# Job Creation and Demand for Skills in Kosovo

What Can We Learn from Job Portal Data?

Calogero Brancatelli Alicia Marguerie Stefanie Brodmann

WORLD BANK GROUP Social Protection and Jobs Global Practice June 2020

#### Abstract

In Kosovo, employers report significant skill shortages, which limits firm growth and job creation. To understand the labor market dynamics and employer needs in real time, this paper analyzes the content of job postings using data from major online job portals from 2018. The findings show that the skills that are most in demand are socioemotional skills (especially related to extraversion), foreign language skills, and computer skills. The importance of these skills is transversal, cutting not only across occupations and industries, but also universally demanded in all education fields. The need for these skills is expressed more often and more explicitly in postings for jobs requiring higher levels of experience. Moreover, job platforms are used almost exclusively for filling high-skill occupations, especially in Kosovo's capital city, Pristina, whereas many low- and medium-skill jobs and jobs outside the capital are filled through informal channels. Overall, online data can be a useful tool for policy makers and other stakeholders to help align career services, training programs, and educational curricula with the skill needs of firms in real time.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

This paper is a product of the Social Protection and Jobs Global Practice. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://www.worldbank.org/prwp. The authors may be contacted at amarguerie@worldbank.org, sbrodmann@worldbank.org and brancatelli@econ.uni-frankfurt.de.

## Job Creation and Demand for Skills in Kosovo: What Can We Learn from Job Portal Data?<sup>1</sup>

Calogero Brancatelli, *Goethe University Frankfurt* Alicia Marguerie, *World Bank* Stefanie Brodmann, *World Bank* 

JEL classification codes: J23, J24, J60

Keywords: Jobs, Skills demand, Online job portal, Kosovo.

<sup>&</sup>lt;sup>1</sup> Calogero Brancatelli, Goethe University Frankfurt: <u>brancatelli@econ.uni-frankfurt.de</u>; Alicia Marguerie, Social Protection and Jobs, World Bank: <u>amarguerie@worldbank.org</u>; Stefanie Brodmann, Social Protection and Jobs, World Bank: <u>sbrodmann@worldbank.org</u>. The authors acknowledge financial support from the Multi-Donor Rapid Social Response (RSR) Trust Fund. The authors are grateful for technical support from Arion Rizaj, Shpat Ferizi (KosovaJob/HumanPower) and Fitim Krasniqi, and for comments received from Shpetim Kalludra (Employment Agency of Republic of Kosovo), Kevin Hempel, Mattia Makovec, Harry Munoz, Monica Robayo, Abla Safir and Mauro Testaverde (World Bank), Visar Rexha and Timothy Sparkman (EYE project), as well as participants from the World Bank Youth Employment Conference held in Pristina in May, 2019. This paper is a product of the staff of the International Bank for Reconstruction and Development/The World Bank. The findings, interpretations, and conclusions expressed in this paper do not necessarily reflect the views of the Executive Directors of The World Bank or the governments they represent. The World Bank does not guarantee the accuracy of the data presented in this work.

### I. Introduction

Understanding firms' needs and constraints in terms of workforce and related skills is key to designing better education and labor market policies. In Kosovo, a perception of skill shortages among firms and an oversupply of labor, especially youth entering the labor market, coexist. Aligning policies, training programs and curricula to the current and future needs of the labor market is challenging in the absence of relevant and real-time data. Data on labor supply and the results of job matches are available from standard surveys such as the Labor Force Survey, while detailed data on job creation and skill needs are often scarce. This paper exploits online vacancy data from the four largest job portals in Kosovo in 2018 to better understand its potential for addressing important information gaps between labor market demand and skills.

The key contribution of this paper is to provide an analysis of skill demand with a high level of granularity and precision, such that our results can directly inform policy making. We conduct a textual analysis of the job descriptions and job titles to identify the incidence of skills, education, and experience requirements across industries. For this purpose, we construct search dictionaries of specific keywords, key phrases and text patterns to approximate demand for skills, education, and work experience with the occurrence of these in job portal ads. For the skill analysis, we develop a skill taxonomy that consists of three layers, providing detailed information on skills in demand. First, we employ a general taxonomy that is in line with the literature on skills to show how demand varies across socio-emotional, cognitive, and technical skills.<sup>2</sup> Second, we disaggregate these three broad types of skills into 15 categories (one cognitive skill, five socio-emotional skills, and nine technical skills) to investigate how demand for specific skills varies across industries and occupations. Third, each of the 15 categories is defined by 152 skill requirements, which offers the highest level of granularity in the analysis of skills in demand. This layer is especially useful for providing youth with information about the skills employers seek, and for informing curricula on the specific skills that are needed in the labor market (e.g., which software or coding language is used in the computer science field). To precisely measure each skill requirement, we build a set of language-specific key words, key phrases and text patterns, which allow to identify the skill requirement in the job descriptions. Notably, such granularity and flexibility in an analysis of skill demand is only possible when using job ads directly and accounting for various possible ways of expressing skills in the text.

The main findings can be summarized as follows. To begin, extraversion, computer and foreign language skills are most in demand across all industries and occupations. Socio-emotional skills as a whole have a high

 $<sup>^{2}</sup>$  The most oft-used framework in the literature of skill formation focuses on cognitive versus non-cognitive skills (Cunha et al., 2006; Heckman et al., 2006; Kautz et al., 2015). However, given that we are interested in the skills required by employers, we focus on job-related skills following the literature on skills at work. Pierre et al. (2014) and World Bank (2018) use a similar categorization, which reflects skills developed later in life from secondary schooling to young adulthood, through general education, technical education or on-the-job training.

prevalence across all industries and are thus transversal skills for this segment of the labor market. Some industries such as the IT sector express high demand for particular software skills, thereby signaling specific industry needs with which curricula could be better aligned. Firms are more likely to express education or diploma requirements rather than work experience when searching for potential workers through job ads. On job platforms, employers require on average 2.5 years of work experience and tertiary education, suggesting that platforms are of limited use to labor market entrants. Employers who require several years of work experience are also those who express demand for a higher number of skills. Finally, we establish that job platforms in Kosovo represent a specific segment of the labor market. They mostly require a minimum level of experience and are used almost exclusively for high-skill occupations in Pristina, the capital city and are concentrated on permanent contracts and full-time jobs, whereas low- and medium-skill jobs are more likely to be filled through informal channels.

Information from job platforms has great potential as it provides useful real-time data complementing traditional sources on labor and skill demand across occupations and sectors.<sup>3</sup> Our paper relates to the literature analyzing job descriptions to better understand labor demand. Recently, Hershbein and Kahn (2019) have used job descriptions from US posted vacancies collected by Burning Glass Technologies to identify "upskilling" by looking at changes in firms' skills requirement in job postings during recession times. Kureková et al. (2016) analyze the text of vacancies in the Slovak Republic to compare the profiles in student-targeted vacancies with other flexible jobs. Similarly, we use job descriptions to shed light on the skills demanded by employers in Kosovo. Muller and Safir (2019) have also documented the skills in demand for Ukraine, using data from one private portal. Our paper adds to this by offering a rigorous methodology on the aggregation of data across four portals. We also use a highly detailed skills taxonomy so that our findings lead to very clear policy recommendations on skills development and have the potential to inform education curricula. With regard to method, our paper closely relates to Deming and Kahn (2018), who apply text analysis to study heterogeneity in employer skill demand using US job platform data. We follow a similar dictionary-based approach to identify

<sup>&</sup>lt;sup>3</sup> Available data such as labor force surveys are useful to analyze overall variations in employment across sectors and age groups, but these data have limitations in reliably and timely estimating job creation and the characteristics of jobs currently created. Firm registry data can provide estimates of jobs created in the formal sector but omit an important share of the economy that is informal or dominated by small firms, which is especially important for labor markets such as Kosovo's. Measuring skills needs quantitatively and at a granular level is challenging. An attempt at measuring skills needs quantitatively is made in some modules of the Skills Towards Employment and Productivity Surveys ("STEP surveys") directed to employers and implemented in several countries, including in Kosovo in 2015-2016. It identifies the existence of skills constraints across sectors and type of occupations. However, the data on demand for skills are not granular enough for specific feedback for policy makers on skills needs and curricula alignments.

skills in job descriptions. As mentioned earlier, we substantially improve the granularity of the skill taxonomy by adding a layer of language-specific key words, key phrases and text patterns.<sup>4</sup>

Moreover, our paper relates to a broader literature building on job portal data in various contexts. Postings on platforms provide detailed data for general labor market analysis. For example, Azar et al. (2017) use US job vacancies collected by Burning Glass Technologies at the county level to study market concentration. Job platforms allow to study employers and job seekers' behavior. Text analysis of the requirements posted by employers on job ads can inform us on existing gender bias (as shown in China by Kuhn and Shen, 2013). Applications can be sent through platforms to assess the extent of discrimination from employers, in the same spirit as resume audit-studies (Kuhn and Shen, 2015). Using US data from the portal CareerBuilder.com, Marinescu (2017) tracks how job seekers' search behavior is impacted by changes in unemployment benefits duration. The same portal also allows Marinescu and Rathelot (2018) to study the importance of distance-to-home in job search strategies, and the geographical mismatches it can induce.

