to by items in the questionnaire.



There are 84 questions in these sections that are suitable for analysis. Moreover, PIAAC is used in the literature to measure skill content in a global perspective (Lewandowski et al. 2022) and used by the OECD to calculate the risk of automation (Nedelkoska and Quintini, 2018).

Figure 1 Example of a PIAAC question. Question F\_Q02b

Layout	Item group table
F_Q02b (JRA) (B)	<b>How often ^DoesDid your ^JobLastjob usually involve ... instructing, training or teaching people, individually or in groups?</b>  1 Never 2 Less than once a month 3 Less than once a week but at least once a month 4 At least once a week but not every day 5 Every day DK RF

Figure 1 presents the typical structure of a PIAAC question. The question asks how often do you (in your current or previous work) are involved in instructing, training or teaching people individually or in groups. We can identify that the question broadly refers to the semantic field of “teaching”, but still there is no standardised skill expressed.

### **Demand:**

Considering the demand side, we use Online Job Ads (OJAs) from the WIH-OJA dataset of Eurostat and Cedefop. It is a collection of online job ads from the whole European statistical system: 27 European countries + UK, evolving to soon cover EFTA countries, as well. The time series is since 2018Q4.

We have data on occupations and related skills as they emerge from online job postings. Skills and occupations are codified using the ESCO Skill Pillar and ESCO Occupation Pillar respectively.

### **So, how can we link PIAAC and OJA?**

Our first contribution is using ESCO and a mix of methods using Artificial Intelligence.

- The linkage is done using AI in a framework that combines various methods: embeddings, selection of the best embedding, taxonomy alignment and experts' validation
- PIAAC questions are processed to tag the most similar ESCO Skills.
- The embedding is trained on OJA UK data and the matching is done on the English language.



What we obtain is a validated dataset that covers 21 PIAAC questions and the mapped ESCO skills, enriched with alternative Labels. For instance, coming back to our example of question F\_Q02b, many ESCO terms are associated. The results of the matching are shown in Figure 2.

Figure 2 ESCO Skills associated with the PIAAC question F\_Q02b

F\_Q02b: instructing training or teaching people individually or in groups?

- coach young people
- coach youngsters
- educate others
- educate young people
- facilitate young peoples education
- facilitate young peoples mentoring
- instruct colleagues
- instruct others
- instruct young people
- instructing others
- teach others
- teach young people
- train others
- train young people
- train youngsters
- tutoring

Many more questions / skills were matched but the expert validation reduced the list to these 21. Only the high-quality matches were selected and are presented in *Table 1*



Table 1 Mapped and validated PIAAC questions

PIAAC Question Id	PIAAC Question Description	Label	Group	Type of variable
F_Q02b	instructing training or teaching people individually or in groups?	Teaching people	General	Frequency (time units)
F_Q02d	selling a product or selling a service?	Selling	General	Frequency (time units)
F_Q04a	persuading or influencing people?	Influencing people	General	Frequency (time units)
F_Q05a	[...] «Problem solving» [...]. How often are you usually faced by relatively simple problems that take no more than 5 minutes to find a good solution?	Simple problems	Problem solving	Frequency (time units)
G_Q01b	read letters memos or e-mails?	Read letters memos or mails	Literacy	Frequency (time units)
G_Q01g	read bills invoices bank statements or other financial statements?	Read financial statements	Literacy	Frequency (time units)
G_Q01h	read diagrams maps or schematics?	Read diagrams maps or schematics	Literacy	Frequency (time units)
G_Q02a	write letters memos or e-mails?	Write letters memos or mails	Literacy	Frequency (time units)
G_Q03b	calculate prices costs or budgets?	Calculating costs or budgets	Numeracy	Frequency (time units)
G_Q03c	use or calculate fractions decimals or percentages?	Use or calculate fractions or percentages	Numeracy	Frequency (time units)
G_Q03d	use a calculator - either hand-held or computer based?	Use a calculator	Numeracy	Frequency (time units)
G_Q03g	use simple algebra or formulas?	Use simple algebra or formulas	Numeracy	Frequency (time units)
G_Q03h	use more advanced math or statistics such as calculus complex algebra trigonometry or use of regression techniques?	Use advanced math or statistics	Numeracy	Frequency (time units)
G_Q04	you use a computer in your job?	Experience with computer in job	ICT	Yes (1) / No (2)
G_Q05a	use email?	For mail	ICT - Internet	Frequency (time units)
G_Q05d	conduct transactions on the internet for example buying or selling products or services or banking?	Conduct transactions	ICT - Internet	Frequency (time units)
G_Q05e	use spreadsheet software for example Excel?	Spreadsheets	ICT - Computer	Frequency (time units)
G_Q05f	use a word processor for example Word?	Word	ICT - Computer	Frequency (time units)
G_Q05g	use a programming language to program or write computer code?	Programming language	ICT - Computer	Frequency (time units)
I_Q04d	I like learning new things	Like learning new things	Learning strategies	Extents
I_Q04l	I like to figure out how different ideas fit together	Figure out how different ideas fit together	Learning strategies	Extents



As we can see in the Group column, we cover a variety of different domains: general, problem solving, literacy, numeracy, ICT (Internet, Computer) and learning strategies. We don't cover manual skills like dexterity or precision or such. This is one of the limitations of the current mapping.

We then decided to publish these data sets, so it can be downloaded and used freely by all the community: <https://crisp-unimib.github.io/PIAAC2ESCO/>

We also have a publication<sup>1</sup> and this was developed within the PILLARS consortium.

In the framework of this research with these colleagues, we used the PIAAC2 ESCO crosswalk to link data from OJA to PIAAC questions in order to derive some descriptive insights on skills mismatch on 17 European countries in 2019 and to go into the detail of the relation with automation and training.

On the one side we have PIAAC which comprises 250,000 observations (4,000 - 8,000 per country). It was run in 2012 and 2014. On the other side, the demand side we have WIH-OJA including 17,966,812 observations from 2019. Therefore, in order to have comparable data, we extrapolated the PIAAC observations of 2014 and 2017 to 2019 using changes in the US (observed in 2014 and 2017) as inflation parameters.

How can we measure the skill mismatch?

We follow two recent publications<sup>2</sup> that calculate for each skill and occupation, how relevant is that skill compared to other occupations. So, what we derived from PIAAC and OJAs is the relative frequency of observation of one skill per occupation. This is, in OJAs, for each job ad posted, while in PIAAC for each respondent, to compute whether this is more frequent compared to all other occupations.

