# Skills Demanded by The Labor Market And Supplied by Russian Universities: An Investigation

Artem Volgin
Suhas Parandekar

## Introduction

The motivation for this paper is to lend support to the human capital aspect of the Russian Federation development strategy declared in the May decree of 2018. Poverty reduction and the progression to the top five world economies are the key Russia's development goals. The central role of human capital for the achievement of these goals is acknowledged by the government policy statements and the World Bank research. Skills that people possess are central to human capital. To augment national wealth the richest countries in the world learned to harness the skills and qualifications of people.

Valid and reliable information about skills trained by universities and possessed by their graduates is critical for a variety of stakeholders, including (1) students – to navigate their academic trajectories and increase education premiums; (2) universities – to reveal their achievements and failures in the provision of specific competencies; (3) companies and organizations – to hire the most skillful graduates; (4) government agencies – to advance policies stimulating the development of skills and abilities required by the economy.

To obtain such information, we use a sample of CVs data provided to us by HeadHunter - the leading online recruitment platform in Russia and the third-largest in the world[1]. The database of HeadHunter consists of more than 50 million CVs of candidates; companies publish more than 800 thousand vacancies every month.

Employing the HeadHunter data, this paper aims at answering three main research questions. How important are Russian universities for the acquisition of skills? Which specific universities and faculties are better at providing different skillsets? What are the university-based factors that escalate skills learning?

The remainder of the paper is organized as follows. First, we outline findings from past literature that analyzed people's skills possession and exploited data from online recruiting platforms. Second, a thorough description of our dataset and methodology is provided. Third, an exploratory data assessment is conducted, and the observed patterns are compared with the official statistics. Forth, our random effects modeling results are delineated. Finally, we discuss the results obtained and formulate policy recommendations.

[1] https://www.similarweb.com/top-websites/category/career-and-education/jobs-and-employment

# Background

## *Human Capital and Skills*

The importance of knowledge and skills is increasingly growing in the modern era. However, the attained education level is a rough measure of the knowledge an individual possesses. Human capital indicators require not only information about one's educational inputs, but also evidence of their actual skills proficiency. The strategies aimed at amplifying the level of human capital should be tailored to social and economic demands in the country (OECD 2001).

However, the current education systems in many countries around the world experience challenges in providing their students with skills allowing them to effectively compete for job positions. This mismatch between skills formed by universities and skills demanded by the labor market can impede successful economic development and depreciate opportunities for one's professional growth over the life-course (Patrinos 2020). The research harnessing data about people's skills can help overcome the existing disagreement between the skillsets individuals acquire and the skillsets expected by employers in the real context-specific labor market settings.

## *Data from online recruiting platforms*

It is worth mentioning that due to the dynamic nature of skills, traditional surveys or administrative data sources might be fruitless in obtaining extensive skills-related information. Nevertheless, such a type of data can be collected from a different source, namely online recruiting and job posting platforms. These data enable researchers to carry out investigations employing information about job ads and resumes or profiles of users, i.e., the demand side and supply side, respectively.

In the area of education studies, such type of online data was previously utilized in the exploration of people's transition from learning to the working environment (Cabrera and Lloret 2017; Case, Han, and Rimes 2016; Li et al. 2016) in evaluating the quality of education programme (Elexpuru Albizuri and Fresno Anabo 2017), and in the ranking of universities based on the graduates' employability (Moreno-Delgado, Orduña-Malea, and Repiso 2020). Since large samples of such data oftentimes are not publicly available, past research works concentrated their analyses on small subsamples by manually collecting information on candidates' resumes and vacancies posted by employers on such online platforms.

Few exceptions from this can be found in the works of Deming & Kahn (2017) and Hershbein & Kahn (2018) who leveraged Big Data containing job ads collected from an array of online sources. They showed that the US labor market increasingly rewards the possession of social skills.

Another notable research was conducted by the World Bank in collaboration with LinkedIn (Zhu, Fritzler, and Orlowski 2018). Based on the worldwide data from the profiles of LinkedIn users the researchers showed that such data are best for the representation of highly competent labor force and knowledge-intensive job sectors. The study argues that LinkedIn data are complementary to the official statistics since they convey a detailed description of the skills-related dimension of human capital.

*Russian context*

The mentioned discord between competencies demanded by companies and mastered by employees is a key factor preventing the expansion of innovations in Russia (The World Bank 2015). Insufficient supply of high-quality university education, facilitating ample provision of versatile training, leads to the disbalance between professional skills that workers have, and the ones anticipated by the market. For instance, previous research demonstrated that some sets of skills are highly valued by the Russian economy: the possession of well-developed non-cognitive skills is essential for highly qualified employees (Gimpleson, Zudina, and Kapeliushnikov 2020) and such possession correlates with greater education returns (Rozhkova 2019). However, the Russian labor market was shown to be relatively short on social and higher-level cognitive skills (The World Bank 2015).

The volume of studies using online data for both the demand and supply sides of the skills market in Russia is scarce. One example comes from the World Bank investigation in which the HeadHunter data were exploited to examine the skills demand in Ukraine (Muller and Safir 2019). The researchers showed that most of the vacancies on HeadHunter require a multitude of socioemotional skills and contended about the complementary nature of cognitive and socioemotional abilities. However, the information retrieved from CVs of job seekers posted on the website has never been leveraged before.

## Data

The current study leverages CVs data collected from the HeadHunter. Both CVs and vacancies published on the website have a range of features that distinguish these data from regular survey data or official statistics. First, the large size and socioeconomic/demographic granularity of the data allows researchers to scrutinize specific subgroups of interest (e.g., departments at a certain university). Second, new vacancies and CVs are uploaded to HeadHunter on a daily basis meaning that such data can be instrumental for careful consideration of rapid changes in the labor market (e.g., for tracing the effect of COVID-19 measures). Third, the HeadHunter database of CVs contains a lot of unique information which is usually absent in standardized surveys, such as the indication of skills possession, concrete names of universities and faculties.

However, the specificity of the data under discussion imposes particular constraints on the opportunities of statistical analysis and entails a number of shortcomings. The unstructured format of many fields in the data involves quite a tedious pre-processing stage which can also include a lot of manual coding. In the Methodology Section, we will display several examples of procedures that enable to appropriately curate such unstandardized fields. There are several issues that are common to almost all online data. First, not all fields in one’s CV are obligatory to be filled out which ends up in massive missing patterns in the data. Second, the data are not representative of the Russian labor market as a whole, albeit fairly representative of the set of workers who obtain employment through online mediums. As we will show later in the paper, this underrepresentation is by no means drastic, particularly given the focus only on applicants with higher education in the present study.

The research uses a sample of 988,831 resumes, which makes up 4% of all resumes available on HeadHunter from 2010 to 2020. The data encompass a spectrum of different technical and socio-demographic variables, such as a *year of CV creation*, *region* in which an applicant is looking for a job, *gender*, and *age group* (to prevent any possibilities for deanonymization the age of applicants was provided by HeadHunter in the form of age groups). There is also information about *professional areas* of potential employees and their monthly *expected salary* in rubbles. Education-related variables include the attained *education* (unfinished higher education, bachelor, master, and just higher education without specifying the precise level), *university and faculty name(s*) (textual fields), and a *year of university graduation*. The skills data for each user may incorporate up to 30 distinct skills.

In our research, we selected only applicants who completed their education in state universities between 2000-2020 for purposes of comparability with the official statistics. The applicants who submitted their CVs in 2010-2013 were filtered out due to the prevalence of missing values in the skills possession variable for this subsample. The indication of skills was not a popular feature on HeadHunter during this period with only 15-25% of applicants indicating their skills in their CVs but rocketed dramatically starting from 2014 and reached nearly 90% in 2020. Another filtering criterion referred to the presence of non-missing faculty and university names in users' resumes. In cases of several degrees, only the last university indication was retained as we assume it had a larger impact on people's future professional and educational trajectories. The final database for the subsequent analysis encapsulated 259,133 CVs.

