
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Calanca, Federica; Sayfullina, Luiza; Minkus, Lara; Wagner, Claudia; Malmi, Eric
Responsible team players wanted

Published in:
EPJ Data Science

DOI:
[10.1140/epjds/s13688-019-0190-z](https://doi.org/10.1140/epjds/s13688-019-0190-z)

Published: 27/04/2019

Document Version
Publisher's PDF, also known as Version of record

Published under the following license:
CC BY

Please cite the original version:
Calanca, F., Sayfullina, L., Minkus, L., Wagner, C., & Malmi, E. (2019). Responsible team players wanted: an analysis of soft skill requirements in job advertisements. *EPJ Data Science*, 8(1), 1-20. Article 13.
<https://doi.org/10.1140/epjds/s13688-019-0190-z>



Responsible team players wanted: an analysis of soft skill requirements in job advertisements

Federica Calanca^{1*} , Luiza Sayfullina², Lara Minkus³ , Claudia Wagner^{4,5} and Eric Malmi²

*Correspondence:

fede.calanca@gmail.com

¹Sapienza University of Rome,
Rome, Italy

Full list of author information is
available at the end of the article

Abstract

During the past decades the importance of soft skills for labour market outcomes has grown substantially. This carries implications for labour market inequality, since previous research shows that soft skills are not valued equally across race and gender. This work explores the role of soft skills in job advertisements by drawing on methods from computational science as well as on theoretical and empirical insights from economics, sociology and psychology. We present a semi-automatic approach based on crowdsourcing and text mining for extracting a list of soft skills. We find that soft skills are a crucial component of job ads, especially of low-paid jobs and jobs in female-dominated professions. Our work shows that soft skills can serve as partial predictors of the gender composition in job categories and that not all soft skills receive equal wage returns at the labour market. Especially “female” skills are frequently associated with wage penalties. Our results expand the growing literature on the association of soft skills on wage inequality and highlight their importance for occupational gender segregation at labour markets.

Keywords: Soft skills; Job advertisement; Text mining; Gender inequality; Crowdsourcing; Computational social science; Labour markets

1 Introduction

When it comes to jobs and careers, technical abilities and professional qualifications are important factors both from the perspective of an employer and of a new employee. However, as pointed out by recent studies [1–3], more and more attention is focused on soft skills, i.e. qualities that do not depend on the acquired knowledge and that are harder to quantify due to being related to one’s emotional intelligence and personality traits. At the same time, they are extremely important because they facilitate human connections [4]. The Oxford dictionary, for instance, defines soft skills as “personal attributes that enable someone to interact effectively and harmoniously with other people”^a. During the period of 1980 and 2012, jobs with high social skills requirements grew by around 10% as a share of the US labour force [5]. The increasing importance of soft skills at labor markets stems from the growth of the service sector, where interpersonal services are sold, as well as from the introduction of lean-manufacturing, where an integrated skill set, comprised of both hard and soft skills, has gained importance [6, 7]. Observational studies have also shown

that social features potentially related to soft skills (e.g. the variety of friendship connections and position diversity within a community) are positively correlated with economic outputs [8, 9].

The growing importance of soft skills also carries implications for gender inequality in labour markets. Research has shown that certain societal groups are perceived as lacking important soft skills, i.e. evidence was found that black men are characterized as being less motivated than their white counterparts [10]. Additionally, not all types of soft skills are valued equally, e.g. based on gender stereotypes and beliefs about women's inferior status in the workplace, skills that are perceived as "female" are found to be associated with wage penalties [11–13]. On the other hand, recent scholarly debates engage in the discussion of a possible female advantage associated with the rising importance of people skills in contemporary labor markets [2, 14–16].

Despite the growing importance of soft skills and their potential contributions to inequalities in labour markets, to date, we know surprisingly little about the role of "gendered soft skills"—i.e., soft skills that are stereotypically associated with one gender—in the job market [15, 17]. Most prior scientific articles referring to skills and labor market outcomes construct indices of soft skills in which male and female connoted skills get added up, rather than making a distinction between them (see, for instance, [2, 14]). This approach is useful, because the overall increasing importance of soft skills in contemporary labor markets [16] can be measured in an easy-to-grasp, single-index way. However, this coarse-grained measure can mask important differences in labor market outcomes with regard to gendered soft skills. We go beyond this relatively crude measure by introducing a semi-automatic approach for constructing an extensive list of soft skills from job advertisements, which we can use for soft skills detection. Combining this data on soft skills with what prior research has identified as commonly shared gender stereotypes (see, for instance, [18–20]) and official statistics about the proportion of women in various professional fields, allows us to differentiate soft skills depending on their gender connotation. Thus we are able to establish new insights on the association of soft skills related to gender stereotypes and wages.

Additionally, we present evidence on the impact of soft skills on sex segregation in labor markets. Although the existing literature on supply-side mechanisms of occupational sorting, i.e. women making career choices based on potentially biased self-assessed beliefs about interests and capacities, is growing [21, 22], the demand-side process, meaning the allocation of men and women into sex-typed occupations by employers, remains relatively understudied [17]. There is only a limited number of studies examining the influence of gendered wording on occupational choices. These studies use small-scale experiments and thus cover only a limited range of soft skills associated with gender stereotypes [19, 23–26]. Utilizing our newly extracted dataset based on real job advertisements, we are able to examine the impact of soft skills in general and gendered soft skills in specific on occupational segregation.

Based on our unique dataset on soft skills in job ads, we find evidence that female connoted soft skills are associated with wage penalties, while soft skills perceived as being stereotypically male are linked to wage premiums. Our results show further that women are more likely to be found in occupations that are advertised using soft skills associated with female stereotypes and vice versa for men.

This article is structured as follows: in Sect. 2, we present our methodology for extracting soft skill mentions from a large corpus of job advertisements. In Sect. 3, we scrutinize wage premiums and penalties associated with soft skills frequently mentioned in job ads based on a matching study. Next, the role of soft skills in reproducing gender segregation, i.e. the unequal distribution of men and women across occupations, is examined in Sect. 4. Finally, we present conclusions in Sect. 5 with a summary of our findings, their implications, limitations, and suggestions for future work.

2 Methods and data

In this section, we describe the datasets used in this work and our semi-automatic soft skill mining approach. Following this approach we first create clusters of soft skills, grouping similar soft skills together, and then detect soft skills in job ads by searching for the soft skill strings in job descriptions.

2.1 Data

Our analysis is based on a dataset containing 245,000 job advertisements (ads) from the United Kingdom (UK).^b This data is provided by the Adzuna job search engine, which collects job ads from hundreds of different websites. Each job ad entry contains the *title*, *full description*, *job category*, and *salary* of the job, among five other types of fields.^c

Adzuna has classified the ads into 29 job categories, based on the source of the ad and the job's description. Table 1 illustrates the most distinctive soft skills for five selected job categories. Desired soft skills differ considerably depending on the job category. For instance, the three most distinctive skills for *Teaching* are *enthusiastic*, *dedicated*, *professional*, whereas for *Accounting & Finance* they are *accurate*, *responsible*, *analytical abilities*. The soft skill detection algorithm is described in Sect. 2.2.4.

