Procedure of gathering and analysing data

At the last day of each month during the studied period, all job offers were downloaded from the studied websites (see Table 2 in the Appendix of the manuscript) and stored in a local database in the original HTML format. First, we applied parsing to the data: the data stored in the standard HTML format were converted into plain text. Parsing removes such information as font size, colour, and other formatting tags.

After parsing, we lemmatised the data, that is, we identified the basic forms of all the words in the text (for example, verbs were transformed to their infinitives while nouns to their nominative cases in singular forms). We lemmatised only Polish words. For this, we developed a tool based on the Morfologik-stemming-1.9.0 library. The results were shown in lemmatisation.xlsx.

[see lemmatisation.xlsx]

Because job offers can contain various words to describe the same competence, the algorithm should deal with the following situations:

a) the competence is mentioned in the sentence;

b) the competence is mentioned in the sentence, but with different words than in the dictionary.

To solve this problem, we first prepared a list of competences (based on the European Commission list of transversal competences, likely the most detailed such classification). Our approach enables us to obtain data on individual competences. Its another advantage is that it helps us detect various terms companies use to describe competences, enabling us to increase the dictionary's size. For this, we created a dictionary of synonyms based on three dictionaries (www.slownik-synonimow.eu, www.synonimy.pl, and sjp.pwn.pl). With a larger dictionary—with more phrases (including colloquial ones) describing competences and their groups—we will obtain more accurate representation of the competences demand in the labour market.

For some words forming the competences, this library provides inconsistent lemmatised words. To deal with such situations that we spotted during the analysis, we created an additional dictionary of exceptions. At this stage, our most important task was to find specific competences in the texts of job offers.

[see dictionary.docx]

Now job offers and the dictionary of competences had the same form, so we were able to search the offers for the competences. We noticed that some of the competences were too specific for employers, for example, 'explain mathematical findings,' or 'make appropriate use of eye contact'; instead of them, employers demanded competences defined in a general way. Thus, we also searched job offers for classes of competences. For the two above examples, these were 'communicate mathematical work process' (which includes 'explain mathematical findings') and 'use non-verbal cues' (which includes 'make appropriate use of eye contact'). While none of these two competences occurred in job offers, the classes they are from did occur. We created an example.xlsx file, wherein we show (all lemmatised):

i. a job offer id;
ii. job position description;
iii. job position details;
iv. a list of identified phrases (transversal competences).

[see example.xlsx]

We present a number of competences per job offer. In the case of categories, subcategories, and classes of competences, one job offer can include more than one competence of a given type. That is why the number of competences per job offer in these groups of competences cannot be interpreted as shares. If a single competence occurred twice in one job offer (for example because the offer was for two positions), we counted it once. So, for a given competence, we can interpret its number per job offer as a share of job offers that contain the competence.

[see data.xlsx]

We study co-occurrence of competences by counting how many times their pairs (over 32 thousand) and triplets (over 513 thousand) occurred together in the same job offer.

[see pairs_of_phrases.csv and triplets_of_phrases.csv]

To analyse similarities in time changes between classes of competences, we apply Pearson's correlation coefficient for monthly occurrences of the classes in job offers. We also supplied additional figures, showing changes in time of chosen competences and classes of competences.

[see additional_figures.docx]

The data were analysed with exploratory data analysis and graphical methods in Microsoft Excel and R (R Core Team 2016).