Gender Differences in Returns to Skills

Evidence from Matched Vacancy-Employer-Employee Data

Mathias Fjællegaard Jensen*

Copenhagen Business School

September 2021

Abstract

Recently available data from online job vacancies have enabled analyses that move beyond across-occupation variation to also include within-occupation variation in workers' task-specific skills. However, analyses of job vacancy data are limited by the fact that information on the hired worker(s) is hidden. To overcome this issue, I develop a novel, pseudo-individual match between Danish job vacancy data and register data. With data on the hired worker(s) for each online job vacancy, I can test how the employment of skills and the returns to skills depend on the gender of the worker. I use the matched employer-employee-vacancy data to show that women face significantly lower returns to cognitive, character, customer service, financial, and specific computer skills when compared to men while controlling for both occupation and firm fixed effects.

JEL classifications: J16, J24, J31, J71

Keywords: returns to skills, tasks, wage differentials, gender pay gap.

*Department of Economics, Copenhagen Business School, mfj.eco@cbs.dk, +4538155620. Thanks to Fane Groes and Moira Daly for data access, advice, and support. Thanks to Herdis Steingrimsdottir for helpful discussions. Thanks to Oliver-Alexander Press, Tim Schurig, and Peter Sundquist for their research assistance with the job vacancy data. This work was supported by the Novo Nordisk Foundation grant number NNF16OC0021056.

1. Introduction

Already in early work on gender inequalities in the labour market, both researchers and feminists focused on whether or not women and men received equal pay for equal work (e.g. Edgeworth, 1922; Fawcett, 1918). As a result, the ambition of securing equal pay for equal work for women and men also received political attention, e.g. in the US Equal Pay Act of 1963. Despite these efforts, a multitude of modern economic studies show that women continue to receive substantially lower hourly wages when compared to men, although some convergence between labour market outcomes of women and men has been observed internationally over the last few decades, both in terms of hours worked, earnings, and educational attainment (Blau and Kahn, 2017; Goldin, 2014; Lindley and Machin, 2012; Olivetti and Petrongolo, 2016).

Family formation and parenthood are emphasised as key drivers of the persistent gender inequalities in the labour market (e.g Kleven, Landais, and Søgaard, 2019). Mothers typically undertake more childcare than fathers, which affects labour market participation in the form of career interruptions (maternity leave or stay-at-home moms) and in the form of part-time or family friendly employment (e.g Bertrand, Goldin, and Katz, 2010; Gupta and Smith, 2002; Nielsen, Simonsen, and Verner, 2004; Joshi, Paci, and Waldfogel, 1999). Partly as a result of the gendered division of childcare, gender segregation into different types of jobs (e.g. occupations and industries) and into different employers (e.g. private vs. public) is also pervasive and plays a central role in explaining the persistent gender inequalities in pay (Blau and Kahn, 2017; Levanon and Grusky, 2016; Olivetti and Petrongolo, 2014; Jarman, Blackburn, and Racko, 2012; Nielsen et al., 2004; Card, Cardoso, and Kline, 2016).

For example, Card et al. (2016) show that women in Portugal are less likely to work in high paying firms, but even if they do, they are likely to receive a smaller share of the firm specific pay premiums from which their male colleagues benefit. Similarly, numerous studies have found that conditional on working in the same occupations and industries, women still receive lower wages than their male colleagues (Blau and Kahn, 2017; Goldin, 2014; Lindley, 2016). Depending on the definition of "equal work," however, these findings do not necessarily imply that women do not receive equal pay for equal work. It could be the case that women and men employ different task-specific skills, even within firms and occupations, and that these skills are renumerated differently. Task-specific skills refer to the type of skills that are associated with specific tasks, such as social skills, cognitive skills, and computer skills, but not education levels.

Exactly what individuals do at their jobs, however, has to a large degree remained a "black box" as data on the job-level composition of task-specific skills rarely are available. The few datasets that contain individual- or job-level information on task-specific skills are relatively small surveys, and thus, when using these data to estimate returns to skills, one cannot control for sorting into firms, either because data on firm affiliation is not available or because the samples are too small to include firm fixed effects in wage regressions (e.g. the OECD Survey of Adult Skills PIAAC, or the UK Skills Surveys as used by Lindley, 2012). At the same time, the US Equal Pay Act of 1963 defines equal work as requiring *"substantially equal skill … within the same establishment"* (U.S. Equal Employment Opportunity Commission, 1997). Thus, if the aim is to estimate whether or not women and men face equal pay for equal work by the definition in the US Equal Pay Act, firm fixed effects should be included in wage regressions. Also from an economic perspective, when estimating gender differences in pay, it can be argued that one should control for sorting into firms by including firm fixed effects in wage regressions if workers face non-pecuniary firm-specific benefits or costs.

To be able to control for sorting into both firms and occupations when estimating gender differences in returns task-specific skills, I develop a matched vacancy-employer-employee dataset from Danish online job posts covering the period 2007-2017. Information on task-specific skills can be extracted from the text from each job post (using an approach similar to Deming and Kahn, 2018). Uniquely, the Danish job vacancy data can be matched with Danish register data at the firm*occupation*month-level. This exercise can only be undertaken because Danish register data include monthly information on employment, including earnings, occupation codes, and firm identifiers, for the universe of Danish employees. The resulting pseudo-individual match between vacancy data and register data make it possible to evaluate gender differences in returns to skills both *across* and *within* occupations and firms. Due to the match with Danish register data, I can control for factors that are usually highlighted as contributing to the gender pay gap in the literature, such as parental status and sorting into firms and occupations. By doing so, I return to the traditional question: Do women and men receive equal pay for equal work? Or in my terminology: Do women and men face equal returns to the same task-specific skills, e.g. social skills, cognitive skills, and computer skills?

When answering this question, I contribute to the literature as follows: 1) I provide a validation and operationalisation of skills data derived from job vacancies matched with register data, and I provide a description of their complementarities. To my knowledge, this is the first paper that matches such data at an individual level and at a large scale. 2) I show the advantages of individual-level matched vacancy-employer-employee data by estimating returns to skills and their heterogeneity across genders while controlling for firm and occupation FEs. 3) I provide individual-level tests of various hypotheses on returns to interactions of skills from the existing literature, including that on technological change.

I find that task-specific skills do not yield particularly high returns to men beyond what can be explained by occupation and firm fixed effects with the exception of cognitive and financial skills. However, I find that there is significant heterogeneity in returns to skills across genders, even when controlling for firm and occupation FEs along with a long list of other controls. With these FEs and controls in the model, returns to 5 out of 9 task-specific skills are significantly lower for women when compared to men. Importantly, the gender differences in returns to task-specific skills are pronounced for cognitive and specific computer skills; skills that have been emphasised as being technology-complementing in the existing literature. In contrast to Deming and Kahn (2018), I do not find any positive significant effects of the interaction between social and cognitive skills.