We identify the degree of specialisation of occupation by observing the cases in which skills are relatively more requested in an occupation compared to the average demand in all occupation. This measurement can be done on the demand side (OJA, skills

---

<sup>1</sup> Guo, Y., Langer, C., Mercorio, F., & Trentini, F. (2022, September). Skills Mismatch, Automation, and Training: Evidence from 17 European Countries Using Survey Data and Online Job Ads. In *CESifo Forum* (Vol. 23, No. 5, pp. 11-15). Institut für Wirtschaftsforschung (Ifo). [[link](#)]

<sup>2</sup> Alabdulkareem, A., M. R. Frank, L. Sun, B. Al Shebli, C. Hidalgo and I. Rahwan (2018), "Unpacking the Polarization of Workplace Skills", *Science Advances* 4(7), DOI: 10.1126/sciadv.aao6030.

Giabelli, A., L. Malandri, F. Mercorio and M. Mezzanzanica (2022), "GraphLMI: A Data Driven System for Exploring Labor Market Information through Graph Databases", *Multimedia Tools and Applications* 81, 3061–3090. <https://dx.doi.org/10.1007/s11042-020-09115-x>



requested by employers in online job ads) and on the supply side (PIAAC skills used at work). Then we can calculate for each occupation-skill pair the distance (a simple difference) between the importance of skill requested on the market and the importance of that skill on the job.

We use the Relative Comparative Advantages (RCA) adapted to the case of survey and OJA<sup>3</sup>. RCA is a measure of specialisation. Formally, for Occupations  $\bar{O} = \{O_k, k=1, \dots, m\}$  and skills  $\bar{S} = \{S_j, j=1, \dots, l, \dots, p\}$  we calculate the relative skill-frequency  $sf$  for the occupation  $k$  – skill  $j$  pairs:

$$sf(o_k, s_j) = \frac{\sum_{i=1}^n I(o_i = o_k) \cdot I(s_i = s_j)}{\sum_{i=1}^n I(o_i = o_k)}$$

The skill frequency for OJA counts the number of ads of an occupation  $k$  that mention the skill  $j$  (numerator) on the number of total job ads observed in that specific occupation. For PIAAC we map the Likert scale<sup>4</sup> to an indicator variable such that if skill is used at least once a month it is set to value 1, 0 if it is never used.

The RCA of an occupation  $k$  - skill  $l$  couple is:

$$rca(o_i, s_l) = \frac{sf(o_i, s_l) / \sum_{j=1}^p sf(o_i, s_j)}{\sum_{k=1}^m sf(o_k, s_l) / \sum_{k=1}^m \sum_{j=1}^p sf(o_k, s_j)}$$

Therefore, the RCA compares the relative frequency of skill  $l$  in an occupation  $k$ , with the relative frequency of that same skill  $l$  in all other occupations. Its domain is  $[0, +\infty)$ , being 1 its neutral point: for values below this threshold the occupation is not specialised in that skill, while for values above the threshold it is specialised, i.e. the skill is more requested or used compared to the average in all occupations.

The RCA is not linear, so we map it to its percentile rank of belonging. For instance, with the example of the skill “teaching others” for occupation “Teaching professionals” we rank the skill among all occupations and find, that it ranks in the top

---

<sup>3</sup> Giabelli, A., Malandri, L., Mercorio, F., & Mezzanica, M. (2020). GraphLMI: A data driven system for exploring labor market information through graph databases. *Multimedia Tools and Applications*, 1-30.

<sup>4</sup> The response scale in PIAAC differs by the type of variable that is considered. In the cases relevant to us, there are three scales. The first, which is used to measure frequency in time units, is a 5-point Likert scale (1 - Never, 2 - Less than once a month, 3 - Less than once a week but at least once a month, 4 - At least once a week but not every day, 5 - Every day). The second scale is used to measure the extent to which the respondents think the statement represents them, again measured on a 5-point Likert scale (1 - Not at all, 2 - Very little, 3 - To some extent, 4 - To a high extent, 5 - To a very high extent), and the third scale is simply binary (Yes/No).





5% of the demand (OJA) for that specific skill. Analogously, on the supply side (PIAAC) we see that it also ranks in top 3% .

Our measure of skill gap is the difference between the RCA percentile rank in supply (PIAAC) and RCA percentile rank in demand (OJA) . In case of positive skill gaps we talk about skill shortage; vice versa we talk about skill surplus. [Indeed, it is appropriate to define them as under-skilling and over-skilling respectively]

PIAAC reports the occupation at the ISCO08 II digit, which constraints us to run the analysis at this level — OJA are currently disclosed at occupation ISCO08 digit IV. Concerning the geographical location, we have information at NUTS2 level on the PIAAC side — OJA have potentially detail at the city level, despite they have complete information at the NUTS2 level.

Skill gaps are therefore calculated at the ISCO08 II digit, skill, and country (or NUTS2 level) level.

[The subsequent part is described in detail in the publication Guo, Y., Langer, C., Mercorio, F., & Trentini, F. (2022, September). Skills Mismatch, Automation, and Training: Evidence from 17 European Countries Using Survey Data and Online Job Ads. In *CESifo Forum* (Vol. 23, No. 5, pp. 11-15). Institut für Wirtschaftsforschung (Ifo). [\[link\]](#)].

Through this method we have calculated the average skill gap by occupation (ISCO 2 digits), for the pool of 17 countries and mapped each occupation to a broad category of belonging in terms of task context (cognitive non-routine, cognitive routine, manual non-routine, manual routine).

What we observe is that there is a relation in terms of the belonging to a category and the intensity of the skill gap. In this case, the skill shortage is more prevalent among manual workers than cognitive workers.

We have also started to inspect the relation between skill shortages and on the job training, so trying to understand if there are relations between the distance in terms of importance in some skills between demand and supply. If there are mechanisms that are related to this divergence. What we observe is that on aggregate (pool of 17 countries), there is a negative relation between the share of participants in on-the-job training in 2012 (PIAAC survey) and the skill shortages that we observe 7 years later. The relationship with the type of tasks of workers is still very present, for example, teaching professionals have high surplus of skills compared to what is demanded and a high share of these workers participate in training, so there is a negative relation.

In the past, PIAAC has also been used by the OECD to calculate the risk of automation, so we have used the PIAAC items of the questionnaire to assign a likelihood of



possibility of automation. On aggregate, we put in relation the two measures and we obtain a positive relation between the skill shortage and the risk of automation. So, some occupations are more prone to changes (tasks changing inside jobs) and this changing content of jobs, makes the skill gap more pronounced.