All experiments in this paper are conducted using the UK dataset, except for a crowdsourcing experiment needed for collecting an initial list of soft skills, which is described in the next Section 2.2.1. For this crowd-sourcing experiment, a dataset posted by the Armenian human resource portal CareerCenter consisting of 19,000 online job postings in a period from 2004–2015 is more appropriate, because job requirements are listed in a separate field. Thus the workers do not need to read through the full ad, allowing us to annotate more ads and to collect a longer list of soft skills.^d

2.2 Soft skill mining

Our semi-automatic soft skill mining approach consists of the following steps: first, crowdworkers generate an initial set of potential soft skills, second, skills that seldom refer to candidates are removed, third, soft skills with a similar meaning are clustered into groups of skills, and fourth, soft skills are detected in new ads. These steps are summarized in Fig. 1 and explained in more detail in the following sections.

The resulting soft skills and their clusters are available at <http://dx.doi.org/10.7802/1707>.

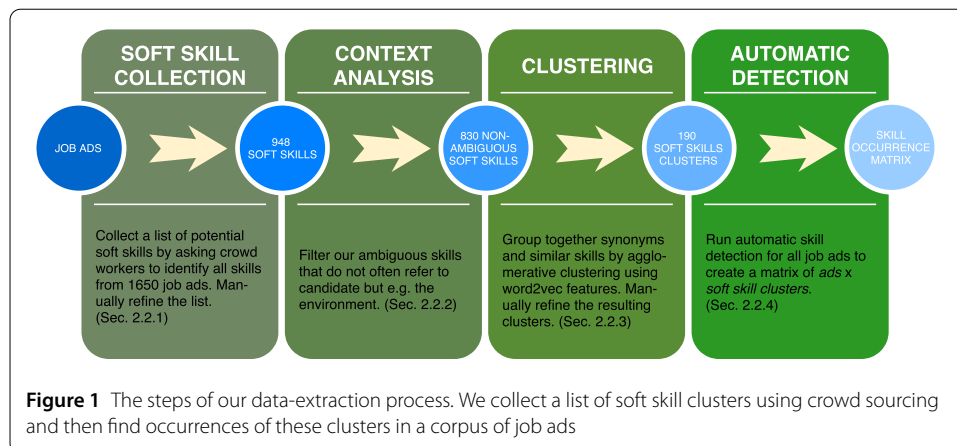
2.2.1 Crowdsourcing a list of soft skills

The collection of soft skills was done through Figure Eight (formerly known as Crowd-Flower),^e a crowdsourcing platform that allowed us to speed up our data collection process by submitting annotation tasks to online crowdworkers.

Table 1 The most distinctive soft skills for five job categories

Social work	$\Delta\%$	%	Accounting & Finance	$\Delta\%$	%	IT	$\Delta\%$	%	Teaching	$\Delta\%$	%	Creative & Design	$\Delta\%$	%
team player	+7.3	22.7	accurate	+7.5	14.1	problem solving	+4.6	8.9	enthusiastic	+12.1	20.2	creative	+24.8	30.3
ability to work with children	+6.6	7.0	responsible	+6.0	34.7	communication skills	+3.5	27.8	creative	+5.9	11.5	innovative	+5.3	11.2
positive	+4.3	9.8	communication skills	+4.6	28.9	innovative	+3.1	8.9	positive	+5.9	11.4	attention to detail	+4.9	9.8
flexible	+1.5	13.4	analytical skills	+3.2	5.9	team player	+2.4	17.8	leadership	+5.0	11.4	management skills	+4.4	14.2
leadership	+1.5	7.9	attention to detail	+2.9	7.8	analytical skills	+2.1	4.8	confident	+4.2	11.1	responsible	+3.6	32.4
patience	+0.8	0.9	ability to work within deadlines	+2.1	4.7	management skills	+1.8	11.6	hard working	+3.4	6.3	confident	+3.0	9.8
people skills	+0.8	2.3	interpersonal skills	+1.5	6.0	creative	+1.5	7.0	innovative	+3.1	8.9	presentation skills	+1.6	3.4

Distinctiveness ($\Delta\%$) is defined as the absolute difference between the percentage of job ads that contain the skill within the given category (%) and the percentage of job ads that contain the skill within all categories.



First, each worker was given the following definition of soft skills:

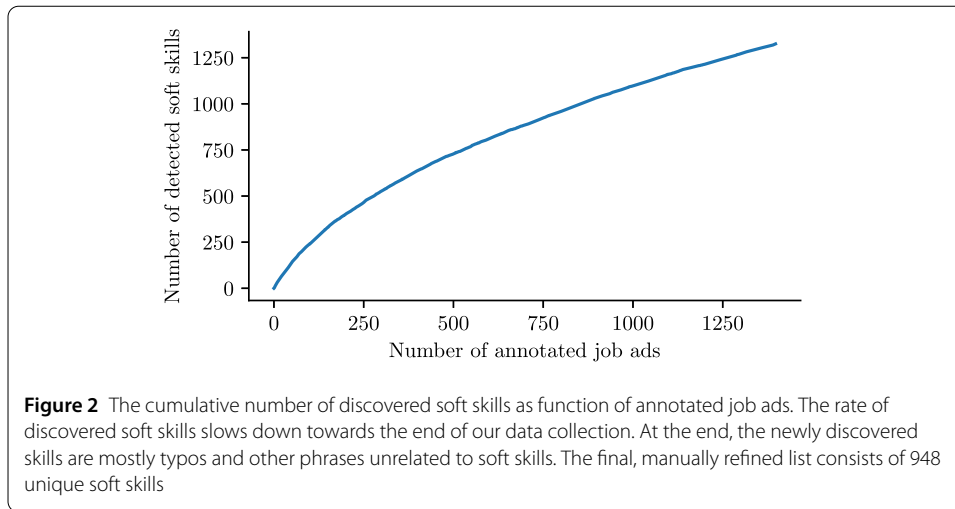
In a nutshell soft skills can be identified as qualities that do not depend on acquired knowledge; they complement hard skills (also known as technical skills). According to Wikipedia soft skills “are a combination of interpersonal people skills, social skills, communication skills, character traits, attitudes, [...] social intelligence and emotional intelligence quotients”.

This was followed by a list of soft skill examples and instructions for completing the tasks. In particular, the workers were instructed to read the presented text, consisting of the “job description” and “required qualifications” fields, select whether the text contained any soft skills, and, if that was the case, they were instructed to copy and paste the smallest relevant part of text denoting each skill to an answer field. Additionally, the workers were instructed to remove unnecessary adjectives and complements, but not to alter the text in any other way. For instance, *excellent communication skills with customers and partners* had to be reported as *communication skills*.

Before the actual annotation phase, the workers were supposed to pass a training phase and answer a set of test questions, for which we had provided the correct answers: they had to obtain an accuracy level of at least 60% to proceed further. These test questions also showed up randomly during the actual annotation phase to ensure that the minimum accuracy level of 60% was maintained.

In total, we annotated 1650 job ads by at least 3 different workers. The annotation effort was conducted in two batches. After both batches we computed the number of distinct soft skills as a function of the number of annotated ads, plotted in Fig. 2. The results show that the rate at which new soft skills are discovered slows down, although new skills were still found at the end of the data collection. However, when examining the skills found last, most of them turned out to be typos and other phrases unrelated to soft skills (these include “ability to work as a part of PSD team”, which is a hard skill since PSD stands for *personal security detail*, and “unquestioned behaviour”, which is highly ambiguous). Therefore, we decided to stop the annotation task after the second batch.