In addition, the data collected by the Russian government from the website graduate.edu.ru were utilized for comparison with our findings. Graduate.edu.ru is the official portal for monitoring the employment of graduates; it was launched by the Russian Ministry of Education in 2015. The website allocates information (obtained from the Pension fund) on employment records of people who graduated from the Russian state universities and vocational education colleges. In the current investigation, several indicators from graduate.edu.ru were leveraged: the *total number of graduates* in 2013-2015 for each state university, the *official salary of those graduates* for the same period, and the total number of graduates in 2013 by *specializations* of their faculties. More details on the data one can find in the recent research dedicated to vocational and private returns by education organizations (Melianova, Parandekar, and Volgin 2020).

## Methodology

The core data in our study include three crucial attributes of the HeadHunter applicants: the university of graduation, the attended faculty, and their skills. As was mentioned, this information is structured as open textual fields which inevitably provokes a lack of unity in the names people use to describe identical entities and also leads to the incidence of too small or nearly duplicated skills categorizes. To illustrate, Moscow State University can be called as "Московский государственный университет", "Moscow State University", "МГУ", "МГУ имени М.В. Ломоносова" etc. All these names are valid in denoting the MSU, but we need to merge them into one entity to make the grouping of graduates feasible. The recommended classical string-matching algorithms (e.g., fuzzy string searching) work poorly for the given task due to the presence of abbreviations in our data (e.g., "Московский государственный университет" has only 3 out of 38 symbols in common with "МГУ"). In the beginning of this section, we suggest an approach of how to deal with these issues and then outline the specification of our main analytic models.

## *University names*

To deduplicate university names, we exploited a geocaching procedure. Geocaching is a term that refers to the process of obtaining precise coordinates of a geographical entity or vice versa. The tools for geocaching implementation are provided by IT companies such as Google or Yandex in the form of APIs: searching on Google maps for “МГУ” and “Московский государственный университет” one is able to retrieve the same set of coordinates. We exploited Google Geocoding API[2] to obtain coordinates of the university names most frequently observed (reported more than 10 times) in our dataset. This allowed us to join identical universities based on their coordinates. After removing university branches, private universities, and universities with less than 30 CVs from the sample, we ended up with 403 unique organizations of higher education.

## *Faculty names*

To utilize information about faculties, textual responses were coded employing manually selected keywords (see Table 1). We stuck to the categorization of specializations elaborated by the Russian Education Ministry with 8 domains: Mathematics and Natural Science, Engineering and Technology, Health and Medicine, Agricultural Science, Social Science, Art and Culture, Education and Humanities.

**TABLE 1.** Key Words for The Categorization of Faculties

| **Faculty specialization** | **Key words** |
|---|---|
| Mathematics and Natural science | математи \| физик \| химия \| химичес \| статисти \| геолог \| географ \| биолог \| гидро \| эколог \| физмат \| кибернет \| земля \| недропольз \| естествен \| природный |
| Engineering and Technology | технолог \| техничес \| инженер \| строител \| информа \| транспорт \| компьютер \| энергети \| программир \| бурен \| оптика \| энерго \| вычислите \| трактор \| материал \| прибор \| производ \| промышлен \| автомат \| связь \| механик \| механич \| эксплуат \| электро \| техник \| машин \| газос \| минерал \| механиз \| металл \| робот \| нефтян \| нефть \| газовый \| аэро \| космиче \| двигател \| ядерны \| архитек \| автомобил \| дорожн \| горный \| авиаци \| аппарат |
| Health and Medicine | медицин \| медико \| сестринск \| здравоохран \| фармацев \| лечеб \| стоматоло \| фармация |
| Agricultural science | агроном \| садовод \| лесное \| зоотехн \| лесной \| лесозаготов \| дерево \| сельское \| почвовед \| природо \| овощевод \| ветеринар \| растительный \| водоснабжение |
| Social science | социол \| социал \| бизнес \| эконом \| финанс \| политолог \| психолог \| менеджмен \| бухгалте \| товаровед \| политич \| урбанис \| торгов \| регионовед \| отношени \| политик \| реклам \| обществен \| журнал \| маркет \| логист \| банков \| демограф \| менеджемент \| туризм \| сервис \| гостини \| юридич \| юриспру \| юрист \| право \| медиа \| аудит \| таможен \| коммерц \| налог \| судебный \| муниципальный |
| Education | педагог \| детство \| дефектолог \| логопед \| педиатр |
| Humanities | археолог \| философ \| региовед \| теолог \| лингвист \| истори \| филолог \| гуманитар \| язык \| перевод \| африка \| восток \| архив \| физический культура |
| Art and Culture | дизайн \| искусств \| хореограф \| музыка художествен |

[2] https://developers.google.com/maps/documentation/geocoding

## *Skills categories*

The amount of unique skills observed in the sample was fairly condensed (due to the restriction on mentioning no more than 30 skills and due to the popularity-based hints popping up for a user when he/she enters skills information) but relatively large. In the current work, we delved into the analysis of sizeable clusters of abilities, rather than distinct skills, and concentrated on hard skills (e.g., Programming or Analytical) as well as soft skills (e.g., Negotiation or Teamwork and Leadership), which could be affected by the university studies. Particularly, we coded 3,000 most frequent skills to 15 skill categories (see Table 2). For instance, "SQL", "Git", "Python" skills were moved to the category "Programming", whereas "Leadership qualities", "Teamwork", "Ability to work in the team" skills – to the category "Teamwork and Leadership".

**TABLE 2.** Skills Categories and Examples of Skills Related to Them

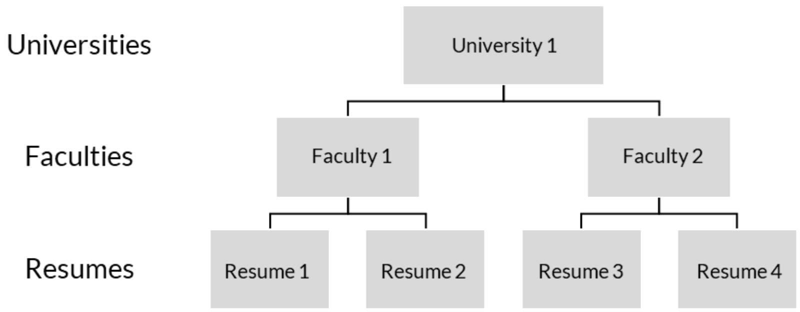
| Skill Category | Example of skills (translated from Russian) |
|---|---|
| Analytical | Business Analysis, Data Analysis, Data Collection and Analysis, Business process modelling |
| Financial | Tax reporting, Budgeting, Cost-benefit calculation, Procurement and supply chain management |
| Law | Claim work, Civil law, Labour Code of the Russian Federation, Corporate law |
| Medical-psychological | Pharmacology, Dentistry, Pediatrics, Child psychology |
| Negotiation | Negotiation, Crisis communications, Conflict management, Interviewing |
| People Management | People management, Team management, Assessment of candidates, Staffing strategy |
| Presentation and Public speaking | Presentation skills, Public appearances, Public speaking, Conducting presentations |
| Programming | SQL, Git, JavaScript, C++ |
| Project Management | Project Management, Agile, Organization of PR campaigns, Scrum |
| Social media marketing | Social Media Marketing, SEO, Digital Marketing, Targeting |
| Software for design, architecture and planning | 3D Max, Adobe After Effect, AutoCAD |
| Software for finance, management and analytics | CRM, Consultant Plus, SAP, 1C |
| Teaching and Training | Staff training, Ability to teach others, Mentoring, Coaching |
| Teamwork and Leadership | Teamwork, Leadership qualities, Team player, Ability to work in groups |
| Writing | Proofreading, Writing articles, Fast typing speed, Copywriting |

Some skills inputs such as general social skills (e.g. "sociability") or personal characteristics (e.g., "responsibility") were excluded from the analysis: we assume such skills do not introduce any meaningful variability to the data and are not substantially influenced by the university learning curriculum.

## *Specification of Multilevel Models*

In this analysis, skills possession variables were used as dependent variables in a binary form. We expect that applicants have more similar skills within universities and faculties than between them. To explore how important the role of universities and faculties in providing competencies to their students is, a set of multilevel models that enable us to account for such a nested structure was utilized. We erected separate three-level logistic regression models with random intercepts for each skill as a dependent variable, where individuals are nested within faculties and faculties are nested within universities (see Figure 1 for the graphical representation of the hierarchical structure).