The remainder of this paper is structured as follows. In Section II, we present the labor market in Kosovo and highlight its challenges. Section III describes the data, followed by the methodology in Section IV. We present our key findings on vacancy characteristics in Section V, and our results on skill needs in Section VI. We conclude in Section VII. An online appendix is available for additional details.<sup>5</sup>

#### II. The Kosovo labor market context

Kosovo is the youngest and one of the poorest countries in Europe in terms of statehood and demographics, and GDP per capita. Despite stable economic growth in the last 10 years, job creation in Kosovo is at a low level. Between 2012 and 2018 about 44,000 new jobs were generated and the employment rate increased by 3.4 percentage points, from 26.6 percent to 29.4 percent (World Bank, 2020). Over the same time, the increase in the employment rate averaged 8 percentage points, from 44.4 percent to 52.5 percent in the Western Balkan region. While commerce and construction have been the main drivers of employment growth over the past decade, these sectors are characterized by low productivity and low wages. Still, employment

<sup>&</sup>lt;sup>4</sup> Our work also relates to a literature using text mining techniques due to the increased availability and scale of job portal data. For instance, a literature around Wowczko (2015) uses clustering methods to identify key skill requirements in job titles and descriptions of online vacancies. Despite their usefulness, the results are fully inductive and thereby often result in high error rates, which are especially problematic when findings are used for policy recommendation. We circumvent these challenges, by following a dictionary-based approach with pre-defined and validated key words, key sentences and dynamic regular expressions.

<sup>&</sup>lt;sup>5</sup> Appendix A is a technical appendix providing the detailed methodology for the analysis of platform data. Appendix B presents in more detail the skills taxonomy and the methodology for the textual analysis on skills. Additional graphs are provided in Appendix C. A comparison of STEP survey data with job platform data is provided in Appendix D. Instructions to replicate the analysis are provided in Appendix E.

growth in Kosovo has not been sufficient to absorb the new labor market entrants. Between 2012 and 2018, the working-age population (15-64 years) increased by 3.4 percent, according to data from annual Labor Force Surveys.<sup>6</sup>

The limited employment creation induces high unemployment rates, especially among youth (although there have been improvements since 2014). In 2017, there were around 157,000 unemployed Kosovars, among which 50 percent were youth aged 15-29. Among the labor force population, approximately 3 out of 10 were unemployed and this rises to 1 out of 2 among youth aged 15-24 (Figure 1). Kosovo has the highest unemployment rate of the region, the Western Balkans average was 16.9 percent in 2017. The lack of opportunities triggers high long-term unemployment rates, reaching 72 percent of the 15- to 64-year-old unemployed in 2017 (Figure 2). This is similar to other Western Balkan countries but in comparison, the average rate in the EU28 was 35.5 percent.<sup>7</sup>





Source: World Bank calculations based on LFS 2012-2017.





The shortage of jobs is especially challenging for youth entering the labor market with no experience. Among the young (15-24 years old), the employment rate was around 10 percent over the period 2012-2018 and unemployment has been consistently higher than 50 percent in recent years (reaching a peak of 61 percent in 2014 and decreasing until 2017). In comparison, the youth unemployment rate in Kosovo has been about 4 times higher than the world's average youth unemployment rate, reaching 53 percent in 2017 for the 15- to 24-year-olds. It also exceeds the average youth unemployment rate in the Western Balkan region and the euro area.

<sup>&</sup>lt;sup>6</sup> SEE Jobs Gateway Database, based on data provided by national statistical offices and Eurostat.

<sup>&</sup>lt;sup>7</sup> Source: Eurostat, long-term unemployment (12 months or more) as a percentage of total unemployment in 2017. Note that the Eurostat indicator is computed over unemployed aged 15-74 and not 15-64.

Kosovo exhibits the highest unemployment rates in general, and youth unemployment rates especially, in the region.

The share of youth not in education, employment or training (NEET) also reflects the challenges they face entering the labor market (Figure 3). The share remained relatively constant between 2012 and 2017, ranging from 30 to 35 percent, which indicates that around 1 out of 3 youth were not investing in their human capital, not working and not looking for a job. Given the lack of jobs for many young people, emigration is increasingly considered a viable option. A recent Gallup survey on migration provides insights into the willingness of youth to leave Kosovo: The Youth Potential Net Migration Index estimates that if all 15- to 29-year-olds who desire to move in and out of Kosovo did so, the youth population would decrease by 48 percent. Among the total population, this rate is 42 percent and the highest in Europe.<sup>8</sup>









Source: World Bank calculations based on LFS 2012-2017.

*Source:* SEE Jobs Gateway Database, based on data provided by national statistical offices and Eurostat. See World Bank (2019b).

On the firm end, employers report significant hiring constraints due to shortages of skilled labor. According to the Skills Towards Employment and Productivity (STEP) Survey conducted in Kosovo (World Bank, 2019a), skill and experience gaps were among the major labor constraints reported by firms respective to potential growth. Firms considered the lack of workers with relevant experience to be the primary labor-related concern,

<sup>&</sup>lt;sup>8</sup> Based on Gallup surveys 2015-2017, Gallup Migration centers report three indices for 152 countries: the Potential Net Migration Index (Total PNMI), Potential Net Brain Gain Index (Brain gain PNMI) and Potential Net Youth Migration Index (Youth PNMI). Kosovo ranks among the lowest (bottom 15) on all those indices: - 42% PNMI, -43% Brain gain and -48% Youth PNMI. A -42% PNMI indicates that if all adults who desire to move in and out of the country did so, the adult population would decrease by 42%. The Potential Net Brain Gain Index is measured on those who have completed four years of education beyond high school or have the equivalent of a bachelor's degree. The Potential Net Youth Migration Index is measured on total population of the country aged 15 to 29.

followed by labor availability and the quality of formal education and training related to the production of relevant skills (World Bank, 2019a). From the firm perspective, skill deficits related mainly to new labor market entrants who were primarily jobless youth (World Bank, 2019a).

Skill constraints affected all hiring firms, across type of occupation and across sector. Depending on the specific position to fill, between 59 percent and 77 percent of hiring firms encountered problems because of applicants' lack of skills or experience (Figure 5). Skill constraints affected all hiring firms but had a stronger effect on large, dynamic and innovative firms.<sup>9</sup> Therefore, the firms that are more likely to be competitive and productive are also the ones facing the largest skill constraints on recruitment.

Among Western Balkan countries, Kosovo stands out when it comes to skill gaps (Figure 5). A majority of firms in Albania and Serbia also experienced skill-related problems in recruiting workers for higher skill occupations, but the share of firms affected in Kosovo is even higher.

The low level of skills is related to the poor quality of the education system. Although enrollment in secondary education is high in Kosovo, a large majority of students have not acquired basic cognitive skills: 70 to 80 percent of students are not proficient in math, science and reading according to PISA tests.<sup>10</sup> In 2014, Kosovar children completed an average of 12.8 years of school by age 18. Once the quality of learning is taken into account, however, this is equivalent to just 7.7 years.<sup>11</sup>

On the employer side, between 40 and 50 percent of firms report that general education does not adequately prepare students for the workplace – either in terms of up-to-date knowledge or soft skills (World Bank, 2019a). Consequently, education credentials are not thought to accurately reflect skills during recruitment.<sup>12</sup>

Our analysis in this paper provides additional insights into the skills required by firms, which are likely to be the bottleneck for recruitment. We use a detailed taxonomy of skills, so we can precisely identify and communicate over the skills in demand, by sector or industry.

Figure 5: Share of hiring firms that experienced difficulties in hiring due to lack of experience or lack of skills, by type of occupation to fill (%).

<sup>&</sup>lt;sup>9</sup> More precisely, firms in the Business Services sector face recruiting constraints due to skills gaps more frequently than in other sectors. This is also true for large firms compared to smaller ones, for firms that invested in R&D and for foreignowned firms compared to other firms. See World Bank (2019a).

 $<sup>^{\</sup>rm 10}\, Results$  from PISA 2015.

<sup>&</sup>lt;sup>11</sup> Learning adjusted expected years of schooling, Human Capital Project data set (World Bank), based on 2014 data.