To remove the typos as well as recurrent superfluous adjectives,^f results were cleaned using a script. The script removed additionally extra whitespace and punctuation, and it corrected simple typos and misspellings by comparing the detected skill tokens to a



whitelist of valid skill tokens. Thereafter, we manually reviewed the skills to remove all non-soft skills and to prune out tokens not relevant to the skill.

The final manually curated collection included 948 unique soft skills.

2.2.2 Removing ambiguous soft skills

The focus of this work is to analyze soft skill requirements for job applicants. However, often soft skill phrases in job ads do not refer to the required applicant characteristics, but they may also describe the working environment or something else. For instance, *independent* could be used to describe an “*independent business*” or a *home care assistant* might be required to “help people to remain *independent* in their own homes.” Therefore, it is crucial to be able to detect soft skills that refer to the candidate rather than something else.

To tackle this problem, we created another crowdsourcing task, instructing crowdworkers to annotate soft skill phrases in the context they appear, i.e. the job ads. We noticed that skills consisting of multiple tokens usually unambiguously refer to the candidate and therefore we only annotated the skills consisting of at most three words, that is, 582 out of the 948 skills found in the previous steps.

More specifically, for each one of these skills, we extracted 10 randomly sampled text snippets where the skill occurs, including 25 words before and after the skill. Then we asked crowdworkers to classify each snippet to one of the following three categories: *Candidate*, *Company/Company environment*, or *Other*. At least three answers were recorded for each text snippet.

Based on the annotations, we computed the following confidence score⁸ for each soft skill

$$\text{Conf}(s) = \frac{\sum_{w \in W_c(s)} T(w)}{\sum_{w \in W(s)} T(w)},$$

where $W_c(s)$ denotes the workers who classified an occurrence of skill s to refer to a candidate, $W(s)$ denotes the workers who assessed an occurrence of skill s , and $T(w)$ is the trust of a worker w . Trust is calculated by the crowdsourcing platform as the contributor’s accuracy level in the current job, determined by his/her accuracy during the training

phase—as explained in Sect. 2.2.1. Thus, the confidence score measures the proportion of votes for the *Candidate* category weighted by the trusts' of the workers who gave the votes.

We included the skills with a confidence value of at least 0.7 into the final list of soft skills. This value allowed us to retain 81.3% of the annotated skills (8.3% of trigram, 10.3% of bigram and 40.1% of single-word skills were discarded) while still having a relatively high confidence that the retained soft skill phrases actually refer to the candidate.

2.2.3 Soft skill clustering

Many of the soft skills collected by the crowdworkers are synonyms or near-synonyms. The different versions of a skill result, e.g., from diverse ways of expressing the concept (team-worker, ability to work in a team), or from slightly different spellings (able to work in team). To unify the different variants, the collected soft skills were clustered by first employing an algorithmic approach and then refining the clusters manually. After experimenting with a small subset of soft skills, different algorithms and parameter settings, we decided upon the following procedure.

Each soft skill was first represented in the vector space by averaging the word2vec [27] embeddings of its tokens, excluding stopwords. We used 300-dimensional embeddings pre-trained on the GoogleNews dataset.^h Then, we employed agglomerative clustering algorithm to cluster the embedding vectors using the *average linkage* cosine distance measure. The clusters were finally reviewed and manually improved by split and merge operations and by reassigning some of the skills to more appropriate clusters, obtaining a final list of 190 clusters.ⁱ

2.2.4 Soft skill detection

In the final phase, our goal was to detect skill clusters in each job ad.

First, we preprocessed the job descriptions and the list of soft skills by lowercasing and removing stop words.^j We also removed the competence terms (able, skills, etc.) from most soft skills, if they were perceived as not being fundamental for skill identification, to avoid false negatives (e.g. capable of handling multiple tasks should match with abilities in handling multiple tasks). Still, for some skills, we kept the competence terms if they would have become too ambiguous, resulting in false positive detection (e.g. communication skills without the word skills would match with communication technologies).

Thereafter, we searched for each soft skill *s* in each job description. If *s* consisted of multiple tokens, we allowed for at most two extra words to occur before each token in addition to stop-words, that were allowed to be removed from certain skills without making them ambiguous. We also experimented with more liberal ways of matching skills, ignoring the word order of the skill tokens or lemmatizing the tokens, but these were found to decrease the precision of the detected skills significantly.

Soft skills were detected in 78% of the ads, with 45.5% mentioning at least 3 soft skills, attesting to the importance of soft skills in the labour market.

2.2.5 Related work on soft skill mining

The curation of *hard* skills has been addressed by LinkedIn [28], whereas Kivimäki et al. [29] proposed a system for automatic detection of new skills in free written text using a

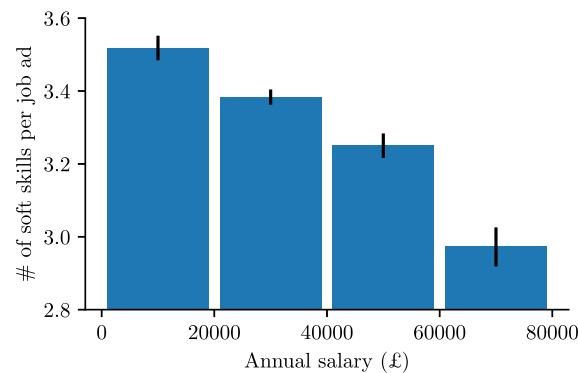


Figure 3 Low-paid job ads contain, on average, more soft skills than high-paid job ads. Bars show the average number of soft skills for the ads in four different salary groups, and the error bars indicate the 95% confidence intervals obtained via bootstrap re-sampling with replacement

spread-activation algorithm. Recently, Haranko et al. [30] suggested a novel approach for collecting data on skills and gender imbalances through LinkedIn's advertising platform. Automatic classification of soft skills referring to a candidate vs. something else (e.g. the work environment), has been studied by Sayfullina et al. [31], using the crowdsourced data collected in this work as described in Sect. 2.2.2.

3 Salary and soft skills

One of our main research questions is how the presence of certain soft skills may affect wages.

Analyzing annual salaries of job ads, we found that low-paid job ads contain, on average, more soft skills than high-paid job ads. This is illustrated in Fig. 3 which shows the average number of soft skill mentions per job ad in four different salary groups. The ads with a salary (s) of $s \leq £20,000$ have 3.52 soft skills on average, whereas ads with a salary of $£60,000 < s \leq £80,000$ have only 2.97 soft skills on average. All paired differences between the salary groups are statistically significant ($p < 0.001$; two-tailed t -tests with unequal variances).

While the higher prevalence of soft skills in low paid jobs is interesting by itself, it does not reveal which soft skills tend to be associated with wage premiums and which ones with wage penalties. To address this question we conduct a matching study.