They have also made a first inspection of the regional variation/ heterogeneity in these skill gaps. Not all regions are covered, but we observe a high heterogeneity between regions but also within country. There are a few countries that in general display skill surplus (BE, DE, FR, IE, SE), while for Greece for example there are skill shortages (except in the region of Athens). There are countries with a mix (ES, IT, LT, PL), some regions have skill shortages and in others just the opposite.

## Conclusions

- Web data can enrich and complement existing datasets (enrich the data and use AI to add new taxonomies and make the data linkable)
- AI-methods can simplify complex tasks and support human experts (putting in relation almost 14,000 ESCO skills and 80 skills not explicitly stated in a PIAAC questionnaire)
  - There is a cost reduction and bounded risk of non-systematic errors
  - We also need a transparent design and account of the process
- What we can learn is very relevant:
  - From the application we learn that jobs are changing in terms of tasks and the skill composition of jobs changes. Skill shortages are negatively related to training and positively related to automation risk.
  - There is a high heterogeneity at the country and regional level calls for more depth in understanding the role of institutional features of the countries labour markets.

## OPEN DISCUSSION

Claudia Plaimauer (3s Unternehmensberatung GmbH) understands that they are using two different measuring units, in PIAAC, there are respondents, real people but in OJAs we don't know how many vacancies there are, how many people might fill the advertisement. So how do they reconcile this?

Francesco Trentini (University of Milano-Bicocca, Italy) responds that on the job-ads side, we observe the number of posted ads with a certain characteristic, so we can actually count the number occurrences among job ads. The skill frequency is how many times we observe a specific skill-occupation pair in all job ads. So, we don't see the



number of people that apply but the number of ads employers post in that specific occupation that stated a specific skill.

Lorenzo Malandri (University of Milano-Bicocca, Italy) adds that in job ads we are interested in how many companies need a skill, the number of open positions, not the number of people that apply. Also, we observe job ads, not vacancies. Because inside the demand there a lot of phenomena occurring and we observe only a small part of the vacancies (not filled internally, etc.). So, we need to focus more on the intensity in the demand of specific skills rather than others which is more relevant than a matter of representativeness in terms of which vacancies will be filled.

Joanna Napierala (Cedefop) asks for clarification, if in the calculation of the RCA index did you include only the 21 skills that are matched, or all skills extracted. Francesco Trentini (University of Milano-Bicocca, Italy) replies that they used them all [Amend: on the OJA case we have calculated the RCA using all the skills and then filtered the relevant matches, while on the PIAAC side we first filtered the questions and then calculated the RCA. Therefore we will correct the procedure to make it consistent.]. Moreover, we have a great deal of relevant information that we are currently not using: the unmatched questions/skills in both PIAAC and OJA. We have not actively considered how to take them into account in the analysis but is one of the things that that they will further explore.

How many more skills are in the PIAAC database? In those 4 modules there are 84 questions, 70 could be used, all of them were matched but then only 21 were validated by the experts. So, they only kept those with higher quality.

Michel van Smoorenburg (UWV) is rather surprised by the results, especially the negative relation between on-the-job training participation and the skill shortages. He would expect a positive relation because when there are large skill shortages you may be motivated to improve and extend training participation within a company. Francesco Trentini (University of Milano-Bicocca, Italy) replies that on-the-job training is observed in 2012 and skill shortages are observed 7 years later, so there is a delay in the observation. So, we could interpret that the low share of participation in on-the-job training in 2012 led to a more pronounced skill shortage. So, it allowed to identify which sectors were less prone to skill shortages. The mechanisms are not clearly inferable from the results, we would need to go more into detail in the institutional context of each country, how firms behaved in that period... but one thing that emerges the more the workforce participates in the training, the more it is up to the request of the market. Which occupations are those that employers take more care of, there are also some dynamics, how training is deployed, the stability of contracts, many mechanisms that could be further explored. I would not say that training is effective in reducing skill shortages, that would be a bit too much, but there are some clear relationships that we could further explore.



Aleksandra Webb (University of the West of Scotland, UK) would want to know more about the expert validation process. Francesco Trentini (University of Milano-Bicocca, Italy) replies that there are embeddings involved. For each PIAAC question, the expert received the top 10 most similar ESCO skills suggested by the system and the expert had to vote on a Likert scale to which extent the ESCO skill correctly represents the PIAAC question. The text of the question was manipulated beforehand. So, the expert voted and then all the matches that were below the top level of the Likert scale were discarded. So, we kept those where all experts had the same judgement that the match was reliable. And who were the experts? They were a group of colleagues of the PILLARS consortium. The system could be further extended to include other stakeholders in the validation.

Michel van Smoorenburg (UWV) has a question about the graph of Europe, he is surprised by the results because when you look at the tightness of the labour market and the unemployment figures, you see major skill shortages in the northern and central countries, and a high level of employment in Spain, South of Italy, Greece... And in this graph, it is just the opposite. Francesco Trentini (University of Milano-Bicocca, Italy) replies that here what we observe is how is the workforce ready for the request of the market, we are not observing unemployed individuals. PIAAC helps us identify the skill use on the job, we are not observing the unemployed. This graph shows the relationship between the skills of workers and the demand of the market, so the importance that is given to the skills to perform a job and the demand of employers. Michel asks if we could say that in the southern countries the employee is underskilled and in the northern and central they are overskilled. Yes, Francesco says that this could be derived from the observation. Michel suggests validating these results by other research, through employer surveys, as this would be very interesting. Francesco thinks that there already exists research about this but doesn't know if there are already surveys. If anyone knows about this kind of surveys/ research, please let Francesco know.

Lorenzo Malandri (University of Milano-Bicocca, Italy) adds that an underskilled worker could be underskilled because of the education system or because the online job market requested too many skills (more than needed). Francesco Trentini (University of Milano-Bicocca, Italy) says that there is dimension that they had to discard when tagging which occupations do not use specific skills. They discarded all the part about the intensity on jobs which can be an interesting part of the research, this could help understanding if there is a relationship between the frequency of the use of a skill and how essential it is for an occupation. To use the gradient to know more about the essential skills, maybe different evidence could emerge.

Christa Larsen (IWAK Goethe University Frankfurt am Main, Germany) says that it is very interesting because the countries that have surpluses have strong VET systems, so



this could be further explored. She also wants to thank Francesco Trentini (University of Milano-Bicocca, Italy) for this approach because up until now we had some approaches either on one side or the other and this one is matching both. We always assume that there is a matching going on in the labour market, and with this methodology we try to reconstruct this matching, combining different type of data sets and developing methodologies to make them comparable. So, from the methodological side it is very interesting for us in the Network, we can learn a lot from such an approach.