**FIGURE 1.** Graphical Representation of the Multilevel Structure




The Intraclass Correlation Coefficient (ICC) was computed for each null (without predictors) random intercept model. The ICC denotes the proportion of variance of the dependent variable explained by the multilevel structure. The ICC ranges between 0 and 1, higher values of it are associated with greater importance of the underlying hierarchy, i.e., of universities and faculties. The ICC is expressed by:

$$ICC = \frac{S_b^2}{(S_b^2 + S_w^2)}$$

$S_b^2$ = between group variation,

$S_w^2$ = within group variation.

Further, we built models with control attributes and made an emphasis on random intercept effects. The following covariates were controlled for in the analysis: *year of CV creation, graduation year, experience, region, professional area, gender, age,* and *education level*. Finally, we carefully examined the effects of *education level, faculty type*, and *university specialization* on the chances of skills acquisition.

**Full Models Specification:**

**Resume (1st level):**

$$ln\left(\frac{P\left(Y_{ijk}^{s}\right)}{1-\ P\left(Y_{ijk}^{s}\right)}\right) = b_{0jk} + b_{1jk}(CV\ creation\ year) + b_{2jk}(graduation\ year) + b_{3jk}(experience) +$$

$$b_{4jk}(region) + b_{5jk}(professional\ area) + b_{6jk}(gender) + b_{7jk}(age) + b_{8jk}(education\ level)$$

**Faculty (2nd level):**

$$b_{0jk} = \pi_{00k} + \pi_{0jk}(faculty\ type) + r_{0jk}$$

$$b_{ijk} = \pi_{i0k} \text{ for } i \neq 0$$

**University (3rd level):**

$$\pi_{00k} = \gamma_{000} + \gamma_{00k}(university\ specialization) + u_{00k}$$

where a resume *i* is nested within a faculty *j*, which is nested within a university *k*, $Y_{ijk}^{s}$ is a dummy variable for a skill, *s* denotes a distinct skill of interest, $r_{0jk}$ and $u_{00k}$ are the second- and third-level errors respectively.

## Data Exploration and Validation

In this section, we lay out an exploratory overview of the main variables of interest, discuss their limitations, and compare the HeadHunter data with the official statistics.

### *Socio-Demographic characteristics*

Table 3 depicts the descriptive statistics of the main socio-demographic characteristics in our sample, namely gender, age group, education, and region of the job search.

It can be noticed that the proportion of CVs published by females (58%) is greater than by males (42%). The age distribution is highly skewed towards younger people: most of the applicants in the sample are aged 18-44, there is only around 3% of individuals older than 45 years. This can be explained by two factors: (1) generally, younger people use online services more often, and (2) more experienced (and therefore usually older) applicants change their work less frequently and prefer to look for new positions relying on their accumulated professional networks. We elaborate on this aspect later along the lines.

The sample under consideration is representative of CVs with unfinished higher education and higher. It should be mentioned that a lot of applicants marked the level of their education as Higher, which may imply both a Specialist and a Bachelor level. However, given a substantial drop in the percentage of resumes with education level labeled as Higher from 92% in 2000 to 37% in 2020, we may assume Higher education holders mostly have gained a Specialist diploma, although the share of Bachelor holders among them can be pronounced. Additionally, the Unfinished Higher category implies two groups of job candidates: those who are still studying at a university on the date of a CV placement and those who studied at a university but have not completed it.

Another observation is that applicants who are seeking a job in Moscow and Saint Petersburg account for around 40% of the whole sample; however, many of those people may actually reside in the nearest regions (e.g., Moskovskaja oblast').

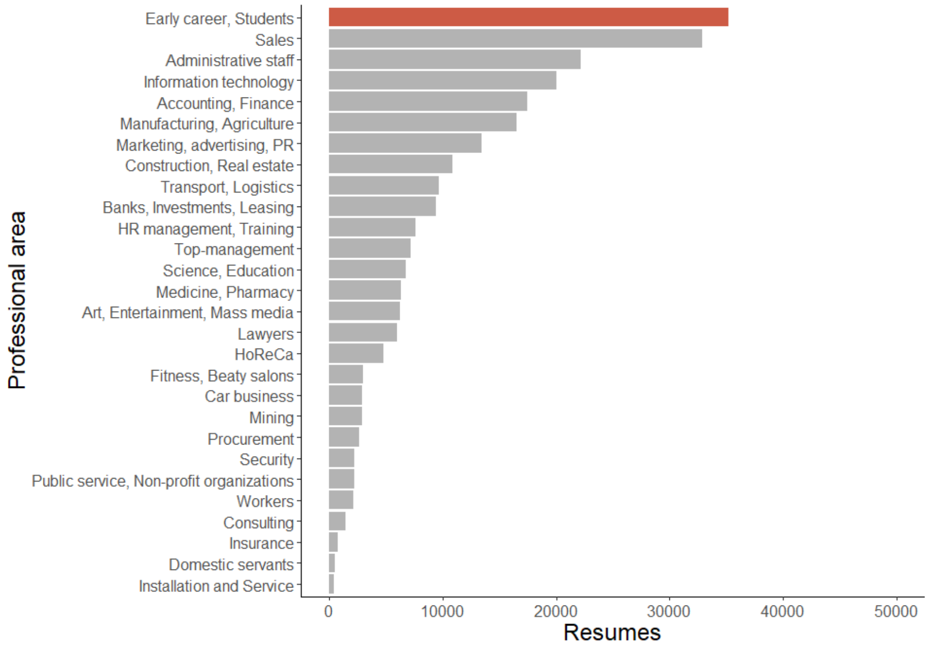
**TABLE 3.** Description of the Main Socio-Demographic Variables

| Variable | Group | Number of resumes | Percentage of resumes |
|---|---|---|---|
| Gender | Female | 147191 | 58.0% |
| | Male | 106383 | 42.0% |
| Age group | 18–24 | 87264 | 34.4% |
| | 25–34 | 118657 | 46.8% |
| | 35–44 | 40829 | 16.1% |
| | 45–54 | 6229 | 2.5% |
| | 55+ | 588 | 0.2% |
| Education level | Unfinished Higher | 27212 | 10.7% |
| | Higher | 190930 | 75.3% |
| | Bachelor | 20837 | 8.2% |
| | Master | 12852 | 5.1% |
| | Candidate | 1743 | 0.7% |
| Region | Moscow | 68068 | 26.8% |
| | Saint Petersburg | 32467 | 12.8% |
| | Moskovskaja oblast' | 11368 | 4.5% |
| | Krasnodarskij kraj | 10564 | 4.2% |
| | Respublika Tatarstan | 7860 | 3.1% |
| | Sverdlovskaja oblast' | 6948 | 2.7% |
| | Novosibirskaja oblast' | 6679 | 2.6% |
| | Nizhegorodskaja oblast' | 6429 | 2.5% |
| | Samarskaja oblast' | 6195 | 2.4% |
| | Respublika Bashkortostan | 5929 | 2.3% |
| | Other regions | 91067 | 35.9% |

## *Professional areas*

Figure 2 displays the distribution of professional areas in CVs on HeadHunter. The most popular categories represent early-career workers/students and employees in sales. The administrative staff, applicants from the IT sector, accounting, finance, manufacturing, agriculture, and marketing are also quite prevalent on HeadHunter. Since our sample was confined to people with unfinished higher education and above, and online job-posting platforms are usually skewed towards highly knowledgeable specialists, such areas as mining, workers, and security are logically underrepresented in our dataset.