3.1 Matching study

In order to study the link between a job ad's soft skill requirements and their respective salary,^k we conduct a matching study [32]. The benefit of matching is that, in pairing a treated job ad (i.e. an ad with a given job title and job category that contain a specific skill) with its counterfactual (i.e. an ad with the same title and category but without the specific skill), we can control for a range of unobserved job category characteristics [33]. These characteristics include, for instance, work experience, since job titles often include qualifiers, such as *head*, *senior*, *junior*, or *intern*.

The specific matching strategy applied in this article is as follows: first, we group ads having the same job category c and job title t , ignoring stop words and the word order of

the title. We picked all titles occurring at least twice, resulting in 34,071 distinct titles and 158,658 ads. Given a soft skill s , a normalized salary reward is defined as

$$r_{s,c,t} = \frac{M_{s,c,t} - \bar{M}_{s,c,t}}{\bar{M}_{s,c,t}} \times 100\%, \quad (1)$$

where $M_{s,c,t}$ and $\bar{M}_{s,c,t}$ are the average salaries of job ads belonging to job category c , having job title t , and containing or not containing skill s , respectively.

For example, in our dataset there are 210 “Java Developer” job ads in the IT Jobs category out of which 28 contain the soft skill `communication skills`. The average salary of these 28 positions is £46,536 per year, whereas the average salary for the other 182 positions is £43,170 per year. This means that the salary reward for `communication skills` in Java Developer / IT Jobs category is

$$\begin{aligned} & r_{\text{communication skills, IT Jobs, Java Developer}} \\ &= \frac{£46,536 - £43,170}{£43,170} \times 100\% \\ &= 7.8\%, \end{aligned}$$

suggesting that Java developer positions that require communication skills usually pay 7.8% more than other Java developer positions.

Given the individual salary rewards, the overall salary reward r_s of soft skill s is obtained by averaging the rewards over all possible job titles and categories

$$r_s = \frac{\sum_c \sum_t r_{s,c,t} \min(C_{s,c,t}, \bar{C}_{s,c,t})}{\sum_c \sum_t \min(C_{s,c,t}, \bar{C}_{s,c,t})}, \quad (2)$$

where $C_{s,c,t}$ and $\bar{C}_{s,c,t}$ are the number of job ads belonging to job category c , having job title t , and containing or not containing skill s , respectively. Individual rewards are weighted by the number of ads to avoid letting infrequent job titles have disproportionately large effect on the overall reward. In most cases, $\min(C_{s,c,t}, \bar{C}_{s,c,t}) = C_{s,c,t}$ since typically less than half of the ads from any category contain a given soft skill. Thus, the individual rewards are typically weighted by the number of ads containing the skill.

A positive reward r_s indicates that job ads that mention skill s have on average a higher salary than other job ads from the same job category and the same job title that do not mention s .

To compute the statistical significance of an observed reward value, r^{obs} , we conduct a permutation test as follows: each job ad consists of (i) a set of soft skills mentioned in the job description, (ii) job category and title, and (iii) salary. We shuffle the soft skill sets (i) between the ads and keep everything else ((ii) and (iii)) fixed. This shuffling is repeated 1000 times and after each shuffle, we compute a new reward r^{rand} . The p -value for the null hypothesis that $|r^{\text{obs}}| \leq |r^{\text{rand}}|$ is given simply by the fraction of $|r^{\text{rand}}|$ values that are greater than or equal to $|r^{\text{obs}}|$. If the fraction is below or equal to a threshold of $\alpha = 0.05$, we conclude that r^{obs} is statistically significant and mark the reward with a ‘*’. A reward with $p \leq 0.01$ is marked by ‘**’.

3.2 Results

The soft skills that are associated with the highest wage premiums or penalties are shown in Table 2. Most of the soft skills associated with wage premiums can also be considered a requirement for higher occupational positions. Soft skills such as delegation skills, team building skills and leadership imply that a certain kind of supervision and authority toward others is required [34]. In contrast, listening skills, willingness to learn, as well as being punctual, describe skills that entail a certain degree of subordination.

Our empirical observation that soft skills associated with wage premiums are also closely tied to leadership positions is in accordance with sociological occupational class theories. Previous research on occupational classes has identified the magnitude of a job's authority as one of the key determinants in assessing the job's position in the occupational class system [35, 36]. Jobs that entail a high degree of authority also occupy a strategic position in the labour market: by monitoring their subordinates, employees in leadership positions are ensuring that a firm produces surplus. Given this powerful position, high degrees of authority entail a significant degree of bargaining power and thereby the possibility to demand higher than average wages [36]. Empirical research indeed supports this notion and shows that leadership skills are associated with wage premiums [37, 38].

Additional supporting evidence for this particular reading of the results comes from psychology. We find that character traits associated with wage premiums, for instance delegation skills, team building skills, and strategic planning are closely connected to skills psychological research has identified as leadership character-

Table 2 Skills with the highest and the lowest overall salary rewards (r from Eq. (2))

Skill cluster	r	Count
maturity	11.9**	112
delegation skills	10.2**	53
team building skills	9.8*	50
strategic planning	9.1**	608
ability to work in a fastpaced environment	8.0*	51
leadership	7.4**	4743
constructive feedback	6.9*	74
proposal writing	6.2*	84
ability to improve skills	6.0**	108
discretion	5.7	309
results driven	4.9**	541
presentation skills	4.5**	1464
telephone skills	-7.3**	227
polite	-5.9**	339
dynamic person	-5.2	70
dedication	-4.6**	467
friendly personality	-4.6	97
listening skills	-4.3**	355
punctual	-4.1*	248
ability to identify problems	-3.1	132
calm	-2.8*	787
professional manner	-2.6**	2303
willingness to learn	-2.2**	1652
time management	-1.8	2149

** $p < 0.01$, * $p < 0.05$.

The *Count* column shows the denominator from Eq. (2), which can roughly be interpreted as the sample size. Only skill clusters with *Count* ≥ 50 are shown.

istics, i.e. management of personnel, visioning, as well as general strategic skills [39].

What is striking, is that many of the aforementioned skills in Table 2 also correspond to gender stereotypes. Gender stereotypes are generalizations about commonly shared perceptions of female and male attributes. Previous research has shown that while women are described as embodying “communal behavior”, such as kindness, loyalty, and warmth, men are characterized by “agentic traits”, such as competitiveness and aggressiveness [20], and as possessing leadership abilities [18]. Common “agentic” traits, such as competitive and aggressive, have been filtered out as ambiguous (see Sect. 2.2.2), since they typically do not describe the desired characteristics of the job applicant. However, we still find several leadership traits to come about with higher wages in Table 2. Moreover, “communal behavior” seems to come about with wage penalties in Table 2 across the board (for instance: polite, dedication, friendly personality, and being calm).

Thus, Table 2 provides first evidence that male gender stereotypes are connected to wage premiums, whereas female gender stereotypes are connected to wage penalties in the labor market. To scrutinize this issue further, in the following section we examine the association between gender stereotypes and wages in more detail.

4 Gender and soft skills

In this section we scrutinize to what extent soft skills are associated with occupational sex segregation. Thereafter, we explore a possible relationship between wages and gendered soft skills.

4.1 Industry gender composition prediction

In what follows, we test whether soft skills can predict the gender composition of a job category. The proportion of women for each job category was approximated by mapping the job categories in our data to the nearest categories from UK Labour Market statistics¹ as shown in Table 3.