Christa would also want to ask a question about the sample size of the survey data. Was there a problem going to NUTS 2 which such a small sample? Did you have to develop a method/ technology to handle this situation to be able to go to a regional level. Francesco Trentini (University of Milano-Bicocca, Italy) replies that there is of course a limitation from the sample size and that we are bounded to what we have. We left out poorly populated cells.

We did not develop any technology, it is methodological the same, but we can match at the lower level, in terms of occupation (ISCO08 II digit), region (NUTS2) and skills (ESCO Skill fourth hierarchical level).

Francesco Trentini (University of Milano-Bicocca, Italy) adds that we are working on using AI to bridge national classifications to ESCO, to provide a sound methodology that can help us align different classifications to create crossworks and this is one of the dimensions on the use of these technologies which can really help improve and enrich the existing surveys with external data sources. Using web data has the strong limitation of not knowing which population you are talking about, you have to make a lot of assumptions. Surveys as a starting point can be a very valuable and having a standard to enrich the survey.

In 2023 [amend: 2024] there will be a new release of PIAAC.

Lorenzo Malandri (University of Milano-Bicocca, Italy) highlights 3 points:

- Use of data, benefit of having quality data
- Linked data, find the right dimension to match the data
- Interaction between AI and human is necessary, AI can be the telescope and the human can do the validations

Christa Larsen (IWAK Goethe University Frankfurt am Main, Germany) asks if the methodology would be available for Network members to getting inspired, it would be a good basis to explore further. Trying to reconstruct the dynamic between the supply and the demand, it would help a lot to have a description of the methodology. Francesco



Trentini (University of Milano-Bicocca, Italy) replies that concerning PIAAC to ESCO, there is a general description and a technical annex available on the website. Francesco has added more details about the methodology throughout this document so that the Network members can better understand the method.

The next seminar of the series will be delivered by Switzerland, the 23rd of February 2023 under the topics:

- Improving refugee integration through data-driven algorithmic assignment
- Implementation of Risk Assessment Tools in the Public Employment Service

More information will be circulated as always through the Network's newsletter. Also, if you are interested in sharing your experiences with the Network, please do so through the Knowledge Hub <https://bigdatahub.uvt.ro/> and contact Eugenia Atin if you wish to present your project through a Seminar.

*Bilbao, December 2022*

#### **References**

##### **Speaker: FRANCESCO TRENTINI**

University of Milano-Bicocca, CRISP and LABORatorio R.Revelli , Italy

Email: [francesco.trentini@unimib.it](mailto:francesco.trentini@unimib.it)

##### **EUROPEAN NETWORK ENRLMM**

Email: [jenny@jennykipper.de](mailto:jenny@jennykipper.de)

Email: [c.larsen@em.uni-frankfurt.de](mailto:c.larsen@em.uni-frankfurt.de)

Website: [www.regionallabourmarketmonitoring.net](http://www.regionallabourmarketmonitoring.net)

##### **BIG DATA WORKING GROUP**

Eugenia Atin

E-mail: [e.atin@prospektiker.es](mailto:e.atin@prospektiker.es)

Website: <https://bigdatahub.uvt.ro/>

# Skills mismatch across EU countries using job ads and survey data at the micro-level

Francesco Trentini

University of Milano-Bicocca, CRISP and LABORatorio R.Revelli



**CRISP**

*Interuniversity  
Research Centre  
for Public Services*



**LABOR**  
LABORATORIO RICCARDO REVELLI  
CENTRE FOR EMPLOYMENT STUDIES

# Overview

- Skill mismatch. Definition and used measures
- Novel approach using
  - PIAAC – Program for the International Assessment of Adult Competences (OECD)
  - Online job ads from WIH-OJA (Cedefop and Eurostat)
  - Bridging them with ESCO Skills pillar using AI
- Application of the mapping to measure skill mismatch in 17 European countries

The research has been developed in the framework of the EU funded H2020 *Pillars – Pathways towards inclusive labour markets*.



# Skill mismatch

**Def** deviation of workers' competencies from those required by employers to perform a job.

Supply

Demand

# How to identify skill mismatch?

**Identification** of supply and demand in empirical research suffers limitations:

- Proxy for supply is usually highest level of education attained
- Demand is measured with surveys directed to employers to state which occupations or competences are needed

**Approaches** to measurement:

- **Normative:** ask experts to state the needed level of education or qualification to perform a job
- **Objective:** use the distribution of realised matches and measure the distance from, e.g., the median
- **Subjective:** workers are asked to self-assess the coherence between their education/qualification and the task that perform on the job

**Limitations:** low granularity and high heterogeneity within education classes

Berton, F., Devicienti, F., & Grubanov-Boskovic, S. (2017). Employment Protection Legislation and Mismatch: Evidence from a Reform. IZA Discussion Papers No. 10904

# Novel approach using survey and big data

## Data

Supply Program for the International Assessment of Adult Competences (OECD)

Demand Online Job Ads (WIH-OJA, Eurostat and Cedefop)

## Methods

- For each occupation-skill compute the degree of specialisation (Revealed Comparative Advantages) on the demand and supply side
- Calculate the skill gap and average at the occupation level.
- Insight at the occupation 2 digit and NUTS2 in relation to training and risk of automation.

## Limitation

PIAAC and OJA express skills with different taxonomies. How can we bridge the two?

# Skill mismatch: data: supply - PIAAC

## PIAAC – Program for the International Assessment of Adult Competences (OECD)

- Cycle I, all rounds: 2012, 2014 and 2017. Cycle II, currently ongoing; data release expected in 2023.
- Representative samples of working-age individuals
- Background questionnaire + test of the performance in some activities (e.g. writing an email, search for a job online)
- **Background questionnaire:** self-declared intensity (frequency or extent) of **skill use** in different domains:
  - at work (Module F)
  - Literacy, Numeracy and ICT at work (Module G)
  - Literacy, Numeracy and ICT in everyday life (Module H).
  - And a module on «[...] about how you deal with problems and tasks you encounter.» (Module I)
- PIAAC is used in the literature to measure skill content in a global perspective (Lewandowski et al. 2022) and risk of automation (Nedelkoska and Quintini, 2018)

Lewandowski, P., Park, A., Hardy, W., Du, Y., Wu, S. (2022) Technology, Skills, and Globalization: Explaining International Differences in Routine and Nonroutine Work Using Survey Data, *The World Bank Economic Review*, <https://doi.org/10.1093/wber/lhac005>

Nedelkoska, L. and G. Quintini (2018). Automation, Skills Use and Training. Technical report, OECD, Paris.