**FIGURE 2.** Distribution of Professional Areas

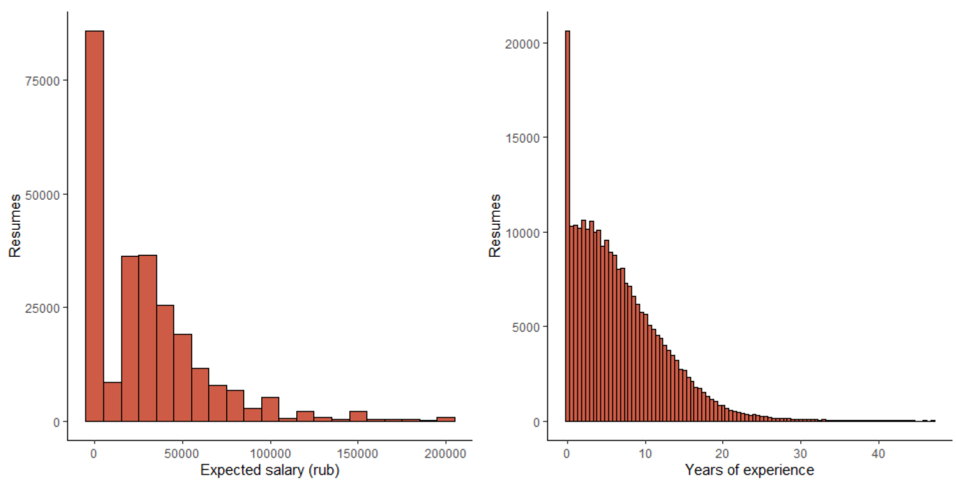



HeadHunter exploits its own classification scheme to define professional categories of applicants which impedes a direct comparison of findings with governmental statistics, although is user-friendly. It should also be noted that the HeadHunter's categorization blends an array of different dimensions such as a position in the organizational hierarchy (e.g., "Early career, Students", "Top-management"), professional sphere ("Administrative staff", "Lawyers"), and business area (e.g., "Car business", "Fitness, Beauty salons"). Such overlapping categories can potentially cause information loss and bring about bias in estimates.

## *Expected Salary and Experience*

Figure 3 shows the distribution of the expected salary and years of experience variables. Both variables discernibly peak around zero, meaning that there is a large group of candidates who are looking for unpaid internships to gain experience: they have little or no practice in their field but are ready to work for free at early stages. We keep this group in our subsequent analysis as these people possess certain skills that they mastered outside their work environment.

**FIGURE 3.** Expected Salary and Years of Experience

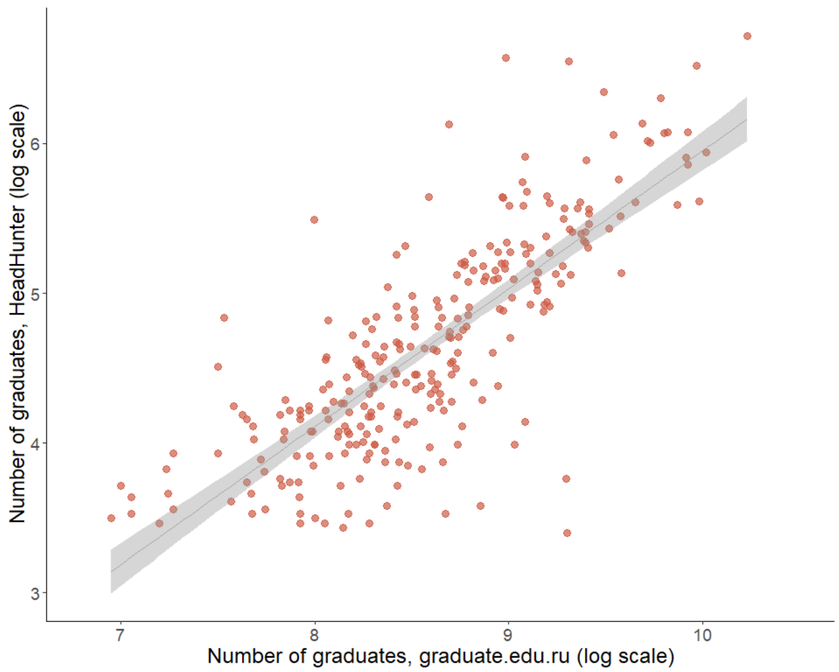


Conversely, one can observe a lack of highly experienced applicants in our sample: CVs reporting 20 years of experience and more are quite rare. Specialists with ample expertise in the field are embedded in various professional networks and informal channels, which may tangibly simplify the job search and make the usage of online platforms unnecessary. Such an underrepresentation is not severe trouble in the current investigation since the effect of a university on professional skills formation fades away over time and after a longer period may become negligible.

### *Applicants and Universities*

Figure 4 demonstrates the number of university graduates from graduate.edu.ru and HeadHunter in 2013-2015, Figure 5 portrays the same comparison but for the university-averaged expected and obtained salaries collected from HeadHunter and graduate.edu.ru, respectively. As we can see, information about the amount of university graduates on HeadHunter rather accurately corresponds to the government data (Pearson's correlation coefficient is equal to 0.79).

**FIGURE 4.** Number of Graduates from Universities in 2013-2015, graduate.edu.ru and HeadHunter comparison




As for the average expected salaries on HeadHunter and the average obtained salaries from graduate.edu.ru, the association is quite strong as well. Interestingly, the average expected salary is rendered higher than the obtained one. This can be interpreted by the fact that fresh graduates may tend to overrate the value of their abilities and professional preparation in the labor market. On the other hand, official statistics can also downplay graduates' actual remuneration.

**FIGURE 5.** Average Salary of Graduates from Universities in 2013-2015, graduate.edu.ru and HeadHunter comparison

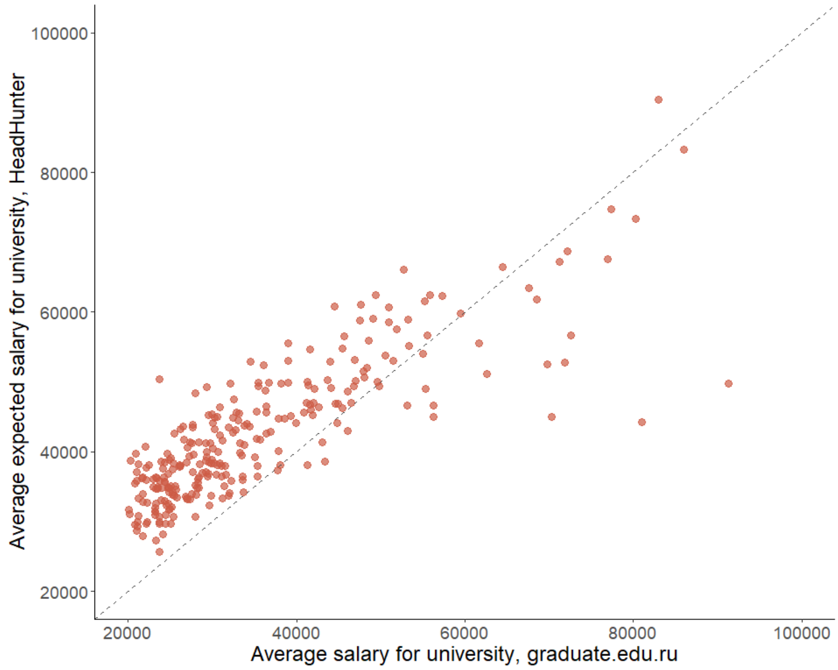
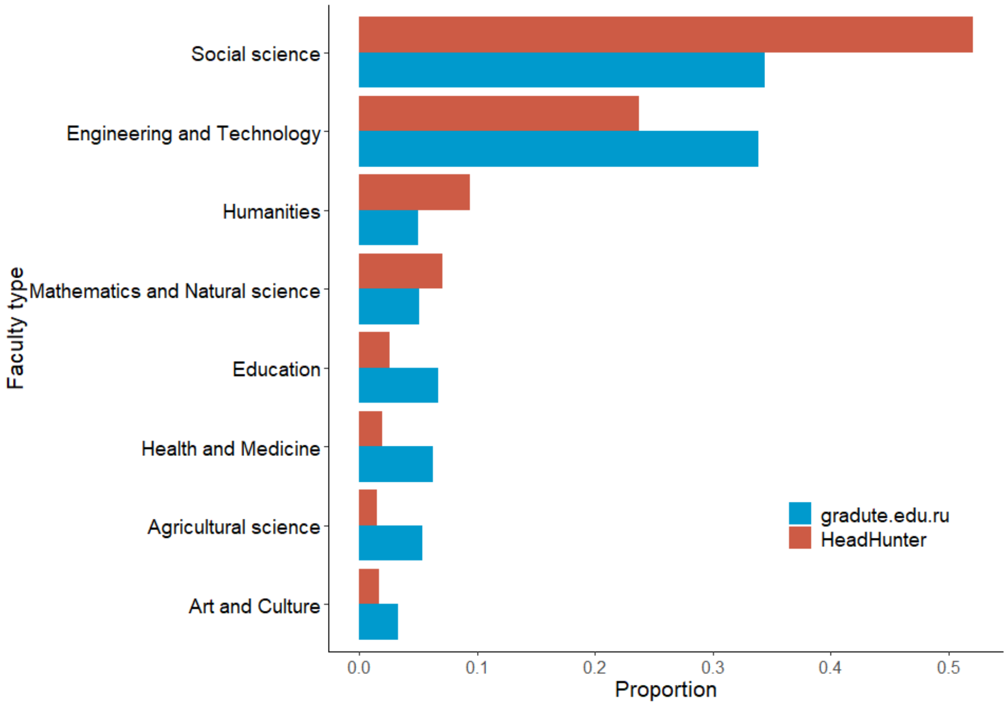