<sup>&</sup>lt;sup>12</sup> Qualitative evidence from employers and stakeholders' consultation confirmed that. Some hiring firms are setting their own entry tests and others are setting up their own training programs to compensate for the deficiencies in the public system. Others are setting up their own training programs to compensate for the deficiencies in the public system.



Source: Employer STEP survey for Kosovo, Albania and Serbia, World Bank own calculations.

#### III. Data

We examine 5,272 job postings (equivalent to 12,321 vacancies) from the four largest online job portals in Kosovo, namely *KosovaJob*, *Portalpune*, *Ofertapune* and *Telegrafi* for the year 2018. The selection of job portals for our analysis is based on information shared from the data providers, and from our own preliminary research, in which 14 main job portals in Kosovo were identified and assessed with regard to content and cross-postings. Most of these job portals are small and tend to follow and repost content from other portals, particularly the four portals used in this analysis, where most of the information in the market is concentrated. The data set for this project is provided through cooperation between the World Bank and KosovaJob, a major job portal in Kosovo. KosovaJob shared two types of data: The first are data on job postings from the back end of their own job portal. Such data are directly filled by firms entering information through a web interface.<sup>13</sup> The second are job posting data for the three other competitors' platforms. These data are collected daily for market research purpose by scraping the competitors' websites, and they contain information on postings already publicly available.

The job postings in our data set are a useful source of information on the vacancies available in the labor market. From a posting, one can directly collect information on characteristics of the position such as the sector or industry, the type of contract, and the location of the job. Typically, the job description also includes requirements in terms of field of education and/or diploma, minimum years of experience, and particular skills demanded by the employer. Other relevant information includes posting and expiration date and the number

<sup>&</sup>lt;sup>13</sup> Note that firms fill two types of fields, mandatory fields (e.g., job title, type of contract) and optional fields (job city). The type of field will impact the completeness of the database for a given characteristic.

of vacancies associated with a job posting. Job descriptions are a key source of data for skill demand, and the paper mostly focuses on analyzing them.

Job postings data, however, may not be representative of labor demand across entire economies. Recent evidence from the roll-out of LinkedIn data for policy use (e.g., at the World Bank) suggests that online job platform data are skewed towards high-skilled jobs and that the degree of labor market representativeness highly varies across countries (Zhu et al., 2018). Correcting for such bias is not straightforward. Kureková et al. (2015) highlight that due to the absence of information on the precise size and structure of the population of vacancies in the job market of many countries, traditional correction methods (e.g., weighting) are often not feasible to achieve representativeness. In the case of Kosovo, we highlight three aspects with regard to representativeness. First, close to 40 percent of the employed population is estimated to work informally (i.e., without a labor contract).<sup>14</sup> Assessing the extent to which this leads to a bias in representativeness is difficult, as traditional data sets at the firm level (such as firm registries) do not reflect jobs created in the informal sector, therefore compromising comparisons with job portals data. Second, and similar to other middle-income countries, even within the formal sector not all vacancies appear on job portals, in part because the majority of jobs tend to be filled through informal channels (Cojocaru, 2017; World Bank, 2019a). Employers report that informal channels are their main recruitment channel, with 64 percent using such channels to recruit for high-skilled positions, and 59 percent for medium- to low-skilled positions (World Bank, 2019a). Finally, because online job advertisement is carried out by both private and public providers, even among online vacancies the true number and characteristics of posted jobs are not known. Kureková et al. (2015) stress that vacancies tend to occur in the presence of an unfulfilled demand or where employers have preferences for a selection process. Therefore, online vacancies are useful to learn about jobs for which employers face difficulties to fill through internal or informal search channels.<sup>15</sup> To examine which segments of the labor market are observed on job portals, we contrast findings from our vacancy analysis with other evidence from the Kosovar labor market using labor force survey data in Section V.16

Despite their limitations, job portal data are a complement to other available data, especially to understand which profiles are in demand. For instance, Labor Force Surveys (LFS) are useful to analyze overall variations in employment across sectors and age groups, but these data have limitations in reliably and timely estimating job creation and the characteristics of jobs currently created. More importantly, LFS data reflect the stock of available skills in the labor force, while job postings directly express employer needs and the skills in demand.

<sup>&</sup>lt;sup>14</sup> World Bank calculation based on LFS 2017. Informality is defined here as workers without a labor/employment contract (legal definition in Kosovo).

<sup>&</sup>lt;sup>15</sup> Anecdotal evidence from counterparty discussions suggests that this is increasingly true in Pristina, the capital city of Kosovo.

<sup>&</sup>lt;sup>16</sup> The 2017 LFS is used to compare the characteristics of jobs on the portals with the characteristics of employment in the economy. STEP Employer survey data (World Bank, 2019a) are used to contrast our findings on skills.

Firm registry data can provide estimates of jobs created in the formal sector but lack general information on job characteristics, and even more on workers' profiles. An attempt at measuring skill needs is made in some modules of the Skills Towards Employment and Productivity Surveys ("STEP surveys") directed to employers and implemented in several countries, including in Kosovo in 2015-2016. It identifies the existence of skills constraints across sectors and types of occupations. However, the data on demand for skills are not granular enough for specific feedback to policy makers on skills needs and curricula alignments. Detailed information on both technical and soft skills demanded by employers for specific types of occupations is not available, thereby limiting its usefulness for students and parents seeking to take career decisions. We show in Appendix D the differences in granularity level between job portal data and STEP data, to further highlight how job portal analysis can substantially contribute to the policy dialogue on skills development.

This paper is thus an example of how the analysis of job descriptions from online vacancies can generate useful and detailed complementary data on current demand for skills in a labor market. Our paper contributes to existing work on skills by (i) using data that directly reflect employers' skills demand rather than the skills stock, and (ii) presenting demand for skills at the broad and granular levels.

### IV. Methodology

First, we define a *job posting* as a unique job advertisement on a given job portal during a specific time interval in the year 2018.<sup>17</sup> This means that in the presence of duplicated postings at the same time across job portals, only a single job posting is taken into account.<sup>18</sup> Furthermore, a *vacancy* is defined as a unique work position that is associated with a job posting. By definition, a job posting can therefore be posted as advertisement for several vacancies. For instance, when a firm is searching for 10 workers for a construction site by advertising only one post, we consider the number of vacancies to be 10 and the number of job postings to be 1. For the analysis of job creation, we are interested in the number of *vacancies*. However, because the numbers of postings and vacancies typically coincide (as shown below), we take a conservative approach by conducting our analysis at the posting level. By doing so, we also address the data provider's concerns that some of the postings advertising several vacancies may overstate the true number of vacancies. Note that we present robustness checks in Appendix C showing that key results are valid regardless of the approach.

While the comprehensive methodology is described in detail in Appendix A, we highlight here two main methodological aspects: (a) the deduplication of job postings across portals, and (b) the text analysis of job descriptions.

<sup>&</sup>lt;sup>17</sup> The time dimension of the postings is defined according to exact start and end dates of each posting.

<sup>&</sup>lt;sup>18</sup> This task is performed using a deduplication algorithm.

One contribution of our approach is the deduplication of the job postings across portals to ensure that only unique postings are considered. Overall, the deduplication leads us to consider only 12,321 vacancies in the final data set rather than the initial 16,888 vacancies, which is around 30 percent less.<sup>19</sup> Our deduplication algorithm consists in iteratively deleting double entries based on different time horizons up to one month of publication. Thereby, it takes into account two different aspects of job postings. On the one hand, it accounts for the possibility of lagged duplications when a posting is duplicated on a different job portal within a one-month frame. Such postings are considered duplicates. On the other hand, the algorithm acknowledges that when firms repost an identical job ad at a later time, this may indeed refer to a different work opportunity. Therefore, if the repost is done after one month (whatever the portal it is posted on), this is *not* considered a duplicate. A detailed overview of this procedure is provided in Appendix A.III. By deduplicating the raw data, we extend previous work on online vacancies in Kosovo, most notably that of Helvetas (2016), who reports raw counts of vacancies from 2015.

The text data used for the analysis, namely the job title and the job description, contain information on skills, minimum years of experience, and minimum education requirements in a raw text format (also called "unstructured" data). That is, the information of interest is specified as raw text in the respective columns from where information is first extracted prior to analysis. A textual analysis is conducted on the job description and job title to identify the incidence of skills, as well as education and experience requirements.