We find that job ads in male-dominated job categories mention 3.20 soft skills on average, while ads in female-dominated job categories mention only 3.00 soft skills. The difference in means is statistically significant ($p < 0.001$; two-tailed t -test with unequal variances).

To predict the proportion of women in the category of a job ad, we used ordinary least squares (OLS) regression over job ads containing at least 3 different soft skills.

Table 4 shows the soft skill clusters that are most predictive of female-dominated jobs (positive coefficients) and of male-dominated jobs (negative coefficients). Only those skill clusters that occurred more than 50 times and whose coefficient is statistically significant ($p < 0.01$) are shown. The table also indicates whether the reward associated with a soft skill is significant or not. The model obtained an R^2 score of 0.11.

A high proportion of women in a job category is associated with soft skills such as empathy, respectful, sensitivity and dedication. Skills such as marketing skills, ability to win new business, ability to lead project teams and analytical skills are negatively associated with women's shares in job categories, meaning they predict soft skills mentioned more frequently in ads for male-dominated jobs. These results illustrate that with a few exceptions (e.g. delegation

Table 3 The percentage of women in job categories

Job category	ONS category	% of women
Social work Jobs	Human health & social work activities	80.62
Healthcare & Nursing Jobs	Human health & social work activities	80.62
Charity & Voluntary Jobs	Human health & social work activities	80.62
Teaching Jobs	Education	71.5
Property Jobs	Real estate activities	57.6
Legal Jobs	Public admin & defence; social security	56.02
Creative & Design Jobs	Other	53.29
Travel Jobs	Other	53.29
Other/General Jobs	Other	53.29
Domestic help & Cleaning Jobs	Accommodation and food services	53.18
Hospitality & Catering Jobs	Accommodation and food services	53.18
Maintenance Jobs	Wholesale, retail & repair of motor vehicles	48.23
Sales Jobs	Wholesale, retail & repair of motor vehicles	48.23
Retail Jobs	Wholesale, retail & repair of motor vehicles	48.23
Accounting & Finance Jobs	Financial & insurance activities	46.45
IT Jobs	Professional, scientific & technical activities	45.72
Engineering Jobs	Professional, scientific & technical activities	45.72
Scientific & QA Jobs	Professional, scientific & technical activities	45.72
HR & Recruitment Jobs	Administrative & support services	44.4
Customer Services Jobs	Administrative & support services	44.4
Admin Jobs	Administrative & support services	44.4
PR, Advertising & Marketing Jobs	Information & communication	29.77
Consultancy Jobs	Information & communication	29.77
Manufacturing Jobs	Manufacturing	24.72
Logistics & Warehouse Jobs	Transport & storage	23.93
Energy, Oil & Gas Jobs	Mining, energy and water supply	21.76
Trade & Construction Jobs	Construction	19.21
Graduate Jobs	–	N/A
Part time Jobs	–	N/A

Data from the UK Office for National Statistics (ONS), according to employment and labour market statistics (2018).

skills and managerial skills), the soft skills that are predictive of the job's gender composition are also closely associated to gender stereotypes.

Thus, not only do skills associated with gender stereotypes about women potentially get lower rewards in labor markets (as suggested by Table 2), but we further find that some soft skills, which are distinctive of the gender composition within a job, are also stereotyped as being female. Put differently, not only does one potentially get paid less if one is carrying out tasks connoted as being female, but occupations carried out mainly by women are also advertised making use of those skills that come about with wage penalties.

Our findings also suggest that there are two deviations from this pattern, i.e. delegation skills and managerial skills, which are soft skills that are associated with leadership (male) stereotypes but still predict a high proportion of women in an occupation. This finding, however, is in line with previous research, providing evidence that women will apply for leadership positions if the remaining part of the job ad is phrased using female stereotypes or gender neutral language [19, 23, 24].

4.2 Occupational segregation and gender-stereotypical soft skills

To more systematically analyze the claim that the gender composition of an occupation is shaped by gender stereotypes, we mapped our soft skill clusters to a list of twenty personality characteristics desired in men and another twenty characteristics desired in women—the so-called Bem Sex Role Inventory [18]. Out of these, we were able to map five femi-

Table 4 OLS regression results predicting the proportion of women using soft skill clusters as predictors

Skill cluster	Coefficient	<i>r</i>	Count
ability to work with children	0.192	0.3	370
delegation skills	0.095	10.2**	117
respectful	0.090	-0.3	254
managerial skills	0.083	3.0	146
reasoning skills	0.072	-10.6**	91
empathy	0.067	-1.3	576
ability to maintain confidentiality	0.059	-0.7	290
sensitivity	0.056	3.0	208
ability to adapt	0.046	4.2**	635
dedication	0.042	-4.6**	923
flexible with hours	0.040	-0.3	1864
attentive	0.039	2.2	124
marketing skills	-0.046	-0.7	337
client skills	-0.045	3.2**	975
ability to win new business	-0.040	2.2	269
ability to lead project teams	-0.037	3.7	200
curious	-0.036	4.1	149
diligent	-0.035	-0.8	241
ability to present ideas	-0.034	2.7	149
courteous	-0.030	0.6	450
methodical	-0.028	-1.1	1035
attention to detail	-0.028	-1.1	7191
self starter	-0.026	3.0*	904
analytical skills	-0.026	2.9**	3972

** $p < 0.01$, * $p < 0.05$.

The first twelve soft skills are the strongest predictors for female oriented job ads (i.e. job ads for professions with a high proportion of women), while the last twelve rows correspond to the strongest predictors for male oriented ads (i.e. job ads for professions with a low proportion of women). Many of the found predictors correspond to common gender stereotypes. The third column lists the salary reward (r , see Eq. (2)), whilst the fourth shows the number of samples from the training set in which the skill clusters appear.

nine and seven masculine characteristics to similar soft skill clusters in our data, shown in Table 5.^m Based on the mappings, we set out to study the prevalence of the gender-stereotypical soft skills in job ads of female and male-dominated industries. The percentage of ads containing a skill within the ads from female- (male-) dominated industries is denoted by P_f (P_m). In the last column of Table 5 we show the percentage difference between these two percentages. A positive value means that the skill is used more in female-dominated industries and a negative value that it is used more in male-dominated industries.

All feminine skills are more prevalent in female-dominated industries, whereas for masculine skills the picture is not as clear. For instance, `analytical skill` is used more than five times more often in male-dominated industries, while `leadership` is used almost twice as often in female-dominated industries, although both of these skills are stereotypically masculine according to Bem [18]. This finding, however, is in agreement with previous research, where evidence was found that although women will make inroads into occupations in which the skill set is in line with typically male features, this is not true the other way around [17, 40]. Hence, although women try to push into male-dominated occupations, men do not do the same with regard to female-dominated occupations.