The paper is outlined as follows. In Section 2, I provide more details on the existing literature on task-specific skills and job vacancy data. In Section 3, I describe the Danish vacancy data and register data which are utilised in the analyses that follow. Furthermore, this section includes some details on the data pre-processing. Section 4 includes descriptive analyses of the data. In section 5, I present regression models and results, as well as robustness checks. Section 6 concludes.

2. Task-specific skills and job vacancy data

Since the seminal work of Autor, Levy, and Murnane (2003) emphasised the link between task-specific skills, technological change, and job polarisation, research on the demand for certain skills and returns to these skills has become increasingly prevalent. Interactions between certain task-specific skills, e.g. cognitive and computer skills, have been highlighted as complementing new technologies, and thus, the demand and returns to these skills should increase with technological change. Recently, Deming (2017) and Weinberger (2014) have emphasised the growing employment and wages in jobs requiring both social and cognitive skills, rather than cognitive skills alone. Although Black and Spitz-Oener (2010) and Cerina, Moro, and Rendall (2020) find that the job polarization patterns noted by Autor et al. (2003) are pronounced for women than for men, little research has looked into differences in skills and in returns to skills between women and men. An exception is Bacolod and Blum (2010) who show that women are particularly well-endowed with people skills and cognitive skills and that increasing returns to these two skills can explain up to 20 % of the decline in the gender pay gap. Similarly, Beaudry and Lewis (2014) find that gender pay gap narrowed with the adoption of PCs from 1980 to 2000, because women are well-endowed with cognitive skills, which complement PC adoption. However, studies on the interaction between technological change and gender typically use skills and task data at the occupation level, i.e. they do not observe within-occupation variation in skills. Before the availability of job vacancy data, researchers typically relied on skills and tasks data from relatively small surveys or from the DOT- and O*NET-databases, which were infrequently updated and provided job characteristics that only varied at the occupation level. Hence, gender differences in skills and in returns to skills could only be inferred from the fact the occupational composition of workers differs between women and men.

With the internet's omnipresence in the Global North, online posting of job vacancies is now an integrated part of firms' recruitment of new employees. The text of each job post is highly informative when studying modern labour markets: Typically, job posts state expected skills, education, and experience of potential applicants, as well as certain characteristics of the job itself, e.g. its occupation, industry, and region. Crucially, the text of digital job posts can easily be scraped from various sources on the web. Many of the studies that utilise job vacancy data also exploit that job vacancies typically include information on skills requirements, and importantly, the derived skill measures vary within occupations. However, these studies do not tend to point at gender differences in outcomes. For example, Hershbein and Kahn (2018) show that during the Great Recession, skill requirements in job posts increase more in areas that were hit harder by the recession. Modestino, Shoag, and Ballance (2016a,b) find a similar relationship between skill requirements and the availability of workers, i.e. that skill requirements increased during the recession and decreased again through the recovery. Cortes, Jaimovich, and Siu (2018) utilise new measures of tasks from job ads in a range of US newspapers from 1960 to 2000 together with DOT data. They find that when social skills become more important within an occupation, the occupation's female share of employment also increases. After merging their skills measure to a sample of US census data, they also indicate that returns to social skills have increased over time, which is consistent with Deming's (2017) findings. Bagger, Fontaine, Galenianos, and Trapeznikova (2021) match Danish job vacancy data from the period 2003-2009 with Danish register data to document the relationship between vacancy posting and a number of firm outcomes. For example, they show that vacancy posting is associated with increasing hiring rates at the firm level. However, they do not extract data on skills from the job posts.

Of the papers utilising job vacancy data, Deming and Kahn's (2018) is the one closest related to my analysis. They use Burning Glass Technologies' online job vacancy data from 2010 to 2015 to extract 10 general skill measures at the firm*occupation level. Next, they match these skill measures to data on individual firms and to wage data from metropolitan statistical areas (MSA). Thus, they can estimate the relationships between skills and wages, as well as between skills and firm performance. Deming and Kahn (2018) find that their skills measures generally correlate positively with both wages and firm performance. High paying and high performing firms require higher levels of social and cognitive skills. When a job requires both social and cognitive skills, they find a particularly high level of wages.

However, job vacancy data are typically constrained by the fact that information on the hired worker is hidden, including information on the worker's gender and earnings. Vacancy data is often matched with firm-level data, for example by using firm names in Deming and Kahn (2018), but matching at the individual level is impossible in settings where only datasets with subsets of workers are available. Although Deming and Kahn (2018) explore variation in returns to skills both across and within occupations, they cannot say whether or not their results hold at the individual level. This follows from the fact that they cannot match their skills and firm data with employees, but only with wage data at the MSA level. With my matched employer-employee-vacancy data, I can check if the findings from the existing literature hold at the individual level, and I can test for gender differences in returns to skills while controlling for firm and occupation FEs.

3. Data sources and pre-processing

The analysis that follows relies on two sources of data. Firstly, Statistics Denmark provides register data on employment, education, demographics, firm characteristics etc. Crucially, these registers include the entire population of both employees and firms in Denmark. Furthermore, it is possible to match the different registers at both the firm level and individual level. Most importantly, monthly employment data are available, and they include a firm identifier and an occupational code for each employment relation. Secondly, Danish online job vacancy data from 2007-2017 are supplied by the Danish consultancy firm, Højbjerre Brauer Schultz (HBS). These data also include a firm identifier and an occupational code for each job post as well as a posting date. Thus, it is possible to match data from the two sources using firm identifiers and occupational codes, and by exploiting the data's time dimension. In the following subsections, I separately describe the register and job vacancy data, and next, the data match.

3.1. Register data

Danish register data contain detailed monthly employment information for the entire Danish population of employees. Monthly wages, job start and end dates, monthly hours, a firm identifier, and an occupational code are provided for each monthly observation.[1] In

[1]The data provider, Statistics Denmark, uses 6-digit Danish versions of the International Labour Organization's ISCO88 and ISCO08 occupational codes, called DISCO88 and DISCO08 codes. The occupational codes have a break in 2009/2010 as Statistics Denmark move from DISCO88 to DISCO08 codes. In order to

what follows, I define a job spell as the period over which a worker remains within the same firm*occupation cell. Thus, a new job spell starts when a worker enters a new role in the same firm (new occupation code), or when a worker gets a job in another firm (new firm identifier). However, in order to avoid possible bias from firm specific human capital accumulation, I only keep new job spells of individuals that switch to a new firm. From this definition, I construct my main dataset as follows. First, I identify new jobs in the employment register, i.e. jobs where workers are registered with a new firm identifier in a given month.[2] I construct a sample of those new jobs with the first 12 months of observations in the employment register (or fewer, if the job spell ends before). Next, I aggregate to get the 12-months-averages of hourly wages, full-time equivalents, and other relevant variables. [3] Thus, this dataset contains all new jobs and information on the first 12 months of employment. I have access to the monthly data from January 2008 until June 2018, and since I need 12 months of observations, the latest job spells included start in July 2017. The constructed dataset yield information on workers only during their first year of employment in a certain job. I impose a number of restrictions on the sample, see Appendix B.1. To complement the employment data, I extract data on demographics, years of education, student status, employment experience etc. from other registers, which completes my register-based dataset. In the following subsection briefly outline the Danish job vacancy data before moving on to describe the match between the vacancy and register data.