The textual analysis follows that of Deming and Kahn (2018) and relies on a dictionary-based approach. The authors study variation in skill demand for professionals across firms and labor markets.<sup>20</sup> They categorize a wide range of keywords found in US job ads into 10 general skill categories and thereby create a skill dictionary (see Appendix B.I). Based on this, we create our own dictionary, which comprises 15 skill categories. For each skill category, there is a set of skill requirements (or sub-categories) which can be identified by keywords or key phrases. In total, the skill categories rely on 152 skills requirements, which are defined by 764 keywords across three languages. We draw on their dictionary of keywords and extend it in various ways documented in Appendix B.II. Finally, the dictionary of 15 skill categories is also mapped to a more aggregate framework

<sup>&</sup>lt;sup>19</sup> Note that this takes into account "trimming" the number of vacancies per posting at the 99th percentile to correct for artificial increase of total of vacancies due to outliers among job postings. In addition, a rough deduplication was performed by KosovaJob in parallel with data collection. Thus, to assess the extent of our deduplication rate, we make use of our own, daily web scraped data for three of the four portals for the overlapping period, i.e. November and December 2018. Comparing initial counts from the scraped data with the deduplicated figures from the data used in this paper suggests a deduplication rate of 20 percent vis-à-vis the raw counts. Thus, at least for November and December 2018, our deduplication rate is consistent with rates estimated, for instance, in Boselli et al. (2018).

<sup>&</sup>lt;sup>20</sup> Focusing particularly on cognitive and socio-emotional skills, the authors find positive correlations between each skill and external measures of pay and firm performance.

highlighting three main types of skills (socio-emotional, cognitive, and technical) to match the current skills literature. Our skill taxonomy with the three different layers is presented in Section VI, Table 1.

To conduct our analysis, we construct a search engine to probe for the predefined set of keywords, key phrases or text patterns that identify our extended skill categories, among which are socio-emotional and cognitive skills.<sup>21</sup> The search engine asserts the occurrence of the underlying keywords and returns the set of skills identified in the job description. We interpret the occurrence of a skill in any job posting as demand for it.<sup>22</sup> Appendix B.III provides more technical details on the search engine.

Finally, three methodological caveats can be highlighted. First, the results are sensitive to the specification of the skills dictionary, which is a combination of predefined and inductively selected keywords. This is because any occurrence of a specified keyword, even if the keyword is just part of another word, is interpreted as demand for the corresponding skill. To address this issue, we conduct a validation test based on random sampling of job postings and calibrate our skills dictionary iteratively to delete keywords that lead to confounding results. Moreover, we make use of regular expressions to account for the context of key words and thereby to enhance precision. We report details on this validation test in Appendix B.<sup>23</sup>

A second caveat stems from the lack of a distinct industry and occupation taxonomy in the data. More precisely, the data used for this analysis are not classified according to any standard taxonomy such as ISCO.<sup>24</sup> Therefore, the classification matches the one used by the data provider, which does not properly distinguish between sectors and occupations. For example, next to sectoral categories such as "Banking" or "IT", we find occupational categories such as "Sales" and "Management". Thus, any sectoral breakdowns reported below are skewed by the fact that firms may arbitrarily report a job under the respective sector or under the respective occupation. Similarly, the provided taxonomy of job type (full time vs. part time) is not separated from the

<sup>&</sup>lt;sup>21</sup> Our dictionary extends the work of Deming and Kahn (2018) through an inductive search of additional relevant keywords, key phrases and text patterns (search expressions) supported by constructing term frequencies and iteratively validating them. Furthermore, while their dictionary consists of two layers (i.e., the described skill categories and the underlying keywords), we incorporate a fourth layer in our analysis. As described in detail below, the fourth layer introduces more flexible search expressions in various languages to improve the accuracy of the search algorithm. These search expressions are the actual inputs of the search algorithm. Hence, we refer to the second layer as *sub-categories or skill requirements* in our analysis below, while the input keywords are referred to as *keys* or *patterns*. Detailed explanation on these and other modifications are provided in Appendix B. Finally, we note that Deming and Kahn use preprocessed skills data while our analysis relies on raw, unstructured text data.

<sup>&</sup>lt;sup>22</sup> As in Deming and Kahn (2018).

<sup>&</sup>lt;sup>23</sup> Similar to prior research (e.g., Humphreys and Wang, 2018), a key objective of the validation procedure is to quantitatively reduce rates of false positives obtained through the automated text search (in this case 10 percent). Consequently, estimated incidence rates appear very low overall. To alleviate the potential downward bias from this conservative approach, we repeat our analysis with a less conservative dictionary and find that although incidence rates increase significantly, key results of our analysis are maintained.

<sup>&</sup>lt;sup>24</sup> Building a new ISCO-type taxonomy and migrating existing job postings to this one is feasible but will require further work, as it requires to analyze the description of the job content more in depth.

taxonomy of the type of contract (permanent vs. full time vs. internship), therefore not allowing us to disentangle these different dimensions.

Finally, the analysis rests on a few assumptions regarding the informativeness of the data. Most importantly, we highlight that by analyzing skills via mined keywords, it is assumed that information on skills in job descriptions is somewhat perfectly specified and hence extractable. However, in the case of imperfect specification along any dimension (e.g., along different sectors or different skill sets), the analysis is biased towards skills that are better specified than others. For example, while the socio-emotional skill requirements of a manager may be specified explicitly and in detail (e.g., "leadership" or "communication skills"), a posting for a worker on a construction site may not explicitly specify the physical skills required to successfully perform the job.

# V. What type of job opportunities are available on job platforms?

In this section, we present our key findings on the characteristics of the 5,272 postings in 2018 job portal data, which is equivalent to 12,321 vacancies.<sup>25</sup>

Most vacancies posted online are for "white collar" jobs. As such, the majority of vacancies posted online are for Sales (14.38 percent), Management (7.95 percent), and Administration (7.55 percent), followed by the public sector<sup>26</sup> (6.75 percent) and IT (6.61 percent) as shown in Figure 6. Across job portals, some specialization is visible. For example, while the incidence of vacancies in Administration, Banking & Insurance, and Management is relatively high on *KosovaJob*, manual jobs are posted relatively more frequently on *Ofertapune*. Compared to global employment trends in the labor market, the manufacturing and construction sectors are likely underrepresented in job portal ads (Figure 7). Given that informal channels prevail in hiring, this signals that only a portion of the job opportunities appear in a transparent manner on portals.

In most cases (78 percent), one posting refers to one vacancy. However, as illustrated in Figure 30 Appendix C, some firms bundle several vacancies within a single posting.<sup>27</sup> This form of collective posting varies across industries. Postings with several vacancies occur more often in specific industries like Education or

<sup>&</sup>lt;sup>25</sup> Recall that results are presented at the posting level. Appendix C presents a set of graphs comparing the two approaches and suggesting that results are similar regardless of the approach. Summary statistics of key variables are provided in Appendix C in Table 5 and Table 6.

<sup>&</sup>lt;sup>26</sup> Note that most of the jobs categorized as "public sector" are not civil servant positions but contractual positions posted by municipalities or ministries. Such jobs are reposted freely on some of the platforms. Their frequent appearance on portals is related to the nature of recruitment in the public sector: jobs must be publicly advertised for a competitive recruitment.

<sup>&</sup>lt;sup>27</sup> Due to plausibility concerns, we truncate the distribution at the 99th percentile and report the full variability of data in Appendix C. Upon cleaning, the distribution still has a pronounced upper tail caused by vacancies relating up to 41 jobs. Summary statistics are provided in Table 5.

Manufacturing. In comparison, in Consulting, Finance and Accounting or Art and Design (among others), one vacancy per posting is the norm (Figure 32, Appendix C).







Source: World Bank calculations based on job portal data.



On average, most job postings are published online in Albanian (78.6 percent), but the share of postings in English is high when compared with other languages (19.9 percent), which is indicative of the types of jobs available on portals (Figure 33, Appendix C).<sup>28</sup> The remaining vacancies are posted in German, Serbian and Italian. Overall, the relatively high share of English postings highlights the importance of English fluency as a key skill requirement for the represented segment of the labor market in Kosovo, namely high-skilled individuals.<sup>29</sup>

Over the course of one year, around 70 percent of firms posted up to two online job offers. Only few firms post several times a year on the platform.<sup>30</sup> More precisely, 5 percent of firms post more than 10 times (Figure

<sup>&</sup>lt;sup>28</sup> To conduct this analysis, we classify the text language of a vacancy according to a non-deterministic Naïve Bayes classifier implemented in the Python Package *langdetect*. Whenever available, the vacancy description serves as basis for the classification, else the title is used. A detailed overview is provided in appendix A.

<sup>&</sup>lt;sup>29</sup> Language distributions are strongly heterogeneous across industries. Figure 32 in Appendix C shows that vacancies pertaining to jobs in Gastronomy & Hoteling, Language, Manufacturing and Public Sector are more than 95 percent of the cases in Albanian. In contrast, vacancies in Consulting are posted in English language in almost 70 percent of the cases.