Our findings have implications for occupational sex segregation, that is, the unequal distribution of men and women across occupations in the labour market. Advertising female or male-dominated jobs in accordance with the associated gender stereotypes reproduces

Table 5 OLS regression results predicting the proportion of women using soft skill clusters as predictors

	Gender stereotype (Bem [18])	Mapped skill cluster	r (%)	P_f (%)	P_m (%)	$\frac{P_f - P_m}{\max(P_f, P_m)} \times 100\%$
<i>Feminine</i>	Compassionate	empathy	-1.3	0.94	0.12	87.1
	Does not use harsh language	polite	-5.9**	0.25	0.22	13.1
	Loves children	ability to work with children	0.3	2.13	0.07	96.8
	Sensitive to the needs of others	sensitivity	3.0	0.22	0.10	52.5
	Warm	friendly personality	-4.6	0.11	0.07	38.4
	Average		-1.7	0.73	0.12	57.6
<i>Masculine</i>	Ambitious	ambitious	1.4	3.11	5.17	-39.9
	Analytical	analytical skills	2.9**	0.59	3.16	-81.3
	Assertive	confident	0.5	6.39	6.09	4.7
	Has leadership abilities	leadership	7.4**	9.85	5.94	39.7
	Independent	capability to work independently	1.9	1.17	1.11	5.4
	Makes decisions easily	make decisions	3.0**	1.25	1.08	13.1
	Self-sufficient	autonomy	1.3	0.99	1.23	-19.4
	Average		2.6**	3.34	3.40	-11.1

** $p < 0.01$, * $p < 0.05$.

The gender stereotypes listed by Bem [18] that could be mapped to one of our soft skill clusters. On average, the feminine stereotypes are associated with a wage penalty ($r = -1.7$), whereas the masculine stereotypes are associated with a premium ($r = 2.6$). The percentage of job ads within female and male-dominated industries that mention a skill cluster are denoted by P_f and P_m , respectively.

cultural beliefs about these stereotypes and upholds the gender-typicality of occupations. Previous research has shown that cultural beliefs about gender stereotypes influence self-assessment of men and women [22, 41]. These biased self-assessments have been shown to be a crucial factor of career choices [22]. Accordingly, empirical evidence employing experiments, suggest that if jobs are advertised using stereotypically male traits, women are less likely to think that they are suitable for the position [25] and, hence, hesitate to apply. Thus, by illustrating that real jobs advertisements that include female stereotypes are dominated by women, we provide large-scale evidence that job ads can be seen as part of a leaky pipeline [42], serving as the first sorting mechanism by which women are crowded out of male-dominated occupations at labor markets [19, 23, 25, 26].

The results thereby suggest the importance of gender stereotypes in the reproduction of occupational segregation, i.e. the demand-side, and the corresponding selection of men and women in different occupations.

However, it is important to note that while our results establish a correlation between the usage of stereotypical soft skills and occupational segregation, studying the causal mechanisms between the two is beyond the scope of this paper. Nevertheless, this work supplements the much richer account of research examining the supply side of the unequal distribution of men and women across occupations, namely the influence of gendered individual preferences and respective assessments of one's own skills and capacities [21, 22], by showing a connection between the demand-side, i.e. job ads, and occupational segregation.

4.3 Gendered soft skills and salary

Results in the previous section illustrated that soft skills corresponding to gender stereotypes are associated with the gender composition of the job category. In what follows, we are going to examine to what extent these gendered soft skills are associated with wage premiums or penalties.

Gender stereotypes may influence wages. More specifically, tasks that are linked to typically “female” responsibilities are often associated with wage penalties [43–45]. An explanation for the devaluation of “female” tasks is found in the ascribed lower status of women, i.e. gender status beliefs. Gender status beliefs are diffuse cultural beliefs on account of which men are rated more competent than women. These beliefs about women’s lack in aptitude and competence are transferred to the labor market and thereby facilitate a devaluation of women and typically “female” tasks in the workplace [11]. Recent evidence, for instance, suggests that women are underrepresented in academic fields where practitioners believe that raw talent is needed in order to succeed. Women are simply seen as less brilliant than men and therefore not hired in academic segments where beliefs about the need for innate talent are salient [46].

The rewards in Table 4 illustrate that soft skills that correspond to gender stereotypes about women, such as *respectful*, *empathy* and *dedication* are predominantly associated with wage penalties (with the exception of *sensitivity*). A similar pattern is found in Table 2, where most of the soft skills related to stereotypes about women are associated with wage penalties, while the ones linked to leadership bring about wage premiums. Hence, our study presents evidence on the devaluation of soft skills related to gender stereotypes based on a large-scale list of soft skills derived from real job ads. We thereby confirm previous small-scale research, in which evidence was found that, net of individual labour-market-relevant characteristics such as work experience, single tasks tied to female gender stereotypes (such as *nurturing* [43]) are associated with wage penalties [44, 45].

Regarding male-dominated jobs, our results show that soft skills that are associated with commonly shared stereotypes about men, such as *analytical skill* and *self starter* [19], predict statistically significant wage premiums. Moreover, Table 4 illustrates that leadership skills, which are also stereotypically ascribed to men, do come with wage premiums (i.e., *ability to win new business*, *ability to lead project teams*, and *ability to present ideas*). However, we find that leadership skills associated with female-dominated occupations such as *delegation skills*, and *managerial skills* are related to wage premiums as well. This means that soft skills that are associated with a high share of women in an occupation are also more often related to wage penalties compared to soft skills that are associated with a high percentage of male incumbents. However, if soft skills required in female-dominated occupations represent leadership skills they can also entail wage premiums.

To further explore the association between sex-typed gender stereotypes and wage penalties or premiums, we calculated the salary rewards r of the soft skills clusters that we found congruent with the personality traits from the Sex Role inventory by Bem [18]. The rewards are listed in Table 5. We find that all masculine skills are associated with a positive reward, whereas 3/5 feminine skills are associated with a penalty. The average rewards for masculine and feminine skills are 2.6 and −1.7, respectively. This difference is statistically significant (one-tailed t -test with equal variances; $p = 0.014$). This suggests

that stereotypically masculine character traits are valued more in the workplace than feminine character traits.

Based on the evidence provided we find that the devaluation of women is mainly realized via gender stereotypes, while skills associated with male stereotypes, i.e. leadership skills, do receive wage premiums.

5 Discussion and conclusions

This study examined soft skills in the labour market and showed that soft skills are a crucial component of job ads, especially of low-paid jobs and male-dominated professions and may therefore potentially perpetuate labour market inequalities. To explore how soft skills influence labor market outcomes, in particular wage premiums or penalties and gendered labour market composition, we developed a semi-automatic approach for mining soft skills from job advertisements.

We would like to highlight three key findings of our study:

1. We found that not all soft skills are valued equally in the labour market, some are associated with wage premiums while others are linked to wage penalties.
2. Some soft skills are significant predictors of a job's gender composition. Utilizing solely soft skills, we can explain 11% of the variation in the gender composition of job categories. Soft skills that are associated with gender stereotypes, such as *empathy* and *sensitivity* for women, are significant predictors for a high percentage of women in the respective jobs, and vice versa is found for characteristics perceived as being “male”.

However, the selection of men and women into different occupations would in itself not be crucial for labour market inequality, as long as this segregation only implies that men and women work in different occupations and no other repercussions are attached. Previous research, however, has pointed out that wages paid in female-dominated occupations are lower than in male-dominated occupations [47–49]. Sex segregation in labour market is thus perceived as being a crucial factor of perpetuating wage differentials between men and women. Therefore, our results suggest that gender stereotypical job ads serve as part of a leaky pipeline upholding gender wage inequality, by contributing to a selection of women into lower paying occupations, on the basis of employing wording that discourages them to apply to higher paid male-dominated jobs in the first place.