3.2. Job vacancy data

The vacancy data is supplied by HBS, who also have provided the initial cleaning of the data. They believe that their data contains the near universe of publicly accessible Danish online job posts from 2007 to 2017.[4,5] They remove duplicates and clean the data before

get consistent codes occupational codes over time, I convert them into 228 consistent occupational groups, which gives a level of detail somewhere between 3- and 4-digit ISCO-codes (see Appendix B.3.1 and Online Appendix D.2 for more details on occupational codes).

[2]"New" in the sense that the worker was not observed in same firm in the previous month. Furthermore, I detect gaps between spells of work in the same firm*occupation cell. If the gap between two spells is less than 6 months, I do not code reoccurring work in a firm*occupation cell as a new job, but include both of them in the same job. I also correct for changing firms identifiers.

[3]A full-time job is defined as 1923.96 hours per year by Statistics Denmark. Hence, full-time equivalents = total number of hours per year / 1923.96 (see https://www.dst.dk/da/Statistik/dokumentation/Times/moduldata-for-arbejdsmarked/fuldtid). This measure of full-time equivalents will be used as weights in the analyses that follows.

[4]For more details, see: https://hbseconomics.com/wp-content/uploads/2017/09/Eftersp%C3%B8rgslen-efter-sproglige-kompetencer.pdf

[5]Due to data collection issues at the data provider, keywords information from the latter half of October, all of November and December 2011 as well as April 2017 are not available, although metadata for these months is available. This represents a very small fraction (2.1%) of the total number of job vacancies.

machine reading the job posts. HBS extracts the date on which a given job vacancy was posted online, a firm ID, and an occupation code.[6] If the firm identifier is not listed directly in the job post, HBS imputes it from publicly accessible registers using the firm name listed in the job post. Importantly, HBS also extract keywords from the raw text in the job post. In many ways, the resulting data is similar to the US job vacancy data supplied by Burning Glass Technologies. In order to be able to match with the register datasets, the vacancy data sample is restricted to include job posts with non-missing firm identifiers and occupational codes only.

When extracting skill requirements from the job vacancy data, I initially follow the method of Deming and Kahn (2018) and map a selection of keywords into skills categories. For example, the keyword "teamwork" is indicative of a job requiring social skills. The nine skill categories as well as the categories' mapping to a selection of keywords can be found in Table 1. Unlike Deming and Kahn (2018) who only map a selection of keywords into skill categories, I assign all keywords either a skill category or a noise tag. This is done as follows: 1) The most frequent keywords (approx. 2000) are assigned a skill category or noise tag manually. These words amount to the vast majority of keyword-observations. 2) Using online dictionary APIs each word's synonyms are obtained.[7] Each word's synonyms are assigned the same category as the word itself. 3) Using online dictionary APIs each word's definition is obtained. 4) Using the definition of the words, the remaining non-categorised words are assigned a category using machine-learning methods (see Appendix B.2 for more details). After these steps, all keywords are assigned either a skill category or a noise tag. The categorised keywords undergo further pre-processing, but only after the vacancy and register data are matched. The matching procedure is described in Section 3.3.

3.3. Data match

As unique firm identifiers and occupational codes are included in both the register data and job vacancy data, the two datasets can be matched along those dimensions.[8] Furthermore, I exploit the time dimension of the data.

In order to match the Danish register data with the job vacancy data, I first assume that vacancies are posted in same month as the vacancy is filled or maximum four months

[6]HBS extracts 6-digit DISCO-codes, which I also convert to the consistent 228 occupational groups as described above, see Appendix B.3 for further details.

[7]Many thanks to the Society for Danish Language and Literature for providing access to these ressources.

[8]For the match on DISCO-codes to be reliable, the codes must be consistently coded across the register data and job vacancy data. In Appendix B.3, I briefly outline how DISCO-codes are coded in the two data sources.

Table 1: Skills categories and examples of their corresponding keywords

Skill	Examples of keywords
Cognitive	problem solving, research, analytical, critical thinking, math, statistics, systematic
Social	communication, teamwork, collaboration, negotiation, presentation, social, extrovert, network, relations
Character	organised, detail-orientated, multi-tasking, time management, meeting deadlines, energetic, busy, engaged, overview
Writing/ language	writing, language, English, German, Swedish, Norwegian
Customer Service	customer, sales, client, patient
Management	management, supervisory, leadership, mentoring, staff, control, planning, implementing
Financial	budgeting, accounting, finance, cost, tender/bids
Computer (general)	computer, spreadsheets, common software, (e.g. Microsoft Excel, PowerPoint)
Computer (specific)	programming, java, python, computer science

Note: Categories and their corresponding keywords are based on Deming and Kahn (2018), Table 1.

prior.[9] For example, if a job spell starts in May, the corresponding vacancy would be posted any time from the beginning of January to the end of May in the same year. With this assumption, I use the job vacancy data to construct a rolling sum of job vacancies for each firm*occupation cell. If a new job spell appears in the employment register, I match it with job vacancies summed over the relevant 5 months. For example, if a firm posts two job vacancies in the same firm*occupation cell, one in January and one in February, a job spell starting in January will only be matched with first vacancy, whereas a job spell starting in February will be matched with both vacancies. Because only 4 months of job vacancy data before job start is needed, my matched data is only limited by the availability of the monthly employment register data, and thus, job spells commencing any time during the period January 2008 to July 2017 are included the final dataset [10]. This matching strategy gives a pseudo-individual-level match between new employees and their corresponding job post. Table 2 shows match rates aggregated to the yearly level for the dataset using the 228 occupational groups.

[9] In job posts, job start dates are often reported as an interval or not reported at all, but the posting date is accurately measured. Considering both the time between the posting date and the application deadline as well as the time from the application deadline to job start, a 5 months rolling window should capture most matches.

[10] Job spells commencing on 1 January 2008 are excluded, as I cannot check if a person was employed in the same firm*occupation cell in December 2007. However, spells commencing 2-31 January 2008 remain included

Table 2: Match rates

Year	New jobs	Matched new jobs	% new jobs matched	Job posts	Matched job posts	% job posts matched
2009	410 850	114 855	28.0	101 241	46 709	46.0
2010	430 916	105 559	24.5	90 232	41 856	46.4
2011	413 976	93 868	22.7	74 623	32 658	43.7
2012	397 809	101 757	25.6	101 114	49 940	49.4
2013	411 235	119 602	29.1	109 475	57 242	52.2
2014	422 277	117 994	27.8	115 159	59 130	51.2
2015	462 099	124 529	26.8	126 497	67 684	53.5
2016	497 670	133 270	26.7	130 171	69 465	53.4
Total	3 446 832	911 434	26.3	848 512	424 684	50.0

Note: As job spells commencing on 1 January 2008 or after July 2017 are excluded, the counts and match rates are not comparable to those reported here and are therefore excluded.