<sup>&</sup>lt;sup>30</sup> To conduct the firm analysis, we correct the data for string misspecifications and differential naming of a firm. Here we distinguish between apparent conflicting firm labels and potential difference due to the firm structure. We define a unique firm as a company title referring to a single entity of a firm structure. This implies, for example, that we distinguish between

31 Appendix C). For these firms, we are interested in knowing if they repeatedly post for the same position or not. Such behavior could express, for instance, repeated needs (e.g. growth of a sales department) or the consequences of hire-and-fire policies and high replacement rates. Our analysis, however, suggests that the number of firms repeatedly posting the same ads is small.<sup>31</sup>

The large majority of jobs posted online are for full-time positions (98.8 percent); part-time jobs are rare in Kosovo (Figure 34, Appendix C) and only a handful of postings are for internships. Of the 5,272 job descriptions, only 23 refer to an internship position. According to job platforms' management, internships are not perceived as equivalent work opportunities by firms, which would rather advertise through other (free) channels such as social media or private networks. The absence of internships on main job portals is of concern for youth searching for experience in a labor market dominated by informal networks.

Most job advertisements do not specify whether a position is temporary or permanent, but they are almost exclusively temporary given the prevalence of such contracts in the labor market. Less than 0.5 percent of job descriptions indicate that a position is temporary, likely because a minority of jobs are permanent in Kosovo. This is consistent with evidence from survey data showing that over 70 percent of the employed population in Kosovo are in a temporary contract in 2017 (Figure 35, Appendix C).

Surprisingly, job portals exhibit low cyclicality of job postings over the year (Figure 37). Minor dips in the posting activity occur in January and July only.

Regarding the dynamics in the market, most vacancies in 2018 remained online for two weeks (Figure 36). Almost a quarter of vacancies expired within the first 10 days (26.91) and only around one-fifth of vacancies (19.44 percent) were posted for a period between 25 and 40 days. Across industries, vacancies in consulting stayed online almost half as long (8 days) as jobs in manual occupations, for example (14 days) when comparing median values. These figures reflect the fact that those job portals recommend firms to leave postings for two weeks by default. While exceeding this period does not typically incur additional costs for the posting firms, firms tend to deactivate a posting after receiving too many applications.

a holding company and domestic and international subsidiaries of it. A detailed overview of this procedure is provided in Appendix A. Overall, we identify around 1,600 unique firms, of which 820 post only once.

<sup>&</sup>lt;sup>31</sup> To assess the number of repostings by the same firm, we investigate the number of *unique* repostings by a firm, i.e. the number of postings when considering a narrower definition of a posting than previously stated. For this purpose, in Table 8, we present summary statistics of unique postings according to four different definitions of uniqueness, next to only deduplicated figures. The figures show that restricting the definition of a posting to, for instance, the job description only reduces the average number of posting by firm from 3.8 to 3.1, while median figures drop from two to one posting per firm. A full frequency table for the distribution of postings by firm is presented in Table 7.

Vacancies published on portals are highly concentrated in the city of Pristina, which means that economic opportunities in other regions are not made publicly available on portals.<sup>32</sup> In Figure 38, Appendix C, we show a ranking of the 10 most frequent cities while accounting for the fact that a vacancy can indicate several potential cities for the job related to it (e.g., when firms have multiple sites in the country). If several cities are mentioned in a posting, Pristina is included most of the time. Other frequent job cities are: Prizren (197), Pejë (184), Fushë Kosovë (168), and Ferizaj (154). Postings on portals do not reflect well the economic dynamics in other regions (comparing with regional employment growth rates, see Figure 39, Appendix C). For example, the incidence of jobs in Pejë and Gjakovë is low despite strong economic activity due to the large retailers maintaining production sites there. One explanation is the type of clients for job portals, which are mostly medium to large firms highly concentrated in Pristina.<sup>33</sup>

#### VI. What are the skills needed by employers?

In this section we present the main findings on skill demand based on job vacancy data in Kosovo. The purpose of this analysis is to explore which skills are sought after by employers who are hiring and to identify specific industry needs, at a granular level. Beyond skills, we are interested in firms' requirements regarding education levels and previous work experience of potential employees. We use a skills taxonomy consisting of three layers, to provide as much detail as possible on the skills demanded by employers (Table 1). The general taxonomy presents three broad types of skills, namely cognitive, socio-emotional, and technical skills.

It is noteworthy that the literature on skills formation and investment in human capital interventions mainly highlights two types of skills: cognitive and non-cognitive skills (also sometimes referred as soft skills, personality traits or socio-emotional skills).<sup>34</sup> However, it is a consensus in the literature that both cognitive and non-cognitive abilities explain socio-economic (including labor) outcomes (Heckman et al., 2006 and Kautz et al., 2015), that there is a multiplicity of skills whether it is among cognitive abilities (Heckman and Kautz, 2012 and Handel et al., 2016) or non-cognitive skills (Almlund et al., 2011 and Sanchez Puerta et al., 2016), and

<sup>&</sup>lt;sup>32</sup> To conduct this analysis, we assume that for the portals where no job city but only a headquarters city was given (all portals except KosovaJob), the headquarters city is equal to the job city. We run several plausibility checks to motivate this assumption and argue that it is justified given that economic activity is very concentrated to the capital in Kosovo. <sup>33</sup> According to local HR experts, most of the jobs in secondary cities tend to be low- to medium-skilled jobs and therefore recruiting relies mostly on informal channels. This is facilitated partly due to pronounced social ties that arise in these smaller cities, while strong growth of Pristina over the last years has made informal recruiting more complex and therefore increased switching to online recruiting. Another potential explanation is that, as most headquarters are in Pristina, the figures reflect a high concentration of recruitment processes in the capital.

<sup>&</sup>lt;sup>34</sup> See Heckman et al. (2006), Cunha et al. (2006), and Kautz et al. (2015) among others.

most importantly, that skills form at several stages from early childhood to adulthood benefitting from a multiplier effect if early investments have been made (Cunha et al., 2006).<sup>35</sup>

Given that we are interested in the skills of the labor force according to employers' demand, we isolate "technical skills" by adding a third type of skills next to cognitive and non-cognitive skills (which we refer to as socio-emotional skills) (Table 1, layer 1).<sup>36</sup> Our technical skills category corresponds to skills built on a combination of cognitive and socio-emotional skills, taught or learned at least partly for labor market use, and developed later in life while at school (general or technical education), in dedicated trainings or on-the-job. By doing that, we also follow recent reports focusing on the use of skills at work, for example Pierre et al. (2014) which sets the taxonomy to analyze STEP surveys (in particular the Employer survey expressing employers' skills requirements), and the World Bank Development Report on Learning (World Bank, 2018).

In addition, we disaggregate this framework into 15 skills categories (cognitive skills, 5 socio-emotional skills - agreeableness, conscientiousness, emotional stability, extraversion, openness to experience, and 9 technical skills) building on Deming and Kahn (2018) to perform a more detailed analysis on the demand for specific technical skills (Table 1, layer 2).<sup>37</sup> For socio-emotional skills categories, we mapped our granular skills to personality traits also referred to as "Big Five" in the psychometric literature (Goldberg 1993), based on Chernyshenko et al. (2018) and Kureková et al. (2016).

Finally, each of the 15 categories comprises a relatively large set of skills requirements (Table 1, layer 3).<sup>38</sup> This layer offers the highest level of granularity in our analysis of skills in demand and is a key contribution of this paper. For example, it enables us to construct word clouds to inform youth on the skills employers look for, and to inform curricula on the specific skills to be taught (e.g. which software or coding language, in the computer science field). It relies on 152 skills requirements, themselves defined by 764 underlying keys across three languages (Albanian, English and German). Details on the taxonomy and text analysis are provided in Appendix B.<sup>39</sup>

<sup>&</sup>lt;sup>35</sup> More precisely, the "complementarity" of skills including across cognitive / non-cognitive categories (i.e. early skills acquisition raises the productivity of investment in skills later) and their "self-productivity" (i.e. skills are self-reinforcing) are two key aspects of skills formation explaining the multiplier effect and justifying the investments in early childhood development programs (Cunha et al., 2006).

<sup>&</sup>lt;sup>36</sup> In Handel et al. (2016), such skills are considered as cognitive skills but distinguished from the cognitive foundation skills (such as writing, reading, problem solving). They are labeled "specific mid-level" and "specific high-level" knowledge/skills, and Table 1.1 in their paper shows the path of skills acquisition for these job-related skills. In the World Bank (2018) report on learning, the terminology "technical skills" is used.

<sup>&</sup>lt;sup>37</sup> Note that due to their rare occurrence certifications are excluded from the analysis below.

<sup>&</sup>lt;sup>38</sup> Recall that in order to identify whether a skill requirement is in demand within a job description, we use the occurrence of underlying keywords, key phrases or text patterns from the fourth layer of our skills taxonomy. See Appendix B for more details.