# Skill mismatch: data: supply - PIAAC

## F\_START Skills used at work

Layout	Item group table
<b>F_Q02b</b> (JRA) (B)	<b>How often ^DoesDid your ^JobLastjob usually involve ... instructing, training or teaching people, individually or in groups?</b>
	1 Never
	2 Less than once a month
	3 Less than once a week but at least once a month
	4 At least once a week but not every day
	5 Every day
	DK
	RF

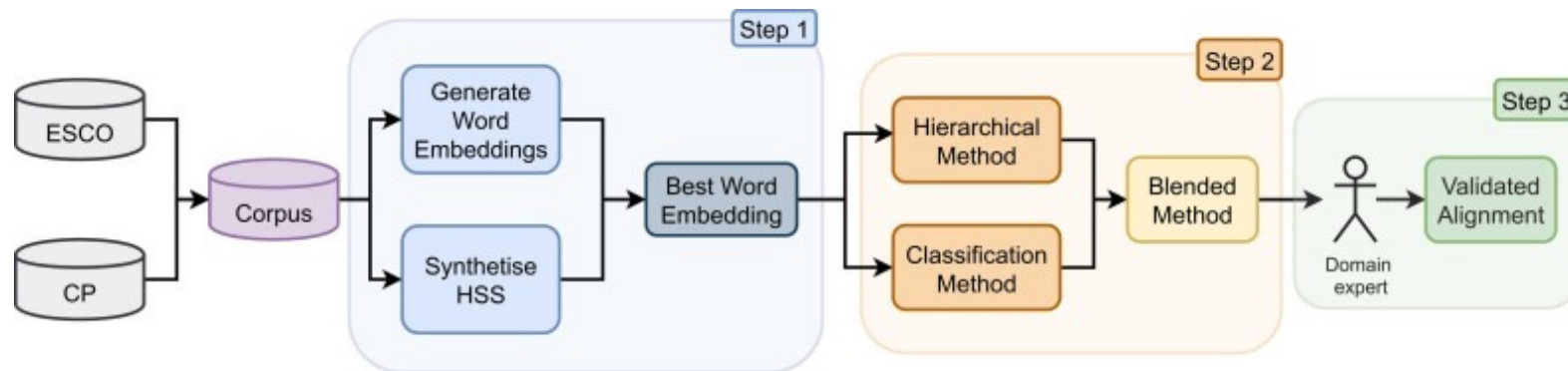
# **Skill mismatch: data: demand - OJA**

## **Online Job Ads (WIH-OJA, Eurostat and Cedefop)**

- Collection of online job ads from 27 European countries + UK and EFTA countries. Since 2018Q4
- Data on occupations and related skills as they emerge from online job postings
- Skills are extracted based on the ESCO Skill Pillar

# How to link PIAAC and OJA? Use ESCO and AI

- The linkage is done using AI in a framework that combines various methods: embeddings, selection of the best embedding, taxonomy alignment and experts' validation
- PIAAC questions are processed to tag the most similar ESCO Skills.
- The embedding is trained on OJA UK data and the matching is done on the English language.



Giabelli, A., Malandri, L., Mercorio, F., & Mezzanzanica, M. (2022). WETA: Automatic taxonomy alignment via word embeddings. *Computers in Industry*, 138, 103626.

# PIAAC2ESCO mapping

PIAAC Question Id	PIAAC Question Description	Label	Group	Type of variable
F_Q02b	instructing training or teaching people individually or in groups?	Teaching people	General	Frequency (time units)
F_Q02d	selling a product or selling a service?	Selling	General	Frequency (time units)
F_Q04a	persuading or influencing people?	Influencing people	General	Frequency (time units)
F_Q05a	[...] «Problem solving» [...]. How often are you usually faced by relatively simple problems that take no more than 5 minutes to find a good solution?	Simple problems	Problem solving	Frequency (time units)
G_Q01b	read letters memos or e-mails?	Read letters memos or mails	Literacy	Frequency (time units)
G_Q01g	read bills invoices bank statements or other financial statements?	Read financial statements	Literacy	Frequency (time units)
G_Q01h	read diagrams maps or schematics?	Read diagrams maps or schematics	Literacy	Frequency (time units)
G_Q02a	write letters memos or e-mails?	Write letters memos or mails	Literacy	Frequency (time units)
G_Q03b	calculate prices costs or budgets?	Calculating costs or budgets	Numeracy	Frequency (time units)
G_Q03c	use or calculate fractions decimals or percentages?	Use or calculate fractions or percentages	Numeracy	Frequency (time units)
G_Q03d	use a calculator - either hand-held or computer based?	Use a calculator	Numeracy	Frequency (time units)
G_Q03g	use simple algebra or formulas?	Use simple algebra or formulas	Numeracy	Frequency (time units)
G_Q03h	use more advanced math or statistics such as calculus complex algebra trigonometry or use of regression techniques?	Use advanced math or statistics	Numeracy	Frequency (time units)
G_Q04	you use a computer in your job?	Experience with computer in job	ICT	Yes (1) / No (2)
G_Q05a	use email?	For mail	ICT - Internet	Frequency (time units)
G_Q05d	conduct transactions on the internet for example buying or selling products or services or banking?	Conduct transactions	ICT - Internet	Frequency (time units)
G_Q05e	use spreadsheet software for example Excel?	Spreadsheets	ICT - Computer	Frequency (time units)
G_Q05f	use a word processor for example Word?	Word	ICT - Computer	Frequency (time units)
G_Q05g	use a programming language to program or write computer code?	Programming language	ICT - Computer	Frequency (time units)
I_Q04d	I like learning new things	Like learning new things	Learning strategies	Extents
I_Q04l	I like to figure out how different ideas fit together	Figure out how different ideas fit together	Learning strategies	Extents



# PIAAC2ESCO validated dataset

- The validated dataset covers **21 PIAAC questions** and the **mapped ESCO skills, enriched** with alternative labels

F\_Q02b: instructing training or teaching people individually or in groups?