It is not surprising that only 26.3 % of new jobs from the employment register can be matched with a job post. Many of the new jobs are likely to be informal hires (the job is not publicly posted), or hires in a job that does not correspond with the job title in the job posts. This will, of course, result in an occupational mismatch. However, 50 % of job posts are matched to new jobs in employment register. This is a very high match rate when compared to, for example, Kettemann, Mueller, and Zweimüller (2018) who undertakes a similar exercise using Austrian data.

It is necessary to assume that new employees' skills levels are reflected in the job posts in their firm*occupation cell just around the start of their job spell. Furthermore, focusing on the first 12 months of wages in a job spell should limit bias from additional human capital accumulation in the firm*occupation cell. Since only few workers tend to start in the same firm*occupation cell in a given month, the level of aggregation is low. However, aggregating the job vacancy/skills data at the firm*occupation*start-month levels is a potential drawback of my data: I cannot separate women and men in the job vacancy data, and thus, I assume that everyone has the same skills at the firm*occupation*start-month level. In other words, the same skills are assigned to women and men in the same cells; I do not observe any gender variation in skills at the firm*occupation*start-month levels. If women and men tend to work in the same cells, this would restrict my analysis. However, as pointed out above, women and men tend to work in different occupations in the Danish labour market, i.e. high levels of occupational segregation are observed (Jarman et al., 2012). Due to the smaller cell sizes, gender segregation is likely to be even more pronounced at the firm*occupation*start-month levels. To explore gender segregation at these levels, I first calculate the female share of hours in each firm*occupation*start-month cell. Next, I graph the cumulative distribution of hours worked for women and men respectively on the cell's female share of hours. Figure 1 shows that women and men rarely get employed at the same time in the same firm*occupation*start-month cell. So, despite the fact that I cannot observe any gender variation in skills within firm*occupation*start-month cells, I still observe considerable gender variation in skills across these cells. Furthermore, I do observe gender differences in wages and in all other characteristics within a cell; these variables vary at the individual level.

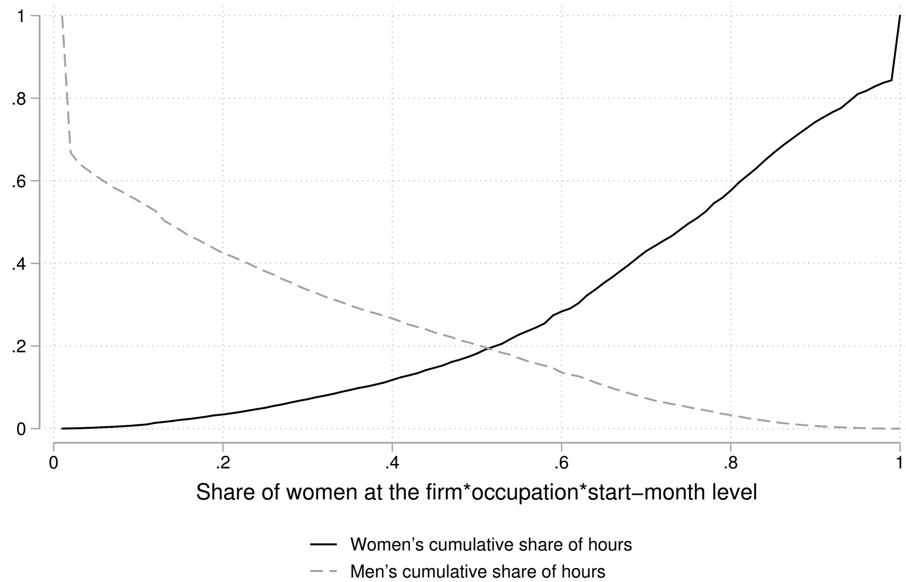
An average match rate of 26.3 % of employment register jobs can be problematic if the matched jobs spell are not representative of the population of new job spells. To check whether or not all occupations and industries are represented in the matched data, I compare the occupational and industrial distribution in the complete employment register data and in the matched subsample. Figures showing the distributions are included in Appendix B.3. The significant overrepresentation of public employees in the matched sample follows from the fact that all permanent public sector jobs by law must be publicly advertised. Thus, public sector job vacancies are also overrepresented in the vacancy data. Importantly, all industries are represented in the matched data. The data analyses includes a control variable indicating whether a job is in the public or private sector. Figure B.2 shows the representation of the 228 occupational groups in the matched data. If a data point lies to the left of the 45 degree line, it indicates that an occupation is underrepresented in the matched data, and if it lies to the right of the 45 degree line, it is overrepresented. Thus, the figure shows that smaller occupations generally are underrepresented and that larger occupations generally are overrepresented in the matched sample. Also for this reason, occupation fixed effects are included in the analyses that follows.

3.4. Skill measures

After matching job spells and job posts, the categorised keywords are revisited. If a job spell is matched with more than one job post, keywords from all the relevant job posts are aggregated. Next, the number of (aggregated) keywords belonging to the nine skill categories as well as noise words are counted for each job spell. Using these counts, the fraction of keywords indicating a certain skill are calculated for each job spell. For example, a job spell may be matched with one job post, which contains 4 % "character" words. Or a job spell may be matched with two job posts, which in total contain 8 % "character" words. However, these skill fractions are hard to interpret, and particularly in regressions analyses.

A more easily interpretable alternative would be to classify each job spell as either "character" or "not character", i.e. to create an indicator variable for each skill category. Indicator variables are easy to interpret, particularly in regression analyses with interaction terms.

Figure 1: Cumulative distribution of hours worked



Source: BFL 2008-2017, excluding observations with missing CVR- or DISCO-codes. *Notes:* Cumulative distribution of hours worked by women and men on the share of women in firm*occupation*start-month cells. Notice that hours worked by men is concentrated in cells with a low share of women and vice versa.

However, almost all job posts include one or more "character" keywords. Hence, there would be little variation in the skill measure if all job posts that include a single "character" keyword were classified as "character" rather than "not character". At the same time, other skill keywords are relatively rare, e.g. keywords indicating "computer (specific)" skills. Therefore, a simple data-driven approach is used to classify each job post as either "character" or "not character", and analogously for the remaining eight skills.