3. Typically “female” soft skills, i.e. prescribed stereotypes about women, are mostly associated with wage penalties, while soft skills associated with *leadership*, and as such stereotypes that are associated with men, come with wage premiums—even after controlling for the job title and job category.

Although, by drawing on empirical research from psychology, we could explain which tasks are associated with being “male” or “female”, we believe that certain soft skills, such as being *respectful* and being *curious* are probably important in any kind of job. Given this assumption, it is the more compelling to find that while the former is associated with a high percentage of women in an occupation and wage penalties, the latter comes about with wage premiums and is found in job ads for male-dominated occupations. This hints, as discussed, at a general devaluation of task carried out by women in labour markets.

One might wonder, if women could not simply apply for jobs that are advertised using “male” soft skills and thereby circumvent possible wage penalties. Current evidence however shows that the solution is not that simple: women are less likely to be successful when

applying for a male-dominated job and when violating female gender stereotypes [20, 50, 51].

This study was not without limitations. Therefore next we discuss these restraints and briefly consider how these limitations can be addressed in the future research.

First, distinguishing between when a given soft skill is a necessity for a job or merely a useful asset is beyond the scope of this paper. The accuracy of the soft skill detection method, as well as the distinction of a soft skill being an asset or a necessity, could be improved by considering part-of-speech features.

Second, although we were able to account for a considerable degree of unobserved occupational heterogeneity by using matching techniques, in order to rigorously test the impact soft skills on wages, one would need to analyze if wage premiums or penalties associated with certain soft skills hold, net of individual labor-market-relevant attributes. More to the point: we believe that work experience and job tenure serve as relevant confounders in our study. The particularly large premiums for *leadership* are very likely also connected to senior positions requiring professional expertise and longstanding on-the-job experience. While work experience is to some extent controlled by using the words of the job titles (e.g. *senior* and *intern*) as matching criteria, in some cases, the expected work experience can be indicated merely in the job description, which is not used for matching. Given previous evidence that finds that tasks associated with being “female”, such as “nurturing skills” do pose a penalty on wages, net of individual characteristics [43], it is plausible that our results would be stable net of individual labor-market-relevant attributes as well. In future research this could be tested by linking the soft skills to individual survey data, which include measures of individual work experience.

Regardless of these limitations, this study has made an important contribution to the impact of soft skills in the labour market. Combining computational methods as well as theoretical and empirical insights from economics, sociology and psychology enabled us to shed more light on how soft skills operate in the labour market. We showed that soft skills are a crucial component of job ads, especially of low-paid jobs and jobs in female-dominated professions. Furthermore, we found evidence that soft skills are associated gender segregation across occupations and reinforce wage inequalities between men and women by rewarding typically “male” characteristics and penalizing “female” traits.

Grugulis and Vincent [6, p. 599] put it this way: “When it is an individual character that is being judged, evaluations based on gender and race are far more likely”. Put differently, personal traits and characteristics, namely soft skills, are hard to evaluate and thus likely subjected to proxies such as gender or race and associated stereotypes, which in turn leads to discrimination. Our results support this observation, as they suggest that soft skill polarize labour market outcomes in terms of wages and occupational segregation. This polarization strikes women, as an already vulnerable group in labour markets, the hardest.

Acknowledgements

We are grateful to Olaf Groh-Samberg, Karin Gottschall, Anne Busch-Heizmann, Matti Nelimarkka, and two anonymous reviewers for their invaluable feedback on previous versions of the article. All remaining errors are our own.

Funding

Not applicable.

Abbreviations

US, United States; UK, United Kingdom; NLTK, Natural Language Toolkit; IT, information technology; OLS, ordinary least squares; UCL, University College London.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the repositories specified in Sects. 2 and 4.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

FC's main responsibility consisted in preparing a dataset of soft skills, crowd-sourcing and data post-processing, as well as preparing the data for the analysis. LS assisted FC in preparing the dataset and running the experiments, and suggested computational approaches. LM's main responsibility was the interpretation of the results employing a sociological perspective, as well as conducting a literature review on the topic. CW provided guidance and directions related to data analysis. EM guided the whole process from data preparation to analysis. All authors participated in paper writing and editing. All authors read and approved the final manuscript.

Author details

¹Sapienza University of Rome, Rome, Italy. ²Aalto University, Helsinki, Finland. ³University of Bremen, Bremen, Germany. ⁴GESIS – Leibniz Institute for the Social Sciences, Cologne, Germany. ⁵University Koblenz, Koblenz, Germany.

Endnotes

- ^a https://en.oxforddictionaries.com/definition/soft_skills
- ^b The dataset from UK is available at: <https://www.kaggle.com/c/job-salary-prediction>.
- ^c Additional variables of the dataset encompass: location, type of contract (full- vs. part-time), length of contract (contract-based vs. permanent), the company name, and the source of the job ad.
- ^d The Armenian dataset is available at: <https://www.kaggle.com/madhab/jobposts>. Using a different dataset carries the risk that some skills might only appear in the UK dataset. However, this most likely only applies to very infrequent soft skills and thus would have little effect on the down-stream analyses.
- ^e <https://www.figure-eight.com/>
- ^f The list of superfluous adjectives includes: `excellent`, `highly`, `very good`, `good`, `strong`, and `high`.
- ^g <http://success.crowdfunder.com/hc/en-us/articles/201855939-How-to-Calculate-a-Confidence-Score>
- ^h Official archive available at: <https://drive.google.com/file/d/0B7XKcWpI5KDYNINUTTISS21pQmM>.
- ⁱ Clusters are available at: <https://docs.google.com/spreadsheets/d/1ke-xjXfwYHG-KPrEnJ90v41A7VzO82P7cx3Ez4lq9fQ/edit?usp=sharing>.
- ^j We used the list of English stop words from the NLTK package (<http://www.nltk.org>).
- ^k The job ads do not mention the exact annual salary but only a range, so we use the median of the range as the job salary.
- ^l <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/datasets/employeesandselfemployedbyindustryemp14>
- ^m Additionally, we found the following four matches: `Act as a leader` → `leadership`, `Self-reliant` → `confident`, `Cheerful` → `cheerful personality`, and `Sympathetic` → `sympathy`. These were, however, left out from our analysis since the former two soft skills had already been assigned to other similar stereotypes and the latter two have insufficient samples sizes of Count = 3 and Count = 4, respectively.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 22 September 2018 Accepted: 1 April 2019 Published online: 27 April 2019

References

1. Lucas S (2015) Retail and food services most at risk from soft skills deficit. <https://developmenteconomics.co.uk/retail-and-food-services-most-at-risk-from-soft-skills-deficit/>. Accessed 30 Oct 2017
2. Bacolod MP, Blum BS (2010) Two sides of the same coin us "residual" inequality and the gender gap. *J Hum Resour* 45(1):197–242
3. Alabdulkareem A, Frank MR, Sun L, AlShebli B, Hidalgo C, Rahwan I (2018) Unpacking the polarization of workplace skills. *Sci Adv* 4(7):6030. <https://doi.org/10.1126/sciadv.aao6030>
4. Bortz D (2014) Soft skills to help your career hit the big time. <https://www.monster.com/career-advice/article/soft-skills-you-need>
5. Bakhshi H, Downing JM, Osborne MA, Schneider P (2017) The future of skills employment in 2030. Technical report, Pearson PLC
6. Grugulis I, Vincent S (2009) Whose skill is it anyway? 'Soft' skills and polarization. *Work Employ Soc* 23(4):597–615
7. Shibata H (2001) Productivity and skill at a Japanese transplant and its parent company. *Work Occup* 28(2):234–260
8. Xie W-J, Yang Y-H, Li M-X, Jiang Z-Q, Zhou W-X (2017) Individual position diversity in dependence socioeconomic networks increases economic output. *EPJ Data Sci* 6(1):10