- coach young people
- coach youngsters
- educate others
- educate young people
- facilitate young peoples education
- facilitate young peoples mentoring
- instruct colleagues
- instruct others
- instruct young people
- instructing others
- teach others
- teach young people
- train others
- train young people
- train youngsters
- tutoring

# PIAAC2ESCO – Open data

<https://crisp-unimib.github.io/PIAAC2ESCO/>

Open access to:

- Dataset – enriched mapping
- Methodological annex

## PIAAC2ESCO - An AI-driven classification of the PIAAC Background questionnaire onto the ESCO Skills Pillar

[View on GitHub](#)

### PIAAC2ESCO - An AI-driven classification of the PIAAC Background questionnaire onto the ESCO Skills Pillar

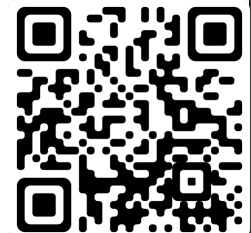
#### What is PIAAC2ESCO?

PIAAC2ESCO provides a characterisation of the [PIAAC background questionnaire](#) on the base of the [ESCO Skills Pillar](#). In practice it associates a list of ESCO skills (v1) to questions of the PIAAC background questionnaire (version 2010), based on their similarity. We use the section F to I of the PIAAC background questionnaire, from which we select the relevant questions (73 questions out of 84) and all the ESCO skills (13600 items). The validated dataset covers 21 PIAAC questions and the mapped ESCO skills, which are enriched using alternative labels.

#### How does PIAAC2ESCO work?

The linkage is done using AI in a framework that combines various methods: embeddings, selection of the best embedding, taxonomy alignment and experts' validation. A description of the adopted methodology is available in the [Technical Annex](#).

The training dataset of the embedding is the representative sample of the job ads collected by Eurostat and Cedefop as part of the [Web Intelligence Hub - Online Job Advertisements \(WIH-OJA\)](#)



# Publication

Yuchen Guo & Christina Langer & Fabio Mercurio & Francesco Trentini, 2022. "[Skills Mismatch, Automation, and Training: Evidence from 17 European Countries Using Survey Data and Online Job Ads](#)," [EconPol Forum](#), CESifo, vol. 23(05), pages 11-15, September.

**econPOL FORUM** 5 2022 September Vol. 23

**ECONOMIC POLICY AND ITS IMPACT**  
What Works? Regional Effects of Universities  
Nicolas Bunde, Nina Czernich, Oliver Falck and Gabriel Felbermayr

**INSTITUTIONS ACROSS THE WORLD**  
Institutions – Moving to the Global?  
Harold James

**BIG-DATA-BASED ECONOMIC INSIGHTS**  
How Do Taxation and Regulation Affect the Real Estate Market?  
Mathias Dolls, David Gstrein, Carla Kralage and Florian Neumeier

**POLICY DEBATE OF THE HOUR**  
**European Labor Markets: How Can We Effectively Manage Technological and Structural Change?**

**Yuchen Guo, Christina Langer, Fabio Mercurio and Francesco Trentini**  
**Skills Mismatch, Automation, and Training: Evidence from 17 European Countries Using Survey Data and Online Job Ads<sup>1</sup>**

Labor markets face challenges due to globalization, structural transformation, and advancing technological change. This can lead to skills gaps and skills mismatch between firms' skill demand and employees' skill supply, which can go in two directions: workers having a skill surplus, where skill supply exceeds demand, or workers experiencing skill shortage, where firms' skill demand is greater than the skills workers actually possess. In light of this, the EU's Agenda for New Skills and Jobs (European Commission 2020) states that creating a more skilled workforce "is a considerable challenge given the rapidly-changing skills needed, and the persistent skills mismatches in the EU labor market." In this report, the Commission also "established the anticipation and matching of labor market and skills needs as a top priority for the EU."<sup>2</sup>

However, empirical evidence on the prevalence of skills mismatch between the skills requested by employers and the skills provided by employees across Europe is scarce. We contribute to the understanding of skills mismatch in the EU by presenting novel evidence on skills gaps across countries, occupations, and skill domains using innovative job ad data and survey data for 16 EU countries and the UK.<sup>3</sup> In particular, we leverage two different data sources: online job vacancy data on skills requested by employers and survey data on skills supplied by workers. We document four key findings: first, skill gaps in the European Union exist, but the extent and direction vary across occupation types: workers in cognitive intensive occupations provide more skills than are demanded (skill surplus), whereas workers in manual intensive occupations face higher skill demand compared to the skills they have (skill shortage). Second, this pattern is consistent across almost all 17 countries that are part of our analysis. This suggests that overall patterns of skills mismatch do not reflect country-specific factors but are rather a European-wide phenomenon. Delving deeper into different skill domains (i.e., digital, numeracy, literacy, and social skills), we document similar skills gaps for different occupation types. Thus, the ob-

erved skills gaps are not driven by a lack of specific skill domains such as digital skills. Finally, we investigate potential mechanisms – i.e., an occupation's automation probability and workers' propensity to participate in on-the-job training – that might underlie the observed patterns of skills mismatch.

**LINKING SURVEY DATA AND ONLINE JOB ADS OFFERS NEW INSIGHTS INTO SKILLS GAPS IN THE EU**

We propose a novel measure of the gap between the skills demanded by employers and the skills supplied by workers. On the demand side, we rely on online job vacancies (OJV) data from the European Center for the Development of Vocational Training (CEDEFOP), collected in 2019, to capture skills de-

**KEY MESSAGES**

- Linking survey data and online job ads offers new insights into skills gaps in the EU
- Matching labor market needs and skill supply remains a Europe-wide challenge
- Manual workers have skill supply shortages, cognitive workers have skill supply surpluses
- Workers at higher risk of automation experience higher skill shortages, potentially because their job tasks are changing more rapidly
- On-the-job training might be a potential measure to meet future skills needs

**Yuchen Guo**  
is a Ph.D. candidate in Economics at KU Eindhoven University and guest researcher at the Ifo Institute.

**Christina Langer**  
is a Ph.D. candidate in Economics at KU Eindhoven University; guest researcher at the Ifo Institute, and Visiting Research Fellow at the Project on Workforce at Harvard University.

CESifo ifo INSTITUTE

EconPol Forum 5/2022 September Volume 23 11

# Skill mismatch across Europe

Descriptive insights on skills mismatch 17 European countries in 2019. Relation with automation and training.

## Samples

- PIAAC comprises 250,000 observations (4,000 - 8,000 per country). 2012 and 2014. Projected to 2019 using changes in US (observed in 2014 and 2017) as inflation parameters.
- WIH-OJA includes 17,966,812 observations. 2019.

# Skill mismatch - measure

We follow Alabdulkareem et al. (2018) and Giabelli et al. (2022) and calculate the relevance of each skill for each occupation.