First, I consider the non-zero fractions of "charater" keywords for each job spell: At which point in distribution does the fraction of "character" keywords predict anything about wage levels? In order to determine this, I do the following: 1) Calculate each percentile in the distribution of non-zero "character" fractions. 2) Construct percentile-dummies indicating whether or not a job spell's "character" fraction is above or below each percentile. 3) Separately regress ln(hourly wages) on each of the percentile-dummies and a constant, but no control variables. 4) Choose the percentile-dummy which yields the most predictive power (the highest r^2). 5) Classify each job spell as "character" if the fraction of "character" keywords equals or exceeds that of the percentile determined by the percentile-dummy. This exercise is repeated for the remaining eight skill measures, giving nine binary skill measures.

As an alternative to the binary skill categorisation, I also develop continuous skill measures by standardising the skill fractions separately for each of the nine skills. Since keywords indicative of some skills are much more common than others, standardisation eases the interpretation and comparison of the effects of different skills on wages. The results using the binary skill indicators are reported in Section 5. In addition, all results using the continuous skills measures are reported in Appendix C.

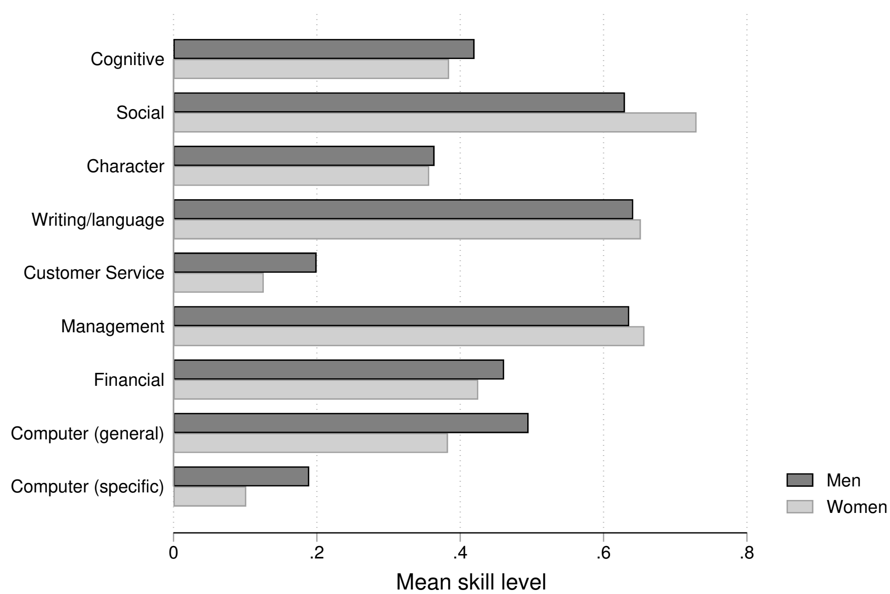
To confirm that the skill measures derived from job posts in fact reflect skill use in the corresponding jobs, I check their correlation with occupation level skill use data from PIAAC, 2011-2012. Generally, the skill measures derived from the job postings data correlate with the relevant measures from PIAAC as one would expect. More details on this validation exercise are available in Appendix D.1.

4. Descriptive statistics

4.1. Gender differences in skills

The vacancy-register data match enables analyses of skills together with the rich sets of variables provided by the Danish registers. In the context of this paper, an essential piece of information to exploit is the gender of workers. Figure 2 maps the average of jobs categorised as requiring each skill by the gender of the hired worker.

Figure 2: Mean skill levels by gender



Notes: Observations weighted by full-time equivalents. For version using continuous, standardised skill measures, see Figure C.1.

Figure 2 shows that women are overrepresented in jobs that are categorised as requiring "social", "writing/language", and "management" skills when compared to men. The opposite is the case for the remaining six skills. Despite some small gender differences, jobs are largely similarly categorised for women and men. The largest relative gender difference observed is in "computer (specific)" skills, where men are more likely to be employed in a firm*occupation*start-month cell that is categorised as requiring "computer (specific)" skills.

4.2. Correlations

Table 3 shows simple correlation coefficients between the skill measures, wages, and gender are important to consider for at least two reasons. Firstly, the skill measures should not be too highly correlated, as that could result in multicollinearity issues in regressions. Second, the correlations themselves may give us some idea of whether or not the skill measures make sense to include in wage regressions. For example, one would expect that high wage workers tend to work in cells with more skill requirements, i.e. that skills measures and wages are positively correlated.

Table 3 includes correlations between ln(hourly wages), a female dummy variable (=1 for women), and finally, all nine skill measures. All skill measures are positively correlated with wages, with the exception of "character" and "customer service" skills. Most skills are positively correlated with each other, although there are a couple of exceptions: "character" and "customer service" are negatively correlated with a few skills. This is an early indication of "character" and "customer service" skills being common in low wage jobs and in jobs with few other skills. Importantly, no skill measures are correlated to a degree that should cause problems of multicollinearity in regression models.

4.3. Variance

Although the correlation coefficients indicate that my skill measures are not correlated to a degree that would cause multicollinearity issues in regression models, the variance of the skill measures should also be explored. Before moving on to regression analyses it must be established that skill requirements cannot be entirely predicted by potential covariates. If so, the skill measures would not add any explanatory to a regression model. Thus, I regress the nine skill measures on various sets of control variables, and plot the adjusted R^2 from each regression. Figure 3 shows that between approx. 35 % and 62 % of the variance in the skill measures can be explained by the most extensive set of covariates. Notice that occupation and firm fixed-effects explain particularly large fractions of the variance in skill requirements. Still, a significant share of the variance in skill demands cannot be explained by even the most extensive set of covariates. Thus, the skill measures appear as suitable regressors in regressions in which similar sets of covariates are included, and the skill measures yield

Table 3: Correlation tables of skills, wages and gender

	ln(wage)	Female	Cognitive	Social	Character	Writing/language	Customer Service	Management	Financial	Computer (general)	Computer (specific)
ln(wage)	1										
Female	-0.150***	1									
Cognitive	0.276***	-0.0356***	1								
Social	0.0933***	0.106***	0.0953***	1							
Character	-0.150***	-0.00747***	-0.194***	-0.0248***	1						
Writing/language	0.108***	0.0108***	0.179***	0.110***	-0.112***	1					
Customer Service	-0.114***	-0.0998***	-0.114***	-0.115***	0.177***	-0.124***	1				
Management	0.148***	0.0217***	0.243***	0.203***	-0.156***	0.175***	-0.134***	1			
Financial	0.122***	-0.0357***	0.224***	0.0387***	-0.123***	0.155***	-0.0323***	0.275***	1		
Computer (general)	0.178***	-0.112***	0.289***	0.0252***	-0.170***	0.197***	-0.0548***	0.183***	0.239***	1	
Computer (specific)	0.172***	-0.126***	0.167***	-0.0107***	-0.0820***	0.0904***	-0.00112	0.0832***	0.0987***	0.218***	1