Figure 5 entails another notable finding: universities scattered in the lower triangle of the graph are characterized by the outstripped obtained salary compared to the expected salary. The most extensive obtained-expected salary discrepancies pertain to *The Russian Presidential Academy of National Economy and Public Administration* with more than 40,000 rubles difference and *The Saint Petersburg State University of Civil Aviation* with around 36,000 rubles difference. Both universities belong to the public sector, hence a lot of such students receive their first job right after graduation. We can maintain that people who were awarded a degree in such public sector universities and at the same time use the HeadHunter platform may experience difficulties in finding a job, thus they anticipate reduced remuneration for their work compared to the market average in the area.

## *Applicants and Faculties*

Figure 6 depicts the share of university graduates in 2013 by faculty specializations on HeadHunter and based on the official statistics. It is visible that Social Sciences and Humanities are overrepresented on HeadHunter, while STEM graduates are slightly underrepresented. This may be rooted in the fact that there are other popular platforms for online recruitment in the IT domain (e.g., https://career.habr.com), which can drag some STEM job candidates from HeadHunter. The underrepresentation of public sector applicants on HeadHunter is again apparent: faculties with an explicit government orientation (Education, Health and Medicine, and Agricultural Science) embrace more modest proportions of graduates on HeadHunter compared to the official statistics.

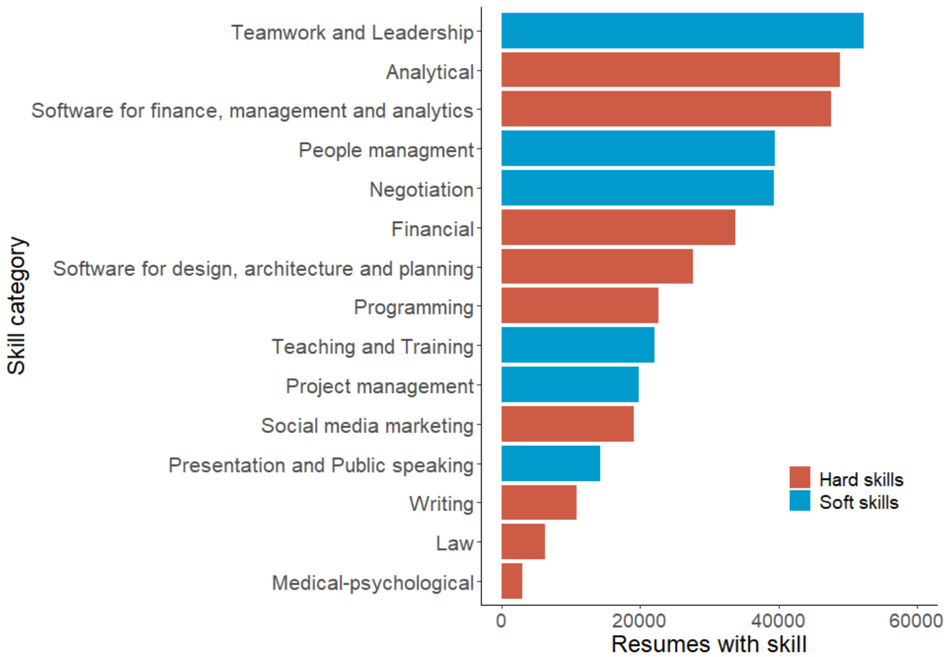
**FIGURE 6.** Proportion of Graduates in 2013 by Faculties, graduate.edu.ru and HeadHunter comparison




## *Skills categories*

Figure 7 presents the distribution of the skills categories under focus. The applicants on HeadHunter report a variety of both "Soft" and "Hard" skills; the most popular amongst them are the following: teamwork-oriented, analytical, and software-related in financial, managerial, and analytical fields. The possession of certain expertise in management and negotiation is also frequently mentioned in CVs of the HeadHunter users.

**FIGURE 7.** Distribution of the Skills Categories



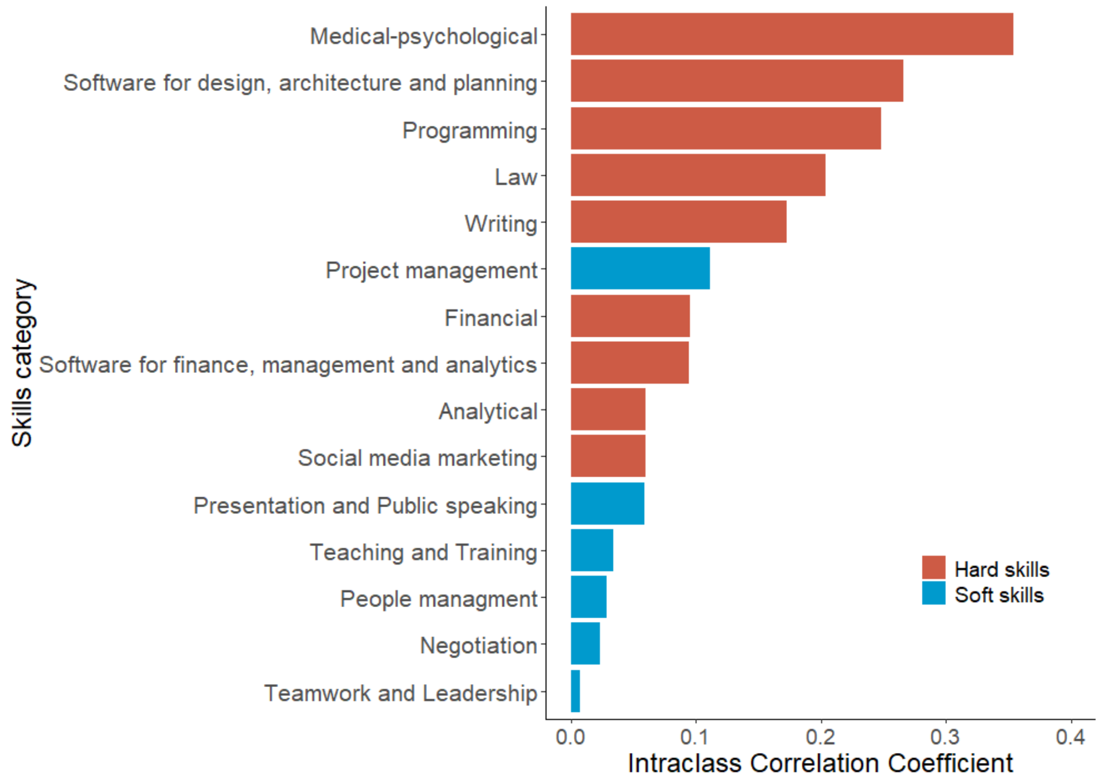
## Modelling Results

### *1. How important are Russian universities for the acquisition of skills?*

To answer the first research question, the Interclass Correlation Coefficient (ICC) for null random intercept models was calculated. Figure 8 arranges different skills categories by the magnitude of their ICC. It is vivid that medicine- and psychology-related abilities are at the top of the ranking. This may stem from the fact that the knowledge delivery in medicine is secured by academia and usually does not take place outside of it. Medical skills are followed up by the knowledge of specific software for design, architecture, and planning. Interestingly, despite the increase in the number of self-taught programmers in recent years, we can claim that the higher education track sustains its significance for mastering the proficiency in programming languages – this skills category occupies the third position based on the ICC ranking.