---

## Skill frequency

**Occupations**  $\bar{O} = \{O_k, k=1, \dots, m\}$

**Skills**  $\bar{S} = \{S_j, j=1, \dots, p\}$

$$sf(o_k, s_j) = \frac{\sum_{i=1}^n I(o_i = o_k) \cdot I(s_i = s_j)}{\sum_{i=1}^n I(o_i = o_k)}$$

PIAAC use number of respondents - by occupation-question.  
From Likert scale to Indicator of use

OJA use number of ads observed - by occupation-skill

---

## Revealed Comparative Advantage

$$rca(o_i, s_l) = \frac{sf(o_i, s_l) / \sum_{j=1}^p sf(o_i, s_j)}{\sum_{k=1}^m sf(o_k, s_l) / \sum_{k=1}^m \sum_{j=1}^p sf(o_k, s_j)}$$

Alabdulkareem, A., M. R. Frank, L. Sun, B. Al Shebli, C. Hidalgo and I. Rahwan (2018), "Unpacking the Polarization of Workplace Skills", Science Advances 4(7), DOI: [10.1126/sciadv.aao6030](https://doi.org/10.1126/sciadv.aao6030).

Giabelli, A., L. Malandri, F. Mercorio and M. Mezzanzanica (2022), "GraphLMI: A Data Driven System for Exploring Labor Market Information through Graph Databases", Multimedia Tools and Applications 81, 3061–3090.

# Skill mismatch - measure

For each skill in demand and supply, the RCA is ranked **among all occupation** and mapped to the percentile of belonging.

Our **mismatch measure** at the occupation level is **the mean RCAs-percentile-rank gap between demand and supply**. Negative values indicate skill surplus, vice versa positive values indicate skill shortage.

An example:

F\_Q02b: “[...] teaching people individually or in groups?” → ESCO skill: “Teaching others”

$RCA_{oja}$  percentile rank: 0.95

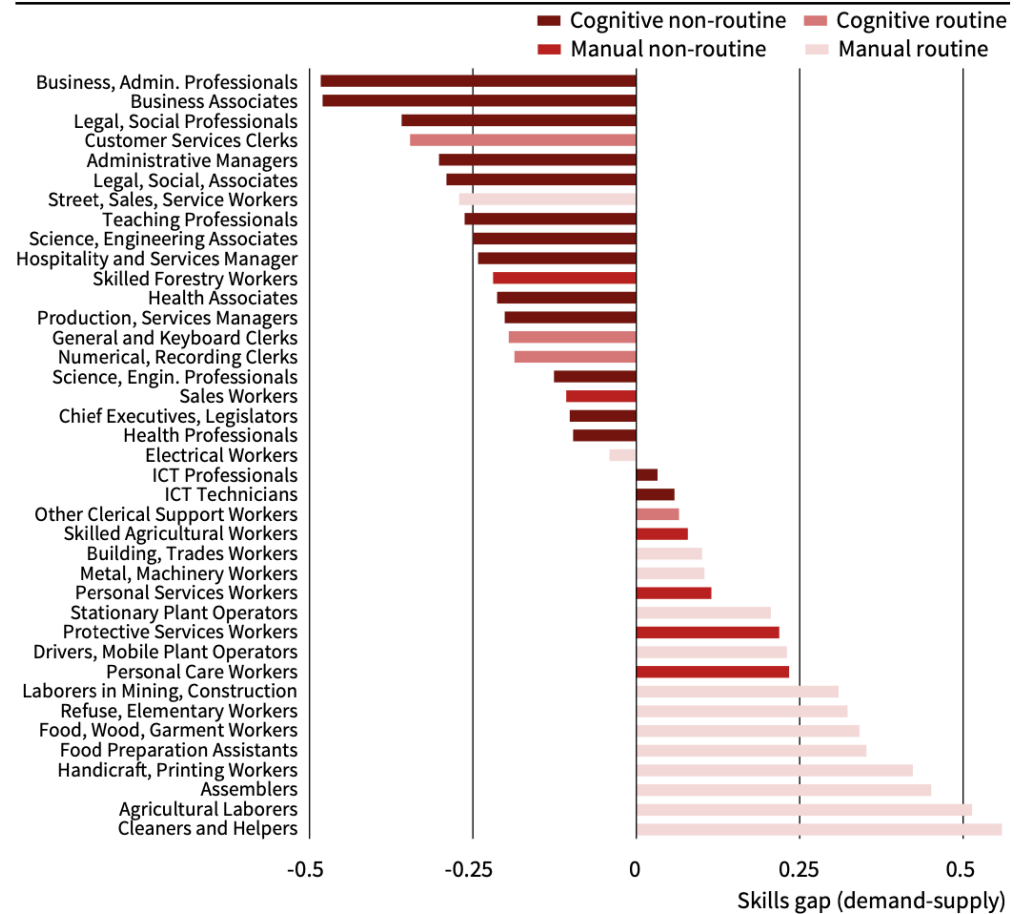
$RCA_{piaac}$  percentile rank : 0.97

Skill mismatch:  $pRCA_{oja} - pRCA_{piaac} = -0.02 < 0$  Skill surplus

# Findings

Skill shortage impacts manual workers more than cognitive workers

Average Skill gap by occupation ISCO08 II digit. 2019 pooled 17 countries.



Autor, D. H., F. Levy, and R. J. Murnane (2003). The Skill Content of Recent Technological Change: An Empirical Exploration. The Quarterly Journal of Economics 118 (4), 1279–1333.