<sup>&</sup>lt;sup>39</sup> Note that for the "Big Five" socio-emotional skills we map granular skills requirements based on Chernyshenko et al. (2018) and Kureková et al. (2016).

Layer 1	Layer 2	Layer 3
Broad type of skills	Skills categories	Skills requirement or Skills sub-categories
Cognitive	Cognitive	Ability to produce solutions quickly, Analytical Thinking, Critical Thinking, Cultivated, Math, Multitasking, Presentation skills, Problem solving, Research, Statistics, Strategic Thinking, writing
Socio-Emotional	Agreeableness	Empathy, Friendliness, Politeness, Positive attitude, Teamwork
Socio-Emotional	Conscientiousness	Attentive, Detail-oriented, Devotion, Discipline, Meeting Deadlines, Motivation, Organized, Process-oriented, Quality orientation, Reliability, Responsibility, Solution- oriented, Time Management
Socio-Emotional	Emotional Stability	Confidence, Patience, Stress Management
Socio-Emotional	Extraversion	Collaboration, Communication, Eagerness, Energetic, Expressiveness, Independent, Interpersonal skills, Leadership, Mentoring, Negotiation, Networking skills, Persuasion, Proactiveness, Show initiative
Socio-Emotional	Openness to experience	Creativity, Curiosity, Flexibility, Open-mindedness
Technical	Certifications	CFA Certification (Chartered Financial Analyst), CISCO Certification, Driver's License, ISO 13485 (Medical Devices Quality Management Systems), ISO 14001 (Environmental Management), ISO 22000 (Food Safety Management), ISO 27001 (Information Security Management), ISO 9001 (Quality Management), Microsoft Certification
Technical	Computer	Computer, Data processing, Database skills, Excel, Linux, MS Office, Outlook, PowerPoint, Spreadsheets, Typing, Word, networks knowledge
Technical	Customer Service	Client-centric, Customer-oriented, Sales, Service-oriented
Technical	Financial	Budgeting, Financial Reporting, auditing, bookkeeping
Technical	Foreign Language	Arabic, Chinese, English, French, German, Italian, Japanese, Portuguese, Russian, Serbian, Spanish, language skills
Technical	Manual	Non-smoker, Physically fit, technical operator
Technical	People Management	Management, Supervisory skills
Technical	Project Management	Project Management, project implementation 3D Max, 3DS Max, AJAX, Angular JS, Apache, Artificial Intelligence, Assyst, Autocad, C#, C+, C++, CAD Programs, CADCAM, CNC Lathe, CNC Sliding, CNC machinery, CSS, Capture One, Cordova, Corel Draw, Geometric Tolerance, Google Adwords, Graphic design programs, HTML, HyperV, Illustrator, InDesign, Instagram, Ionic, Java, Javascript, Jquery, Kaizen, LUCA, LV3, Laravel, Matlab, NEBIM, NEBIM V3, Object Oriented Programming, Odeon, Oracle, PHP, Photoshop, Premier PRO, Python, Rubi, SPSS, SQL, Simulink, Social Media , Stata,
Technical	Specific Software	Twitter
3 broad type of skills	1 otal : 15 skills categories	1 otal : 152 skills requirements (and 764 underlying keys)

Table 1	: Skills	taxonomy	for	textual	analysis.

The analysis is structured as follows. We start by presenting what is the demand for broad types of skills (the first layer of our taxonomy, e.g. technical skills), as well as for more detailed skills categories (e.g., "computer skills" within the technical skills) in subsection (i). In subsection (ii) we explore skill needs by industries at a more granular level, using also the third layer of our taxonomy (requirements, e.g. "Spreadsheets" within the "computer skills" category) to provide guidance on skill profiles for specific occupations. In subsection (ii) we complement our findings on skills with evidence on minimum experience and education requirements, which are usually part of the job descriptions.

#### i. Which skills matter? Exploiting word clouds on aggregate skill demand from job descriptions

Technical skills have the highest incidence of all skill categories across all job postings. More precisely, as shown in Figure 9 among technical skill language skills (mostly English) are the most demanded skills (42.12%), followed by computer skills (33.18%), underlining the importance of these requirements against the light of technological and economic developments of recent years. The high prevalence of English requirements highlights once more the relevance of English language as a key skill requirement for the represented segment of the labor market in Kosovo, namely high-skilled people.

Socio-emotional skills have a high prevalence across all industries. More precisely, across all portals and industries 50 percent of postings report demand for socio-emotional skills (Figure 8). Particularly, extraversion (40.27%, Figure 9) and agreeableness (10.14%, Figure 9) are most demanded. Like in many other labor markets, this finding illustrates that socio-emotional skills are transversal. In the case of job portals, it may stem from the fact that socio-emotional skills are simply specified more explicitly in job descriptions than other types of skills. However, this finding is consistent with findings based on data from other sources (such as STEP data) showing that socio-emotional skills are equally needed in lower, medium and high-skilled occupations (World Bank, 2019a).

Apart from technical and socio-emotional skills, there is substantial heterogeneity in terms of required skills across industries and occupations when looking at a higher disaggregation level. Figure 10 highlights that demand for cognitive skills is relatively high for jobs in IT, Science and Research, Law, and the Public Sector (for the universe of sectors, see Appendix C, Figure 42). In contrast, these industries show little demand for language skills.<sup>40</sup>

<sup>&</sup>lt;sup>40</sup> In appendix C (Table 9) we also explore pairwise skills correlations across industries and occupations, however finding no particular pattern.





Figure 8: Incidence of skills across postings by skill aggregate.

Source: World Bank calculations based on job portal data.

Source: World Bank calculations based on job portal data.







We compare in more details our findings to those obtained in the Employer STEP survey for Kosovo, which is the closest comparable source of data as it expresses employers' views on skills required for certain types of occupations. We find that socio-emotional skills (especially conscientiousness) and cognitive skills (problem solving and numeracy) are on the top of the rankings of required skills for new recruits, while technical skills rank lower (Figure 11). However, the comparison cannot be pursued at a more detailed level given the differences in measurement and taxonomy. The contrast of STEP data with job platform data is deepened in Appendix D, where we highlight the advantages of using platform data when conducting a skills demand analysis.



Figure 11: Ranking of skills important for employers when hiring or taking retention decisions for higher skill occupations

*Source:* Employer STEP survey Kosovo (World Bank, 2019a). *Note:* The ranking is based on an index (0 to 300) to take into account the ranking provided by employers on three most important skills. \* Official language refers here to Serbian only (proficiency in Albanian, the other official language, is not part of the questions to employers). \*\* Foreign language excludes English which is measured separately.

To analyze skill demand more specifically, we construct word frequency clouds that allow us to analyze demand for a specific skill sub-category.<sup>41</sup> For this, we search for key-words, key phrases or regular expressions in job descriptions which are related to specific sub-skills within each broad skill category previously shown (see section IV and Appendix B.III for further methodological details). In Figure 12 to Figure 15, each word cloud reflects the relative importance of a skill requirement within a broad skill category. Underlying word frequencies are provided in Appendix C.

Among extraversion skills, communication skills are the most important. Employers also stress the need for interpersonal skills and the need to work independently. Cognitive skills are associated most frequently with demand for analytical thinking, followed by presentation and problem-solving skills (Figure 13). Demand for general computer skills comprises frequently basic understanding of a computer and operating systems, including the application of standard software, such as MS Office (Figure 14). Besides, Figure 15 shows that among more specialized software knowledge, programming skills in Java, Javascript, SQL, HTML, and CSS are frequently demanded.

<sup>&</sup>lt;sup>41</sup> Recall that this corresponds to the third layer of our taxonomy, see Table 1.

Figure 12: Relative importance of extraversion skills



Figure 13: Relative importance of different cognitive skills

writing problem solving analytical thinking statistics presentation skills strategic thinking research critical thinking Figure 15: Relative importance of different software skills

oracle photoshop jquery social media python (SSS html autocad object oriented programming ocial media phop<sup>ajax</sup> java illustrator witter instagram

Source: World Bank calculations based on job portal data.

The word clouds shown shed light on the underlying skill requirements for each skill category and thereby stress particular skill needs while, however, staying on *the aggregate level* of the labor market. Thus, they do not provide any insights on specific needs faced by industries. In the next subsection we use the same graphical tool to illustrate the relative importance of skills at the industry level. By that, we construct skills profiles for different sectors or occupations that in a next step could be used as evidence to inform curricula and skill development for the youth. Again, underlying word frequencies are provided in Appendix C.

ii. Which skills do I need for this job? Exploiting word clouds on key industries

In this subsection, we explore what skills are needed at the industry or occupation level. Figure 16 to Figure 20 are word clouds showing skills profiles for a few industries. Figure 16 shows that skill requirements for a typical administration job comprise most importantly the ability to communicate effectively, as well as general computer skills. In comparison, IT jobs require more knowledge of special software, e.g. Java and more pronounced analytical thinking (Figure 17), while jobs in management require staff management and communications skills (Figure 18).