However, Russian universities seem to be less essential for most of the soft skills under analysis. One exception with relatively high ICC values refers to Project management skills; the rest of the soft skills such as Presentation, Teaching, People management, Negotiation, and Teamwork are placed at the bottom of the hierarchy. Overall, this may be portrayed as evidence pointing to the insufficient importance of Russian universities in the provision of soft skills rather than hard skills.

**FIGURE 8.** Intraclass Correlation Coefficient for the Different Skills Categories




### *2. Which specific universities and faculties are better at providing different sets of skills?*

Multilevel modeling allows us to inspect random intercepts effects for each university-faculty pair. Here we display examples of the top-10 random intercepts for the five full models: 3 models for top Hard skills by ICC (Medical-Psychological, Software for design, architecture and planning, and Programming skills) and 2 models for top Soft skills by ICC (Project management and

Presentation and Public speaking skills). The computed random intercept lists arrange university-faculty pairs by graduates' chances to acquire a skill holding all control characteristics constant. In other words, these effects show the likeliness of a concrete university faculty to contribute to the formation of the skills under focus.

Figure 9 depicts random intercepts for Medical-Psychological skills. Russian faculties and universities that are most successful in providing the related qualification are the following: Psychology faculties at Saint Petersburg State Institute of Psychology and Social Work and Moscow State Universities of Psychology and Education. Interestingly, medical universities affiliated with the Russian Ministry of Health are amongst the top-10.

**FIGURE 9.** Random Intercepts for Medical-Psychological Skills, TOP-10 University-Faculty


The random effects coefficients for the proficiency in specific design-, architecture- and planning-oriented software programmes are portrayed in Figure 10. It is vivid that an array of different faculties pertaining to STEM degrees such as Engineering and Technology graduates from the Moscow Institute of Architecture are taking the lead.

**FIGURE 10.** Random Intercepts for Software for design, architecture and planning Skills, TOP-10 University-Faculty


A similar picture can be detected for the possession of Programming Skills (see Figure 11). Top-10 positions are mostly represented by STEM faculties with LETI and MEPhI occupying the first two places.

**FIGURE 11.** Random Intercepts for Programming Skills, TOP-10 University-Faculty


Despite the fact ICCs in our models with soft skills as dependent variables are relatively low, some universities can still release students with better soft skills compared to the rest. Top-10 random intercepts for university-faculty pairs specialized in Project Management, Presentation and Public speaking skills can be noticed in Figures 12 and 13.

It is clear that the most prestigious Russian universities including the Higher School of Economics, Lomonosov Moscow State University, Saint Petersburg State University, Moscow State Institute of International Relations, Novosibirsk National Research State University retain the lead regarding the level of soft skills possession by their graduates. Notably, the bulk of such faculties comes from Social Science disciplines; however, there are a few exceptions such as Mathematics and Natural Science faculties at MSU whose graduates are fairly skillful at presentations.

**FIGURE 12.** Random Intercepts for Project Management Skills, TOP-10 University-Faculty


**FIGURE 13.** Random Intercepts for Presentation and Public speaking Skills, TOP-10 University-Faculty

## *3. What are the university-dependent factors that escalate skill learning?*

In this subsection, we elaborate on the effects of the three categorical variables in our multilevel models on skills formation: *education level* (reference group – unfinished higher education), *faculty category* (reference group – social science faculties), and *university specialization* (reference group – classical universities). The full table with coefficients can be found in Annex (Table A1), here only the most remarkable results are discussed.

*Education Level*

One curious finding is that the probability of acquiring skills increases as students move up in their university studies. To illustrate, for bachelor holders the chances of obtaining Presentation and Project management skills are 30-45% greater compared to applicants with unfinished higher education (reference), whereas for the people with master degrees these odds are 2 times as high as for the reference, and for job seekers with a candidate of science degree – almost 3-4 times as high. A similar pattern was captured for the odds of Analytical skills possession. Generally, candidates of science demonstrate the largest odds of indicating skills related to Analytics, Writing and Presentation compared to bachelor and master holders.

*Faculty Type*

Additional insights can be gained from the inspection of the odds ratios for the faculty variable. The estimated coefficients demonstrate that STEM workers (Engineering and Technology, Mathematics and Natural Science) lag behind Social Sciences specialists regarding the predominant part of soft skills categories. Although STEM professionals are usually way ahead of Social Science graduates in terms of technical competencies (such as Programming, Analytical abilities, Software knowledge), the insufficient readiness in soft skills may downgrade the value of their technical preparedness for the prospective employer.

*University Specialization*

It also essential to explore the effect of the university specialization on skills possession. The parameter estimates in Table A1 demonstrate that multidisciplinary universities in Russia are better at delivering preparation in major skills compared to more specialized universities. Strikingly, STEM universities seem to perform less satisfactorily than the multidisciplinary ones in supplying training even in programming or software knowledge – their specialization domains. This association can be explained by the favorable effects of a versatile environment at a university on skills acquisition. It is though not the case for universities with socioeconomic orientation: they are slightly better than the multidisciplinary ones in terms of supplying experts with analytical and financial skills.

## *Limitations*

There are several limitations in this study that are worth discussing.

As was stated, the absence of a detailed and comparable categorization of occupations may lead to the omitted variable bias. However, that problem can be tackled in future research by coding job titles indicated in CVs in a free textual form by HeadHunter users.

Although our analytic sample was quite voluminous, the subsamples for university-faculty pairs were rather small which led to large standard errors in the effects for a range of faculties. The enlargement of the sample size can help deal with the issue.

In general, the type of data leveraged in the present investigation appropriately represents the skilled labor force with higher education. However, some groups of applicants are poorly represented on HeadHunter, for example, public sector workers or STEM graduates. The shortcoming can be reduced by enriching the current dataset with relevant information extracted from other online job posting platforms that are popular in Russia.

The methodology exploited in the paper does not permit substantive interpretation involving causality terminology. Despite the usage of control characteristics, we were able to detect only associations (not causation) between universities and chances of skills formation. Prospectively, causal findings can be inferred by longitudinally tracking changes in the CVs data over applicants' study period.

In this study, as in many similar papers examining online posting data, we dealt with a binary absence or presence of a skill, hence there was no indication on how proficiently individuals can apply their skills. There are also several biases connected with the skills data. First, applicants report only competencies that are relevant to a professional area in which they search for a job. Second, some HeadHunter users may misleadingly mention skills possession without being proficient in them to showcase a competitive strength and make a better impression on a potential employer. Nevertheless, we assume that such bias is partially neutralized in our exploration by the usage of control covariates such as socio-demographic attributes, experience, and professional areas. This can somewhat balance the distribution of skills misspecification between universities and faculties and help avoid rough bias in parameter estimates.

## Discussion and Policy Recommendations

The current paper exhibits both methodological discoveries aiding effective usage of HeadHunter CVs data and a number of substantive findings that are informative for the Russian education policy.

**The resume data on HeadHunter offers a unique opportunity to monitor competencies procured by graduates of Russian universities at a massive scale**. The data size permits researchers to evaluate highly granular statistical models and uncover the learning benefits obtained by small subgroups of applicants. The tracking of skills acquisition can be conducted during the learning process as well as after graduation. It is vital though to regard potential bias and inconsistencies embraced by the data and treat them properly.

**Russian universities require a sharpened focus on soft skills training in the curriculum.** That can be done by adopting successful practices of universities that gained an upper hand in stimulating a favorable learning environment for mastering soft skills.