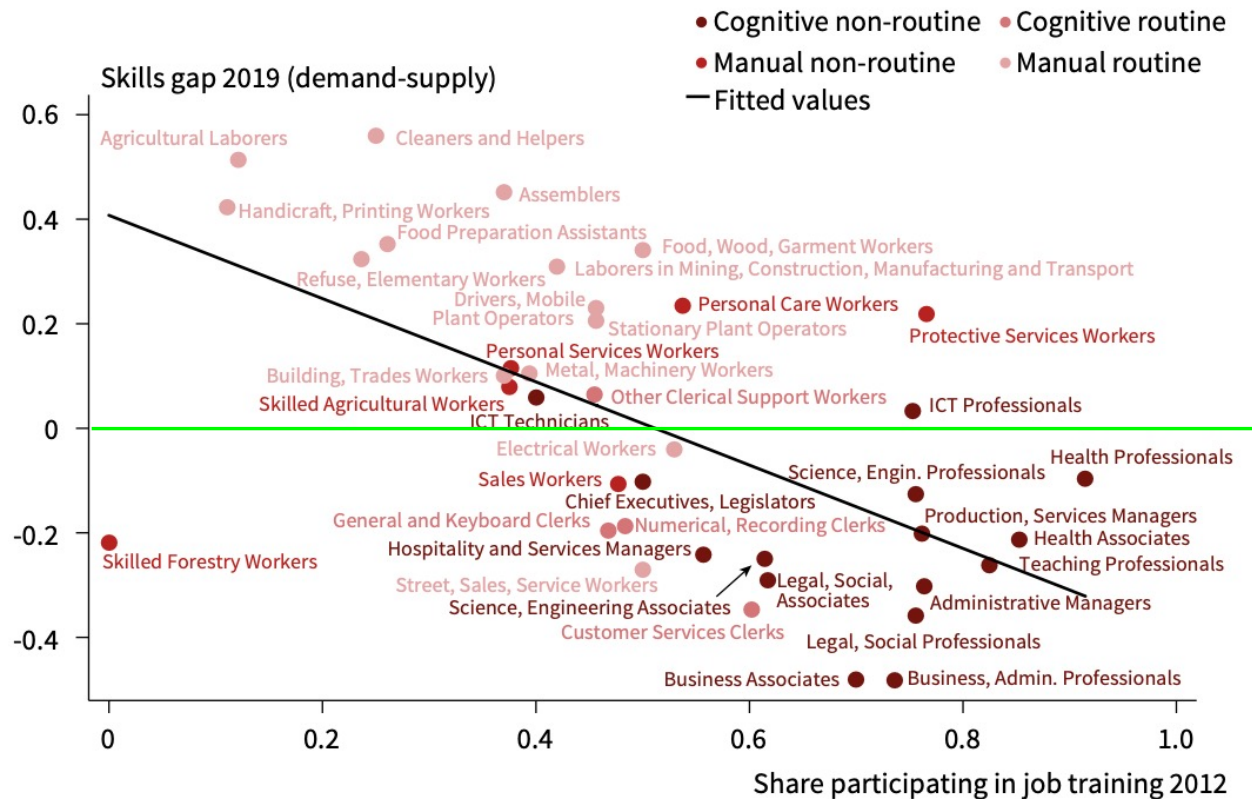
Note: Pooled for 17 European countries.  
Source: CEDEFOP; PIAAC.

© ifo Institute

# Findings

On aggregate, on-the-job training is negatively related to skill shortage

On-the-job training (2012) and skills shortage (2019). Pooled, 17 countries



Note: Correlation between on-the-job training (measured in 2012) and the skills gap (measured in 2019), pooled for 17 European countries.

Source: CEDEFOP; PIAAC.

© ifo Institute



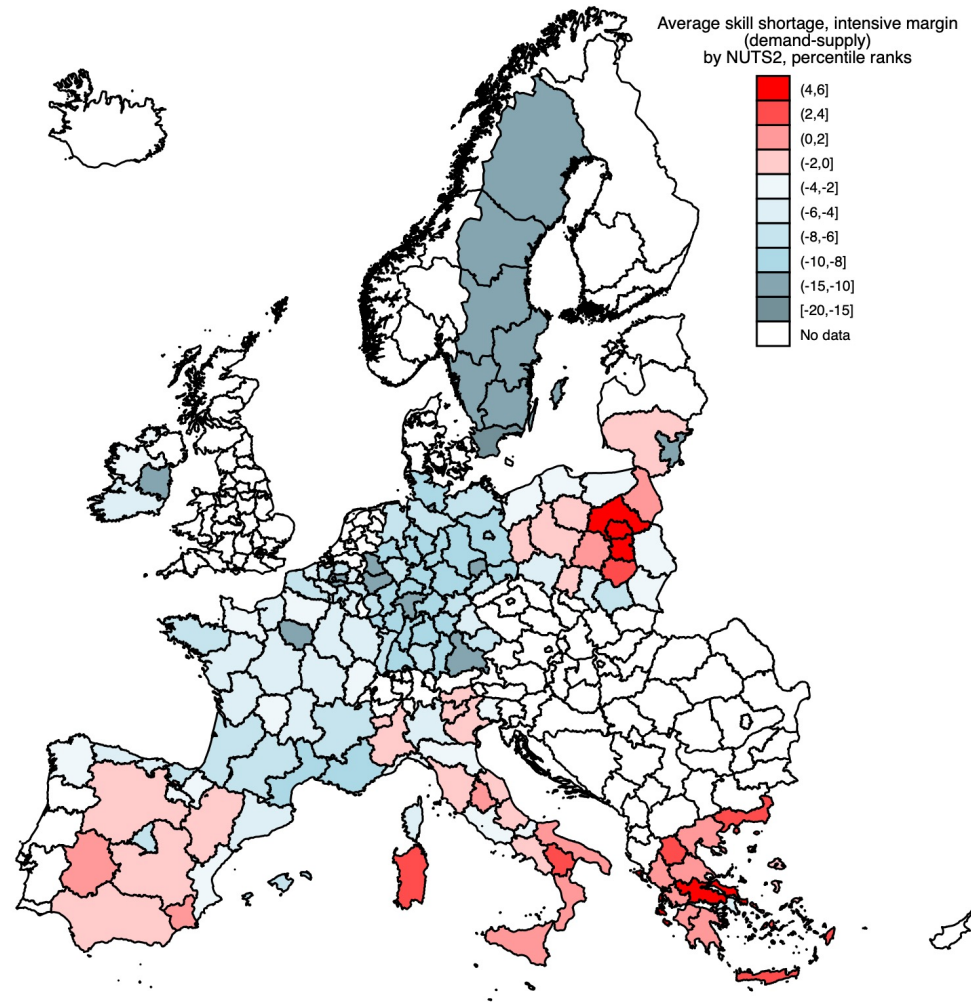


# Findings

Descriptive figures show high heterogeneity among European regions.

Within-country heterogeneity is also relevant

- Surplus: BE, DE, FR, IE, SE
- Shortage: EL
- Mix: ES, IT, LT, PL



# Conclusions

Web data can enrich and complement existing datasets.

AI-methods can simplify complex tasks and support human experts

- **Pros:** cost reduction and bounded risk of non-systematic errors
- **Caveat:** need of transparent design and account of the process

Jobs are changing in terms of tasks and the skill composition of jobs changes. Skill shortages are negatively related to training and positively related to automation risk.

High heterogeneity at the country and regional level calls for more depth in understanding the role of institutional features of the countries labour markets.

**Thank you for your attention! Questions?**

Francesco Trentini

francesco.trentini@unimib.it