Figure 16: Skills demand for jobs in Administration

Figure 17: Skills demand for jobs in IT



Figure 18: Skills demand for jobs in Management



Figure 19: Skills demand for jobs in the Public

Sector<sup>42</sup>

DOW

serbian

negotiation

teamwor

management interpersonal skills

leadersh

bo

ottice



Figure 20: Skills demand for jobs in Sales



Source: World Bank calculations based on job portal data.

These examples show that text analysis can be used to provide guidance and information to youth willing to pursue a career in a specific sector. For sectors with important technical skills requirements such as IT, this can also be used to better identify what are the current needs of the sector and align curricula. In the case of IT, it means identifying the software on which students should be trained.

<sup>&</sup>lt;sup>42</sup> Recall that most of the jobs categorized as "public sector" are not civil servant position but contractual positions posted by municipalities or ministries.

#### iii. Experience and education requirements

On average, firms specified a minimum requirement of 2.5 years of experience. However, overall, only around 58 percent of vacancies actually specify an experience requirement at all. There is large heterogeneity across industries: Table 2 shows that management positions are associated with higher experience requirements ( $\geq$ 3 years) while sales positions tend to be associated with lower experience requirements ( $\geq$ 1 year). These findings suggest that higher experience requirements are specified in cases where job profiles may be more complex, e.g. for leadership positions.

Industry	Mean	SD	Min.	Max.	Median
Administration	2.94	1.75	1	8	2
Architecture & Construction	2.84	1.56	1	8	3
Art & Design	2.22	1.23	1	5	2
Banking & Insurance	2.11	1.20	1	5	2
Consulting	4.86	2.30	1	8	5
Economics	2.96	1.49	1	5	3
Education	2.81	1.46	1	8	3
Finance & Accounting	2.80	1.59	1	8	2
Gastronomy & Hotelling	2.11	1.19	1	5	2
Health & Biotech	1.77	0.93	1	5	2
IT	2.37	1.28	1	8	2
Language	2.59	1.09	1	5	2
Law	3.04	1.62	1	8	3
Logistics & Transport	2.37	1.45	1	7	2
Management	3.17	1.59	1	8	3
Manual occupations	2.00	1.19	1	8	2
Manufacturing	2.43	1.18	1	5	2
Marketing, PR & Media	2.34	1.51	1	8	2
Others	2.67	1.57	1	8	2
Public sector	3.14	1.58	1	8	3
Sales	1.62	0.94	1	6	1
Science and Research	2.20	1.17	1	4	2
Telecommunication	2.21	1.28	1	5	2
Unclassified	2.78	1.03	1	5	3

Table 2: Summary statistics of minimum years of experience

Source: World Bank calculations based on job portal data.

We explore this aspect more in depth below. The extent of under-specification differs across industries as shown in Figure 40 in Appendix C, while jobs in Finance & Accounting, Law and the Public Sector specify experience requirements in more than 50 percent of the cases, hardly any requirements are found among jobs posted in Gastronomy and Hoteling or Telecommunication. For management jobs the specification rate is roughly 50 percent. This general lack of specification is induced by platform policies which do not classify the field on "minimum years of experience" as a mandatory one. Therefore, when a firm enters a minimum requirement it reflects the firm's active decision, signaling the relative importance of experience for that job. To investigate the relationship between skills profiles and the specified experience requirement in job postings more in depth, we construct a skills incidence index for minimum years of experience. The index serves as a measure of skills intensity for postings that mention a particular a requirement for years of experience.<sup>43</sup>

The skill incidence index illustrates that jobs which are associated with a higher minimum experience requirement, on average, tend to have a higher incidence of skills requirements in their job descriptions. For example, jobs with a minimum requirement of 4 years of work experience, on average, tend to have a skills incidence which is 32% higher in comparison with jobs that specify a minimum experience level of only one year. Adding to our previous findings that experience requirements are usually specified where job profiles may be more complex, these results confirm that job postings of jobs with higher experience requirements express skills needs more often and more explicitly.





Note: The index is constructed as the arithmetic average of skill incidences using disaggregated skill categories across postings that mention a particular a requirement for years of experience.

Source: World Bank calculations based on job portal data.

Education requirements are specified more frequently than experience requirements and most jobs posted online require tertiary education. Two-thirds of all postings specify an education requirement compared with

<sup>&</sup>lt;sup>43</sup> The index is constructed as the arithmetic average of skills incidences across all skill categories for a given year of experience requirement. The base value is one year of experience. Underlying data are provided in Table 11 of appendix C.

58 percent of postings specifying an experience requirement. Among the postings listing an education requirement, close to 80 percent require a university degree (bachelor or higher) while only one-fifth of postings specify other forms of education requirements, most notably a high school degree or special certifications (Figure 22). Among the jobs requiring tertiary education, 10 percent of postings specify a master's degree or PhD, while usually a bachelor's or equivalent university degree is sufficient.<sup>44</sup> Overall, these figures highlight that firms in Kosovo barely use job platforms to recruit for jobs requiring medium or low levels of education. These findings confirm previous research pointing to the skewedness of job platform data towards high-skill occupations (e.g., Zhu et al., 2018).



Figure 22: Distribution of education requirements across postings

Given the frequent occurrence of postings with tertiary education requirements, we exploit the granularity of our data to analyze the incidence of different education fields across industries. For instance, Figure 23 shows the relative importance of different fields of education in administration jobs, highlighting that Economics and Business Administration are the most demanded fields of studies. Naturally, jobs in IT are associated with more technical fields of studies, such as Computer Science, Graphic Design or Engineering (Figure 24). The most diverse set of demanded education fields is encountered for public sector jobs (Figure 26).

The findings on education fields favored by the demand side can be contrasted with labor force survey (LFS) data showing education fields among the labor force. A frequent demand for a specific field of education (based on our analysis of job descriptions) does not necessarily mean that following such field of study is associated with higher chances of getting a job. One reason is related to market tightness relative to a given curriculum:

Source: World Bank calculations based on job portal data.

<sup>&</sup>lt;sup>44</sup> The category *university degree (unspecified)* refers to all job postings specifying generally the need for a university degree without specifying whether it is Bachelor, Master or other forms.

the size of cohorts in some fields may be too large to be absorbed by the market. Therefore, to contrast the previous word clouds, one can take into account labor force survey data which document the field of education followed by the current stock of workers (Figure 27). Economics, for example, corresponds to the "social sciences, business and law" field of study. While it was frequently highlighted by employers across occupations (Figure 23 to Figure 26), it is not a field of studies associated with higher chances of being employed, given the large numbers of students following this path (Figure 27 and Figure 28).



Source: World Bank calculations based on job portal data.

These findings complement the analysis on skills profiles for different industries or occupations presented above. While skills profiles can address the skills needs of an industry, education profiles inform students on education requirements for a specific job. This can help to alleviate the mismatch between youth aspirations and labor market requirements. More importantly, they offer another perspective on education and skills requirement as they express the demand-side view, while using survey data such as LFS to analyze education levels per industry or type of occupation would only provide us with the supply-side view.

#### Figure 27 : Distribution of current fields of studies among 15-29-years-old population

Figure 28: Decomposition of the 30-64 years-old population between employed and not employed (i.e. unemployed and inactive people) per field of studies





*Note:* Current field of studies, across levels of education. Data restricted to 15 to 29-years-old to better reflect the current demand for fields of education.

*Note:* Data is restricted on 30 to 64 years-old rather than 15 to 64 years-old given that some youth are still studying up to 29 years-old, especially in specific fields which require longer years of studies, and this can bias the figures on employment.

Source: World Bank calculations based on LFS 2017.

The results presented so far shed light on education and skills profiles expressed by the demand side only separately from each other. However, as education and skills levels may be subject to a certain degree of substitutability, we explore the granularity of our data to investigate interaction effects between education fields, education levels and certain skills. For this purpose, we analyze the incidence of skills in postings specifying specific education levels and certain fields of study of tertiary education.

Skill demand is usually non-specific across demand for education in certain fields of study. In fact, foreign language, computer and extraversion skills are transversal across all education fields, in line with previous results (Figure 29). This highlights that despite the specificities of certain education fields, socio-emotional, computer and language skills are considered as complements across any field of education that firms specify in their job postings. In Appendix C (Figure 41) we provide evidence that this also holds for lower levels of education. Nevertheless, highly technical jobs constitute an exception to this relationship: postings that specify the need for a computer science degree tend to also specify the need for general computer and specific software skills more pronouncedly than postings requiring broader fields of study, such as Business Administration or Economics degrees (Figure 29).