**Soft skills preparation should receive more attention in bachelor's as well as STEM programmes.** The analysis revealed that soft abilities such as presentation skills or project management are better practiced by masters and Ph.D. students. This means that bachelor graduates entering the labor market right after graduation may appear unprepared for job positions demanding soft skills proficiency. Additionally, experts in technical areas will greatly benefit from an environment incentivizing soft skills development.

**Specialized universities will benefit from more multidisciplinary exposure.** Multidisciplinary universities in Russia outstrip specialized universities at delivering high-quality preparation in the predominant amount skills. Therefore, a learning environment stimulating multidisciplinary content seems quite advantageous for successful skills acquisition, especially for technical qualifications.

The present research is a first attempt to approach the examination of the HeadHunter CVs data. Future work can be oriented towards the analysis of specialized skill areas such as the knowledge of distinct programming languages and software. The prospective studies will gain from scrutinizing multiple interaction effects between combined sets of abilities that individuals possess. Other areas for future discoveries can revolve around monitoring professional or migration trajectories of graduates. Finally, scholars can replenish their research by incorporating vacancies data from HeadHunter in addition to CVs data and comprehensively approach the analysis of supply and demand sides in the Russian labor market.

## References


1. Cabrera, Marga, and Nuria Lloret, eds. 2017. From Universities to Private Companies: A Measurable Route of LinkedIn Users. Digital Tools for Academic Branding and Self-Promotion. IGI Global.
2. Case, Han, and Rimes. 2016. ‘Career Paths of Computing Program Graduates: A LinkedIn Analysis’. St. Augustine 7.
3. Deming, David, and Lisa B. Kahn. 2017. ‘Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals’. 33.
4. Elexpuru Albizuri, Iciar, and Icy Fresno Anabo. 2017. ‘LinkedIn as a Tool for Higher Education Programme Evaluation’.
5. Gimpleson, Zudina, and Kapeliushnikov. 2020. ‘Некогнитивные Компоненты Человеческого Капитала: Что Говорят Российские Данные’.
6. Hershbein, Brad, and Lisa B. Kahn. 2018. ‘Do Recessions Accelerate -Routine-Biased Technological Change? Evidence from Vacancy Postings’. THE AMERICAN ECONOMIC REVIEW 108(7):36.
7. Li, Lei, Guangzhi Zheng, Svetlana Peltsverger, and Chi Zhang. 2016. ‘Career Trajectory Analysis of Information Technology Alumni: A LinkedIn Perspective’. Pp. 2–6 in Proceedings of the 17th Annual Conference on Information Technology Education. Boston Massachusetts USA: ACM.
8. Melianova, Ekaterina, Suhas Parandekar, and Artëm Volgin. 2020. ‘Returns to Education in the Russian Federation: Towards Evidence Based Decision Making with Fiscal and Private Returns to Education’. 62.
9. Moreno-Delgado, Alicia, Enrique Orduña-Malea, and Rafael Repiso. 2020. ‘LinkedIn as a Data Source to Rank Universities According to Graduate’s Employability in Top Companies’. Transinformação 32:e200023. doi: 10.1590/2318-0889202032e200023.
10. Muller, Noël, and Abla Safir. 2019. What Employers Actually Want. World Bank, Washington, DC.
11. OECD. 2001. The Well-Being of Nations: The Role of Human and Social Capital. OECD.
12. Patrinos, Harry Anthony. 2020. The Learning Challenge in the 21 St Century. World Bank, Washington, DC.
13. Rozhkova. 2019. ‘Отдача От Некогнитивных Характеристик На Российском Рынке Труда’.
14. The World Bank. 2015. ‘Развитие Навыков Для Инновационного Роста в России’.
15. Zhu, Tingting Juni, Alan Fritzler, and Jan Orlowski. 2018. ‘Jobs, Skills and Migration Trends Methodology & Validation Results’. 98.

# Annex

**TABLE A1.1** Coefficients for education level, faculty category and university specialization

| | Teamwork and Leadership | Negotiation | Teaching and Training | Presentation and Public speaking | People Management | Project Management |
|---|---|---|---|---|---|---|
| Education level - Higher | 1.058*** | 1.200*** | 0.962 | 1.439*** | 1.104*** | 1.559*** |
| | (0.018) | (0.022) | (0.027) | (0.037) | (0.023) | (0.037) |
| Education level - Bachelor | 1.197*** | 1.097*** | 0.942 | 1.362*** | 1.038 | 1.443*** |
| | (0.023) | (0.031) | (0.038) | (0.048) | (0.033) | (0.048) |
| Education level - Master | 1.235*** | 1.316*** | 1.135*** | 2.228*** | 1.206*** | 2.307*** |
| | (0.027) | (0.033) | (0.040) | (0.047) | (0.035) | (0.046) |
| Education level - Candidate | 0.966 | 1.356*** | 1.363*** | 3.720*** | 1.359*** | 2.958*** |
| | (0.065) | (0.066) | (0.071) | (0.076) | (0.063) | (0.073) |
| Faculty category - Agricultural science | 1.052 | 0.829*** | 0.997 | 0.889 | 0.845*** | 0.802** |
| | (0.046) | (0.060) | (0.072) | (0.105) | (0.061) | (0.108) |
| Faculty category - Art and Culture | 0.976 | 0.647*** | 1.036 | 0.891 | 0.744*** | 0.935 |
| | (0.043) | (0.056) | (0.062) | (0.080) | (0.060) | (0.075) |
| Faculty category - Education | 1.051 | 0.860*** | 1.179*** | 1.034 | 0.925* | 0.737*** |
| | (0.034) | (0.042) | (0.043) | (0.065) | (0.044) | (0.078) |
| Faculty category - Engineering and Technology | 0.997 | 0.743*** | 0.920*** | 0.749*** | 0.871*** | 0.895*** |
| | (0.017) | (0.021) | (0.026) | (0.038) | (0.021) | (0.036) |
| Faculty category - Health and Medicine | 1.019 | 0.760*** | 0.782*** | 0.981 | 0.787*** | 0.840 |
| | (0.061) | (0.079) | (0.089) | (0.104) | (0.085) | (0.133) |
| Faculty category - Humanities | 1.134*** | 1.118*** | 1.359*** | 1.272*** | 0.939** | 1.189*** |
| | (0.020) | (0.025) | (0.027) | (0.041) | (0.027) | (0.042) |
| Faculty category - Mathematics and Natural science | 1.025 | 0.771*** | 0.997 | 0.918* | 0.796*** | 0.941 |
| | (0.022) | (0.028) | (0.032) | (0.047) | (0.029) | (0.044) |
| University specialization - Agrarian | 1.010 | 0.793*** | 0.871*** | 0.716*** | 0.897*** | 0.629*** |
| | (0.026) | (0.033) | (0.041) | (0.062) | (0.033) | (0.062) |
| University specialization - Art | 1.118*** | 1.029 | 1.050 | 1.032 | 1.025 | 1.108 |
| | (0.038) | (0.046) | (0.053) | (0.075) | (0.049) | (0.070) |
| University specialization - Medical | 0.870** | 0.887 | 0.871 | 1.166 | 0.938 | 0.919 |
| | (0.059) | (0.075) | (0.084) | (0.099) | (0.081) | (0.126) |
| University specialization - Pedagogical | 1.066*** | 0.925*** | 1.214*** | 0.972 | 0.952 | 0.794*** |
| | (0.024) | (0.030) | (0.031) | (0.049) | (0.031) | (0.051) |
| University specialization - Socio-economic | 1.019 | 0.995 | 0.960 | 1.036 | 1.028 | 1.090* |
| | (0.022) | (0.028) | (0.031) | (0.049) | (0.029) | (0.049) |
| University specialization - Specialized | 0.971 | 1.114 | 1.021 | 1.183 | 1.058 | 1.080 |
| | (0.100) | (0.111) | (0.140) | (0.190) | (0.117) | (0.189) |
| University specialization - Technical | 0.860** | 0.723*** | 0.934 | 0.743** | 0.809** | 0.515*** |
| | (0.072) | (0.090) | (0.093) | (0.142) | (0.094) | (0.169) |
| University specialization - Sport | 0.982 | 0.932*** | 0.934*** | 0.930** | 1.016 | 1.006 |
| | (0.016) | (0.020) | (0.024) | (0.035) | (0.021) | (0.034) |