9. Wachs J, Hannák A, Vörös A, Daróczy B (2017) Why do men get more attention? Exploring factors behind success in an online design community. In: Proceedings of the eleventh international AAAI conference on web and social media (ICWSM 2017), Montreal, CA, pp 299–308
10. Moss P, Tilly C (1996) “Soft” skills and race: an investigation of black men’s employment problems. *Work Occup* 23(3):252–276
11. Ridgeway CL (1997) Interaction and the conservation of gender inequality: considering employment. *Am Sociol Rev*, 62(2):218–235
12. Ceci SJ, Williams WM (2011) Understanding current causes of women’s underrepresentation in science. *Proc Natl Acad Sci USA* 108(8):3157–3162
13. Lester J (2010) Women in male-dominated career and technical education programs at community colleges: barriers to participation and success. *J Women Minor Sci Eng* 16(1):51–66
14. Black SE, Spitz-Oener A (2010) Explaining women’s success: technological change and the skill content of women’s work. *Rev Econ Stat* 92(1):187–194
15. Balcar J (2014) Soft skills and their wage returns: overview of empirical literature. *Rev Econ Perspect* 14(1):3–15
16. Borghans L, Weel BT, Weinberg BA (2014) People skills and the labor-market outcomes of underrepresented groups. *ILR Rev* 67(2):287–334
17. Levanon A, Grusky DB (2016) The persistence of extreme gender segregation in the twenty-first century. *Am J Sociol* 122(2):573–619
18. Bem SL (1974) The measurement of psychological androgyny. *J Consult Clin Psychol* 42(2):155–162
19. Gaucher D, Friesen J, Kay AC (2011) Evidence that gendered wording in job advertisements exists and sustains gender inequality. *J Pers Soc Psychol* 101(1):109–128
20. Rudman LA, Glick P (2001) Prescriptive gender stereotypes and backlash toward agentic women. *J Soc Issues* 57(4):743–762
21. Busch-Heizmann A (2014) Supply-side explanations for occupational gender segregation: adolescents’ work values and gender—(a) typical occupational aspirations. *Eur Sociol Rev* 31(1):48–64
22. Correll SJ (2001) Gender and the career choice process: the role of biased self-assessments. *Am J Sociol* 106(6):1691–1730
23. Askehave I, Zethsen KK (2014) Gendered constructions of leadership in Danish job advertisements. *Gend Work Organ* 21(6):531–545
24. Bem SL, Bem DJ (1973) Does sex-biased job advertising “aid and abet” sex discrimination? 1. *J Appl Soc Psychol* 3(1):6–18
25. Taris TW, Bok IA (1998) On gender specificity of person characteristics in personnel advertisements: a study among future applicants. *J Psychol* 132(6):593–610
26. Born MP, Taris TW (2010) The impact of the wording of employment advertisements on students’ inclination to apply for a job. *J Soc Psychol* 150(5):485–502
27. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems, pp 3111–3119
28. Bastian M, Hayes M, Vaughan W, Shah S, Skomoroch P, Kim H, Uryasev S, Lloyd C (2014) LinkedIn skills: large-scale topic extraction and inference. In: Proceedings of the 8th ACM conference on recommender systems. ACM, New York, pp 1–8
29. Kivimäki I, Panchenko A, Dessy A, Verdegem D, Francq P, Bersini H, Saelens M (2013) A graph-based approach to skill extraction from text. In: Proceedings of TextGraphs-8 graph-based methods for NLP, pp 79–87
30. Haranko K, Zagheni E, Garimella K, Weber I (2018) Professional gender gaps across us cities. [arXiv:1801.09429](https://arxiv.org/abs/1801.09429)
31. Sayfullina L, Malmi E, Kannala J (2018) Learning representations for soft skill matching. In: International conference on analysis of images, social networks and texts. Springer, Berlin, pp 141–152
32. Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1):41–55
33. Angrist JD, Pischke JS (2009) Mostly harmless econometrics: an empiricist’s companion. Princeton University Press, Princeton
34. Goldthorpe JH (2007) On sociology. Stanford University Press, Stanford
35. Hertel FR (2016) Social mobility in the 20th century: class mobility and occupational change in the United States and Germany. Springer, Berlin
36. Wright EO (1997) Class counts: comparative studies in class analysis. Cambridge University Press, Cambridge
37. Kuhn P, Weinberger C (2005) Leadership skills and wages. *J Labor Econ* 23(3):395–436
38. Weinberger CJ (2014) The increasing complementarity between cognitive and social skills. *Rev Econ Stat* 96(5):849–861
39. Mumford TV, Campion MA, Morgeson FP (2007) The leadership skills strataplex: leadership skill requirements across organizational levels. *Leadersh Q* 18(2):154–166
40. England P (2010) The gender revolution: uneven and stalled. *Gend Soc* 24(2):149–166
41. Correll SJ (2004) Constraints into preferences: gender, status, and emerging career aspirations. *Am Sociol Rev* 69(1):93–113
42. Shaw A, Hargittai E (2018) The pipeline of online participation inequalities: the case of Wikipedia editing. *J Commun* 68(1):143–168
43. England P, Herbert MS, Kilbourne BS, Reid LL, Megdal LM (1994) The gendered valuation of occupations and skills: earnings in 1980 census occupations. *Soc Forces* 73(1):65–100
44. Kilbourne BS, England P, Farkas G, Beron K, Weir D (1994) Returns to skill, compensating differentials, and gender bias: effects of occupational characteristics on the wages of white women and men. *Am J Sociol* 100(3):689–719
45. England P (1992) Comparable worth: theories and evidence. Transaction Publishers
46. Leslie S-J, Cimpian A, Meyer M, Freeland E (2015) Expectations of brilliance underlie gender distributions across academic disciplines. *Science* 347(6219):262–265
47. Levanon A, England P, Allison P (2009) Occupational feminization and pay: assessing causal dynamics using 1950–2000 us census data. *Soc Forces* 88(2):865–891

48. Mandel H (2013) Up the down staircase: women's upward mobility and the wage penalty for occupational feminization, 1970–2007. *Soc Forces* 91(4):1183–1207
49. Murphy E, Oesch D (2015) The feminization of occupations and change in wages: a panel analysis of Britain, Germany, and Switzerland. *Soc Forces* 94(3):1221–1255
50. Davison HK, Burke MJ (2000) Sex discrimination in simulated employment contexts: a meta-analytic investigation. *J Vocat Behav* 56(2):225–248
51. Benard S, Correll SJ (2010) Normative discrimination and the motherhood penalty. *Gend Soc* 24(5):616–646

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