Figure 29: Incidence of skills by selected fields of required education

Source: World Bank calculations based on job portal data.

Note: Multiple education requirements can be specified for a given posting, thus fields of education requirements are not mutually exclusive.

#### VII. Conclusion

This paper illustrates how online job portals can be used to inform us on the skills demanded by employers and complement knowledge on which sectors are recruiting, using the example of Kosovo (for the segment of the labor market that is well represented). Online vacancies provide a rich source of information that can complement more traditional surveys and administrative data in two ways. First, it is "real-time" data that can be processed immediately without a time lag (as opposed to surveys). Second, it is directly sourced from employers and detailed job descriptions provide us with very granular data on what skill and education requirements employers are looking for (again, in the segment that is well represented in the online platforms). As such, these data differ from survey or administrative data which infer demand for skills by looking at characteristics of the employed (i.e., when the employment match has already happened).

The findings are illustrative of the segment of the labor market represented in such data. Job platforms are used almost exclusively for high-skill occupations in Kosovo's capital city Pristina, while many low- and medium-

skill jobs are filled through informal channels. Job postings require on average 2.5 years of work experience and tertiary education, are mostly full-time, and exhibit strong regional focus on Pristina. Foreign language, extraversion and computer skills are the most demanded skills across all industries and occupations, which confirms that they are "horizontal" skills requirements. Some industries such as the IT sector express high demand for particular software skills, such as programming skills, thereby signaling specific industry needs. From an applicant's perspective, it points at the importance of more specific abilities that need to be trained beyond these transversal skills before entering the job market.

The approach presented in this paper provides valuable insights for policy makers and other stakeholders on priority skills demanded by employers and this information should be useful in aligning curricula with labor market needs in general education and training programs. The findings on skill needs and education requirements by sectors and occupations are very detailed and directly actionable. Such findings can also benefit parents and youth looking for more information on career and education choices.

Nevertheless, skill needs are dynamic and thereby evolve quickly over time, especially technical skills in rapidly evolving sectors (for example IT and computer science), often linked to technological progress and overall economic development. The construction of time series of skill requirements to follow skill demand over time would be a key extension of this analysis. Several years of such data could allow us to monitor the changes in demand for skills, for policy use.<sup>45</sup> Another main avenue for analysis is the inclusion of data from public platforms, which would improve the coverage of the labor market. Discussions with the public employment agency of Kosovo suggest that the public platform includes more middle and low-skilled jobs than private platforms and has a better coverage of secondary cities in the country. This suggests that private and public platforms are complementary and serve different markets. We encourage further applications of this approach to other Western Balkan countries to compare skills requirements across borders and to account for enhanced regional mobility of predominantly young labor market participants.

This paper was written prior to the outbreak of the COVID-19 pandemic in Europe. While the severity and duration of the pandemic and the resulting impact on labor markets are uncertain, analysis of online vacancy data provides real-time insight into firms' demand and severe distress in the spring of 2020. Kahn et. al. (2020) for example show that job vacancies in the United States collapsed by 30 percent in late March compared with early 2020. The continued monitoring of job vacancy data will signal when labor markets are starting to recover.

<sup>&</sup>lt;sup>45</sup> For a recent example of an analysis on skills evolution in the context of the Great Recession, see Hershbein and Kahn (2019).

### References

Almlund, M., Duckworth, A. L., Heckman, J., & Kautz, T. (2011). Personality Psychology and Economics. *Handbook of the Economics of Education*, 4, 1–181.

Azar, J., Marinescu, I. and Steinbaum, M., Labor Market Concentration (2017). NBER Working Paper No. w24147.

Boselli, R., Cesarini, M., Marrara, S., Mercorio, F., Mezzanzanica, M., Pasi, G., & Viviani, M. (2018). WoLMIS: a labor market intelligence system for classifying web job vacancies. *Journal of Intelligent Information Systems*, *51*(3), 477-502.

Chernyshenko, O. S., Kankaraš, M., & Drasgow, F. (2018). Social and emotional skills for student success and well-being.

Cojocaru, A. (2017). Kosovo Jobs Diagnostic (No. 27173). World Bank. Washington, DC.

Cunha, F., Heckman, J. J., Lochner, L., & Masterov, D. V. (2006). Interpreting the Evidence on Life Cycle Skill Formation. *Handbook of the Economics of Education*, 1(06), 697–812.

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

Goldberg, Lewis R (1993). "The Structure of Phenotypic Personality Traits." *American Psychologist*, 48(1):26–34.

Handel, M. J., Valerio, A. & Sanchez-Puerta, M.L. (2016) Accounting for mismatch in low- and middle-income countries: measurement, magnitudes, and explanations. Directions in Development. Washington, DC: World Bank.

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

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, 24(3), 411–482.

Helvetas (2016). Job Matching Services in Kosovo. Retrieved from: http://helvetasks.org/eye/file/repository/CASE\_ONE\_Job\_Matching\_Services\_EYE1.pdf

Hershbein, B., & Kahn, L. B. (2018). Do recessions accelerate routine-biased technological change? Evidence from vacancy postings. *American Economic Review*, 108(7), 1737-72.

Humphreys, A., & Wang, R. J. H. (2018). Automated text analysis for consumer research. *Journal of Consumer Research*, 44(6), 1274-1306.

Kahn, L.B., Lange, F. & Wiczer, D.G. (2020). Labor Demand in the Time of COVID-19: Evidence from Vacancy Postings and UI Claims. *NBER Working Paper* No. 27061.

Kautz, T., Heckman, J. J., Diris, R., ter Weel, B. & Borghans, L. (2014). Fostering and Measuring Skills: Improving Cognitive and Non-cognitive Skills to Promote Lifetime Success. *OECD Education Working Papers* N. 110, OECD Publishing.

Kuhn, P., & Shen, K. (2013). Gender discrimination in job ads: Evidence from China. *The Quarterly Journal of Economics*, *128*(1), 287-336.

Kuhn, P., & Shen, K. (2015). Do employers prefer migrant workers? Evidence from a Chinese job board. *IZA Journal of Labor Economics*, 4(1), 22.

Kureková, L. M., Beblavý, M., & Thum-Thysen, A. (2015). Using online vacancies and web surveys to analyse the labour market: a methodological inquiry. *IZA Journal of Labor Economics*, 4(1), 18.

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

Kureková L. M. & Žilinčíková, Z., (2016). Are student jobs flexible jobs? Using online data to study employers' preferences in Slovakia. *IZA Journal of European Labor Studies*, 5(1), 1-14.

Marinescu, I., (2017), The general equilibrium impacts of unemployment insurance: Evidence from a large online job board, Journal of Public Economics, 150, issue C, p. 14-29.

Marinescu, I., & R. Rathelot (2018), Mismatch Unemployment and the Geography of Job Search. American Economic Journal: Macroeconomics, 10 (3): 42-70.

Muller, N. & Safir, A. (2019). What Employers Actually Want. Skills in demand in online job vacancies in Ukraine. Social Protection & Jobs Discussion Paper series, N. 1932. Washington, D.C. World Bank Group.

Pierre, G., Sanchez Puerta, M. L., Valerio, A., & Rajadel, T. (2014). STEP Skills Measurement Surveys: Innovative Tools for Assessing Skills. STEP Skills Measurement, (1421), 104.

Sanchez-Puerta, M.L., Valerio, A. & Gutiérrez Bernal, M. (2016). Taking Stock of Programs to Develop Socioemotional Skills: A Systematic Review of Program Evidence. Directions in Development. Washington, DC: World Bank.

World Bank. (2018). World Development Report 2018: Learning to Realize Education's Promise. Washington, DC: World Bank.

World Bank. (2019a). Kosovo Country Report: Findings from the Skills towards Employment and Productivity Survey. World Bank, Washington, DC.

World Bank. (2019b). Western Balkans Labor Market Trends 2019. World Bank, Washington, DC.

World Bank. (2020). Western Balkans Labor Market Trends 2020. World Bank, Washington, DC.

Wowczko, I. (2015). Skills and vacancy analysis with data mining techniques. In *Informatics* (Vol. 2, No. 4, pp. 31-49). Multidisciplinary Digital Publishing Institute.

Zhu, T. J., Fritzler, A., Orlowski, J.A.K. (2018). World Bank Group-LinkedIn Data Insights: Jobs, Skills and Migration Trends Methodology and Validation Results. Washington, D.C. World Bank Group.

# **Online Appendix**

Appendix A: Data preparation

Appendix B: Skills taxonomy and dictionary composition

Appendix C: Additional descriptive statistics Appendix D: Comparing STEP Employer survey data with job platform data

**Appendix E: Replication**