**TABLE A1.2** Coefficients for Education Level, Faculty Category and University Specialization

| | Software for finance, management and analytics | Analytical | Programming | Financial | Social media marketing |
|---|---|---|---|---|---|
| Education level - Higher | $1.141^{***}$ | $1.599^{***}$ | $1.199^{***}$ | $1.247^{***}$ | $1.049^{*}$ |
| | (0.021) | (0.022) | (0.031) | (0.028) | (0.028) |
| Education level - Bachelor | $1.260^{***}$ | $1.589^{***}$ | $1.421^{***}$ | $1.174^{***}$ | $1.180^{***}$ |
| | (0.027) | (0.028) | (0.038) | (0.039) | (0.036) |
| Education level - Master | $1.470^{***}$ | $2.506^{***}$ | $1.858^{***}$ | $1.562^{***}$ | $1.213^{***}$ |
| | (0.030) | (0.029) | (0.040) | (0.039) | (0.040) |
| Education level - Candidate | $0.834^{**}$ | $3.862^{***}$ | $1.862^{***}$ | $1.204^{**}$ | 0.861 |
| | (0.074) | (0.056) | (0.083) | (0.079) | (0.102) |
| Faculty category - Agricultural science | $0.667^{***}$ | $0.852^{**}$ | $0.730^{**}$ | $0.532^{***}$ | 0.942 |
| | (0.062) | (0.064) | (0.134) | (0.078) | (0.085) |
| Faculty category - Art and Culture | $0.472^{***}$ | $0.509^{***}$ | $1.305^{***}$ | $0.465^{***}$ | $0.558^{***}$ |
| | (0.064) | (0.062) | (0.086) | (0.083) | (0.070) |
| Faculty category - Education | $0.691^{***}$ | $0.772^{***}$ | 0.968 | $0.584^{***}$ | $0.846^{***}$ |
| | (0.046) | (0.048) | (0.086) | (0.059) | (0.061) |
| Faculty category - Engineering and Technology | $0.810^{***}$ | $0.836^{***}$ | $1.791^{***}$ | $0.693^{***}$ | $0.797^{***}$ |
| | (0.023) | (0.026) | (0.041) | (0.027) | (0.031) |
| Faculty category - Health and Medicine | $0.576^{***}$ | 0.896 | $0.710^{**}$ | $0.589^{***}$ | 0.879 |
| | (0.098) | (0.079) | (0.160) | (0.116) | (0.114) |
| Faculty category - Humanities | $0.765^{***}$ | 0.995 | $0.887^{**}$ | $0.716^{***}$ | $1.349^{***}$ |
| | (0.028) | (0.030) | (0.054) | (0.034) | (0.032) |
| Faculty category - Mathematics and Natural science | $0.849^{***}$ | $1.122^{***}$ | $2.249^{***}$ | $0.680^{***}$ | 0.954 |
| | (0.029) | (0.031) | (0.048) | (0.036) | (0.038) |
| University specialization - Agrarian | 1.007 | $0.746^{***}$ | $0.534^{***}$ | 1.005 | $0.897^{**}$ |
| | (0.034) | (0.039) | (0.070) | (0.039) | (0.049) |
| University specialization - Art | 0.984 | 0.926 | 0.868 | 0.936 | 1.056 |
| | (0.051) | (0.054) | (0.088) | (0.062) | (0.061) |
| University specialization - Medical | $0.583^{***}$ | $0.820^{***}$ | $0.610^{***}$ | $0.741^{***}$ | 0.858 |
| | (0.093) | (0.076) | (0.149) | (0.108) | (0.109) |
| University specialization - Pedagogical | $0.934^{**}$ | $0.774^{***}$ | $0.857^{***}$ | $0.840^{***}$ | 0.953 |
| | (0.032) | (0.035) | (0.058) | (0.040) | (0.041) |
| University specialization - Socio-economic | 1.048 | $1.105^{***}$ | 1.077 | $1.084^{**}$ | 1.031 |
| | (0.031) | (0.036) | (0.061) | (0.037) | (0.039) |
| University specialization - Specialized | 1.004 | 1.007 | 0.737 | 0.846 | 0.815 |
| | (0.126) | (0.134) | (0.234) | (0.150) | (0.175) |
| University specialization - Technical | $0.682^{***}$ | $0.705^{***}$ | $0.475^{***}$ | 0.881 | $0.693^{***}$ |
| | (0.102) | (0.100) | (0.207) | (0.118) | (0.129) |
| University specialization - Sport | $1.119^{***}$ | 1.038 | 1.020 | $1.104^{***}$ | $0.940^{**}$ |
| | (0.022) | (0.024) | (0.040) | (0.026) | (0.029) |

**TABLE A1.3** Coefficients for Education Level, Faculty Category and University Specialization

| | Medical-psycological | Writing | Law | Software for design, architecture and planning |
|---|---|---|---|---|
| Education level - Higher | 1.448*** | 1.227*** | 2.007*** | 1.447*** |
| | (0.078) | (0.041) | (0.085) | (0.027) |
| Education level - Bachelor | 1.294*** | 1.352*** | 1.777*** | 1.670*** |
| | (0.100) | (0.049) | (0.103) | (0.033) |
| Education level - Master | 1.838*** | 1.587*** | 2.361*** | 2.044*** |
| | (0.104) | (0.053) | (0.101) | (0.035) |
| Education level - Candidate | 1.850*** | 4.901*** | 2.263*** | 1.840*** |
| | (0.177) | (0.086) | (0.165) | (0.082) |
| Faculty category - Agricultural science | 1.959*** | 1.091 | 0.940 | 1.130 |
| | (0.193) | (0.135) | (0.165) | (0.094) |
| Faculty category - Art and Culture | 0.807 | 0.605*** | 0.293*** | 3.705*** |
| | (0.188) | (0.088) | (0.270) | (0.062) |
| Faculty category - Education | 2.318*** | 0.880 | 0.458*** | 0.834** |
| | (0.101) | (0.086) | (0.137) | (0.079) |
| Faculty category - Engineering and Technology | 0.600*** | 0.720*** | 0.499*** | 2.129*** |
| | (0.111) | (0.051) | (0.064) | (0.038) |
| Faculty category - Health and Medicine | 1.364** | 0.972 | 0.438*** | 0.748** |
| | (0.135) | (0.156) | (0.310) | (0.147) |
| Faculty category - Humanities | 0.931 | 2.613*** | 0.650*** | 1.258*** |
| | (0.095) | (0.044) | (0.073) | (0.046) |
| Faculty category - Mathematics and Natural science | 0.836 | 0.967 | 0.486*** | 1.739*** |
| | (0.110) | (0.059) | (0.091) | (0.045) |
| University specialization - Agrarian | 0.546*** | 0.552*** | 0.640*** | 0.647*** |
| | (0.160) | (0.086) | (0.092) | (0.059) |
| University specialization - Art | 0.627** | 1.070 | 0.737** | 1.109 |
| | (0.191) | (0.081) | (0.147) | (0.069) |
| University specialization - Medical | 1.020 | 0.722** | 0.568** | 0.736** |
| | (0.125) | (0.150) | (0.280) | (0.137) |
| University specialization - Pedagogical | 1.363*** | 0.796*** | 0.779*** | 0.889** |
| | (0.094) | (0.058) | (0.085) | (0.052) |
| University specialization - Socio-economic | 0.685*** | 0.949 | 0.971 | 0.957 |
| | (0.118) | (0.059) | (0.069) | (0.056) |
| University specialization - Specialized | 0.00000 | 0.680 | 1.537* | 0.897 |
| | (221.489) | (0.264) | (0.222) | (0.216) |
| University specialization - Technical | 0.915 | 0.342*** | 0.403** | 0.442*** |
| | (0.235) | (0.206) | (0.371) | (0.179) |
| University specialization - Sport | 0.508*** | 0.756*** | 0.849*** | 1.261*** |
| | (0.093) | (0.044) | (0.054) | (0.036) |