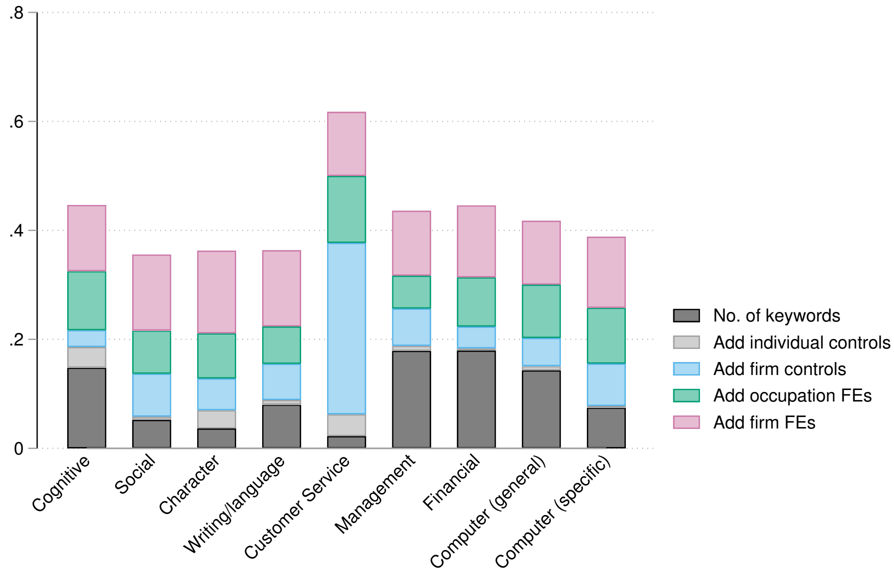
Notes: Observations weighted by full-time equivalents.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

For version using continuous, standardised skill measures, see Table C.1

explanatory power beyond that of standard labour market data (cf. Deming and Kahn, 2018).

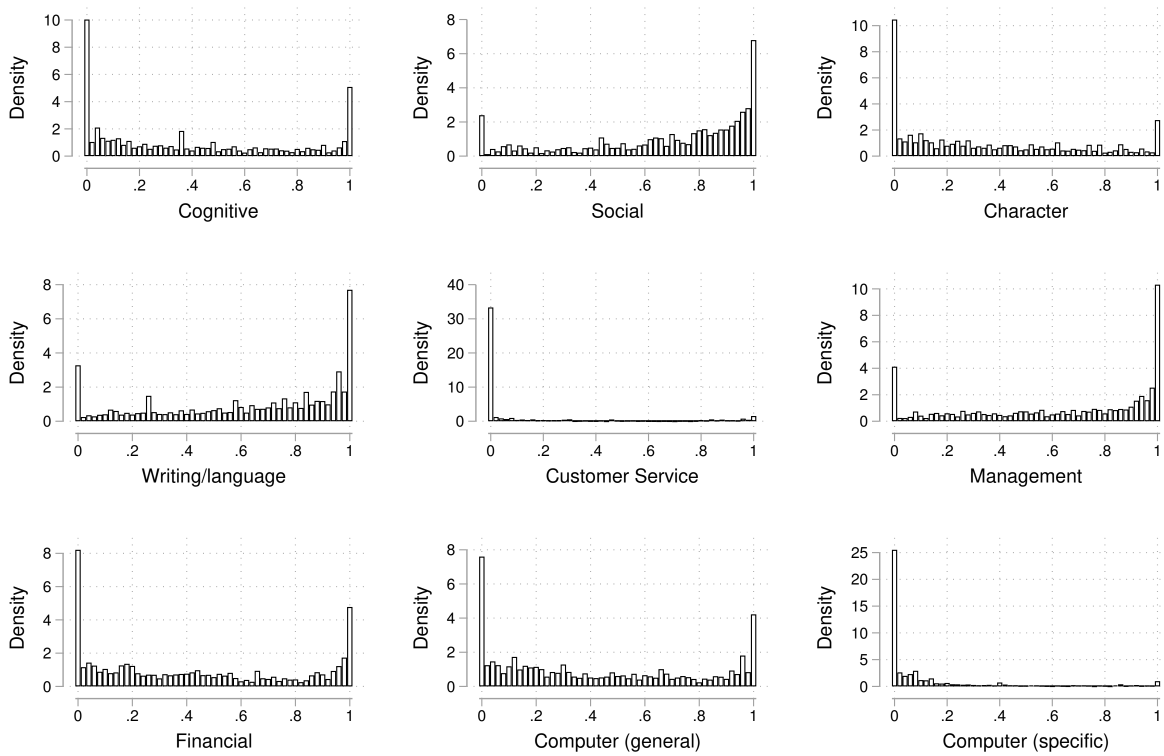
Figure 3: Adjusted R^2 from regressions of skills level on various controls



Notes: Skills are regressed on various sets of controls. Individual controls and number of keywords: *parent dummy, parent*female interaction, age, age*^{2}*, years of experience, years of education, immigrant dummy, marriage dummy, part-time dummy, year FEs, start-month FEs, number of keywords.* Firm controls: *1-letter industry dummies, firm location, number of employees, a private sector dummy.* Occupation fixed effects: *See Appendix B.3.1 for details.* Firm fixed effects: *4682 firms in total.* Observations weighted by full-time equivalents. For version using continuous, standardised skill measures, see Figure C.2.

The next section introduces the regression analyses used to estimate gender differences in returns to skills. An important aspect of the analyses is the inclusion of occupation and firm fixed effects. Hence, variation in the skills measures at these levels is crucial. Figure 4 shows the distributions of the mean skill levels within each firm*occupation cell. One possible limitation of the data could be that all job posts within the same firm*occupation cell would include the same keywords, and thus, consistently be coded as requiring the same skills. However, from Figure 4 it is evident that even at this very detailed level, the mean skill levels vary, and far from all mean skill levels are exactly equal to zero or one. A job is most rarely categorised as requiring “Customer Service” and “Computer (specfic)” skills. Many firm*occupation cells do not include jobs with these requirements, and consequently, the mean in Figure 4 is equal to zero for those jobs. Importantly, Figure 4 reveals that firm*occupation fixed effects do not sufficiently account for the variation in skill requirements across jobs.

Figure 4: Distributions of mean skill requirements within firm*occupation cells



Note: For version using continuous, standardised skill measures, see Figure C.3.

5. Regression analyses

In this section, I first outline the regression models used to estimate gender differences in returns to skills. Next, results are presented, and finally, robustness checks are reported.

5.1. Models

Regressing hourly wages on skills and their gender interactions will indicate whether women and men with same skill requirements also receive the same wage. I regress ln(hourly wages) on skills and female*skills interactions with extensive sets of control variables and fixed effects. Before writing out the relevant regression models, I outline the four sets of control variables and fixed effects that are used:

1. Individual controls and number of keywords: *parent dummy, parent*female interaction, age, age^2, years of experience, years of education, immigrant dummy, marriage dummy, part-time dummy, year FEs, start-month FEs, number of keywords*

2. Firm controls: *1-letter industry dummies, firm location, number of employees, a private sector dummy*
3. Occupation fixed effects (see Appendix B.3.1 for details)
4. Firm fixed effects

The parent dummy equals one if an individual has a child less than 18 years old. The parent*female interaction term is included to control for the gender specific effects of parenthood (cf. Kleven et al., 2019), which may otherwise affect the estimates of gender differences in returns to skills. Controls are additively included in the regression models, so it suffices to write out the full model including all controls and fixed effects:

$$w_{iofym} = \beta_0 + I_{iym}\beta_1 + F_{fy}\beta_2 + S_{ofym}\gamma_1 + S_{ofym} \times g_i \times \gamma_2 + \lambda_o + \phi_f + \theta_y + \delta_m + \varepsilon_{iofym}$$

Where the subscript i indicates variation at the individual level, f at the firm level, o at the occupation level, m at the *start*-month level, and y at the year level. w_{iofym} is ln(hourly wage). I_{iym} is a matrix of individual start-month-varying characteristics and includes a female dummy, F_{fy} a matrix of firm year-varying characteristics, S_{ofym} is a matrix of the nine skill measures that vary at the firm*occupation*year*start-month-level. g_i is a female dummy variable, which equals 1 for women only. λ_o are occupation FEs, ϕ_f firm FEs. Finally, δ_m are start-month FEs and θ_y are year FEs.[11] The vector γ_1 gives the coefficients on the nine skill measures for men and $\gamma_1 + \gamma_2$ for women, i.e. γ_2 is the gender differences in the skill measures' coefficients. Standard errors are clustered at the firm*occupation*start-month level.[12]

5.2. *Identification*

The full specification with both occupation and firm FEs targets the issue that workers with certain skill compositions may sort into high/low paying occupations and firms. Note that including both occupation and firm FEs is not analogue to including firm*occupation FEs, and thus, variation specific to the firm*occupation interactions remains. The estimated coefficients can be interpreted as *within*-firm*occupation returns to skills and *within*-firm*occupation gender differences in returns.

[11] Note that the start-month FEs and year FEs are *not* interacted, and thus, they are not year*start-month FEs.

[12] As the nine skill measures only vary at the firm*occupation*start-month levels, the errors ε_{iofym} are correlated within these cells. Thus, I follow the approach taken by Hersch (1998) and cluster my standard errors at these levels. Such an approach is also recommended by Cameron and Miller (2015). The nine skill measures are perfectly correlated within clusters (they do not vary within), and thus, applying cluster-robust standard errors significantly inflate the estimated errors.

Although the full specification with various controls, occupation, and firm FEs may identify returns to skills, omitted variable/ability bias at the individual level should also be considered. However, not many individuals start more than one job spell in a new firm in the relatively short sample period. Furthermore, there is little variation in skills *within* individuals that do (again due to the short sample period), and thus, a specification including individual FEs is not feasible, although estimates from such a specification would warrant a more causal interpretation.

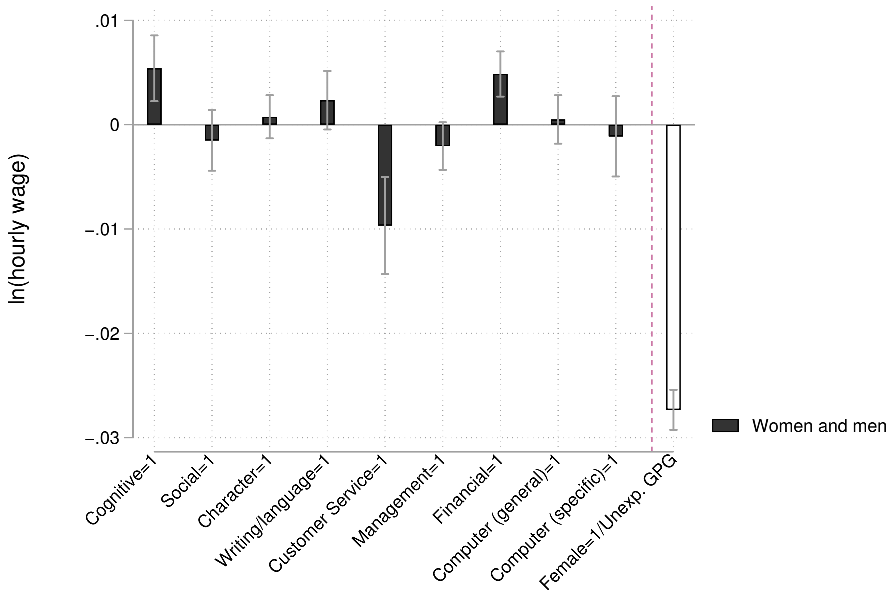
5.3. Results

First, estimation results from the linear regression model outlined above are reported, but *excluding* the gender interactions with the nine skill measures. These results are graphically presented in Figure 5 (see Table A.1 in Appendix A for point estimates and for results with varying sets of controls/FEs). The estimates show that when both occupation and firm FEs are included in the model, the coefficients on the skill measures tend to be insignificant. Exceptions are jobs categorised as requiring “customer service” skills, which are associated with lower wages. In contrast, jobs categorised as requiring “cognitive” or “financial” are associated with higher wages. Although the effects are significant, they are rather small. Thus, one could jump to the premature conclusion that task-specific skills generally do not matter much for wage formation beyond what can be explained by occupation and firm FEs.

However, after including gender interactions with the nine skills measures in the model, a different picture emerges. The results are presented in Figure 6 (see Table A.2 in Appendix A for point estimates and for results with varying sets of controls/FEs). For the model including both occupation and firm FEs, consider the coefficients on the skill measures that are not interacted with the female dummy, γ_1 (dark grey bars in Figure 6). These coefficients can be interpreted as *within*-firm*occupation returns to skills for men. The coefficients on the “cognitive”, “characer” and “financial” skill measures are positive and significant even after introducing occupation and firm FEs. However, the coefficients on the “social” and “management” measures are negative and marginally significant, but also small.

In the model including both occupation and firm FEs, the coefficients on the female*skill interactions, γ_2, can be interpreted as gender differences in returns *within*-firm*occupation cells (black bars in Figure 6). Notice the coefficient on the interaction terms with the “cognitive”, “character”, “customer service”, “financial”, and “computer(specific)” all are negative and highly significant. Thus, for five out of nine skill measures, women face lower returns. The interaction term with the “social” skill measure is positive and significant at the 0.01-level. The interaction term with the “management” skill measure is also positive,

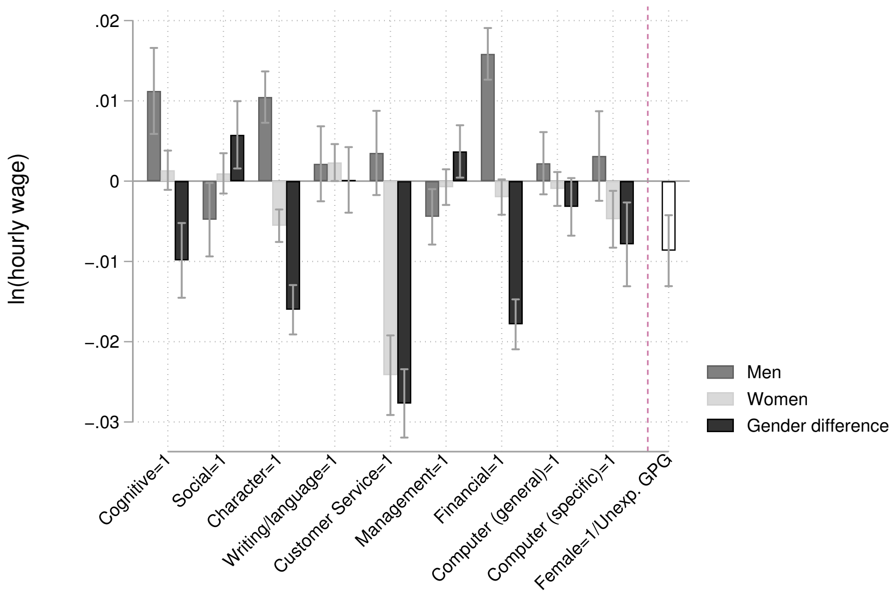
Figure 5: Coefficients on skills *without* gender interaction



Notes: Individual controls and number of keywords: *parent dummy, parent*female interaction, age, age^2, years of experience, years of education, immigrant dummy, marriage dummy, part-time dummy, year FEs, start-month FEs, number of keywords.* Firm controls: *1-letter industry dummies, firm location, number of employees, a private sector dummy.* Occupation fixed effects: *See Appendix B.3.1 for details.* Firm fixed effects: *4682 firms in total.* Observations weighted by full-time equivalents. Standard errors clustered at the firm*occupation*start-month level, 95 % confidence intervals indicated. See Column (4) in Table A.1 in Appendix A for point estimates. For version using continuous, standardised skill measures, see Figure C.4.

but the coefficient is small and only marginally significant. These estimates are the main take away from this paper. If the gender dimension of returns to skills was ignored, I could have concluded that skill generally do not yield returns beyond what can be explained by occupation and firm FEs. Instead, this specification sets the stage for the conclusion that men face positive returns to a number of skills, even within firm*occupation cells, and that women face lower returns than men to most skills, and women's returns to skills ($\gamma_1 + \gamma_2$) are generally near zero (light grey bars in Figure 6). However, before jumping to this conclusion, a couple of robustness checks are considered.

Figure 6: Coefficients on skills for women and men, and their gender difference



Notes: Individual controls and number of keywords: *parent dummy, parent*female interaction, age, age^2, years of experience, years of education, immigrant dummy, marriage dummy, part-time dummy, year FEs, start-month FEs, number of keywords.* Firm controls: *1-letter industry dummies, firm location, number of employees, a private sector dummy.* Occupation fixed effects: *See Appendix B.3.1 for details.* Firm fixed effects: *4682 firms in total.* Observations weighted by full-time equivalents. Standard errors clustered at the firm*occupation*start-month level, 95 % confidence intervals indicated. See Column (4) in Table A.2 in Appendix A for point estimates. For version using continuous, standardised skill measures, see Figure C.5.

5.4. Robustness

A couple of robustness checks are performed. First, the full specification with both occupation and firm FEs is re-estimated for a number of subpopulations in order to check whether or not the results from the previous section are driven by a certain group of workers. Next, potential interactions between skill measures are explored. In the existing literature, computer and cognitive skills have been highlighted as complementing technology and technological change, and thus, yielding positive returns. However, the results from the previous section indicate that this is only the case for men, since cognitive skills are not associated with higher wages for women. Social skills have also been emphasised as complementing technological change, but together with cognitive skills (Deming, 2017; Deming and Kahn, 2018). Lastly, the results using continuous, standardised skill measures are discussed and compared to the results using binary skill measures.

5.4.1. Subpopulations

The first robustness check focuses on the full specification, which include all occupation and firm FEs as well as various other controls. The model is estimated again on the following subpopulations:

a. Professionals [13]
b. Workers in large firms (with 100 or more employees)
c. Workers in small firms (with fewer than 100 employees)
d. Private sector workers
e. Full-time workers
f. Workers that remain employed for at least 12 months after commencing a job spell

Results for each subpopulation are reported in Table 4. The results from the entire sample generally hold for all the selected subpopulations. However, the results for relatively small sample of job spells of workers at smaller firms differ to some degree. For example, the coefficient on "character" skill measure is insignificant, and this is also the case for the coefficient on its interaction term with the female dummy. The number of job spells per firms is naturally lower for small firms, and thus, the firm FEs may account for more of the variation in wages at these firms. For the comparison with Deming and Kahn (2018), the subsample of professionals is particularly important. The results generally hold for professionals, although the coefficient on "computer (general)" skills measure is positive and significant, and the coeffecients for its interaction term between with the female dummy is negative and significant. On the other hand, the coefficient on the "character" skill measure is insignificant for the subsample of professionals. For the subsample of employees in the private sector, the coefficients on the "social" and "management" skills measures are now positive for men, although insignificant. The coefficients on the interactions between the female dummy and the skills measures also differ slightly. For private sector employees, the coefficients on the interaction terms with the "cognitive", "social", and "computer (specific)" skill measures are insignificant, whereas the coefficient on "management" skills is negative and significant.

5.4.2. Interactions

In the existing literature, interactions between certain skills are often emphasised in the context of technological change. Thus, four sets of interaction terms are explored. First, an interaction term between the "social" and "cognitive" skill measures is included. This interaction term is of particular interest after the recent work by Deming (2017), and Deming and Kahn (2018). Next, interaction terms between "cognitive" and "computer (general)", as well as "cognitive" and "computer (specific)" are included respectively. Lastly, both of the

[13] Here, professionals are crudely defined as workers with more than 16 years of education.

Table 4: ln(hourly wage) regressed on skills with gender interaction for subpopulations

	Dependent variable: ln(hourly wages)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	All	Professionals	Large firms	Small firms	Private sector	Full-time	Whole year
Female=1	-0.00866***	0.0180***	-0.00665**	-0.0243***	-0.0339***	-0.0154***	-0.00919***
	[0.00225]	[0.00307]	[0.00243]	[0.00489]	[0.00305]	[0.00272]	[0.00232]
Cognitive=1	0.0112***	0.00873***	0.0108***	0.0132**	0.00786**	0.00999**	0.00814***
	[0.00273]	[0.00210]	[0.00289]	[0.00423]	[0.00282]	[0.00310]	[0.00207]
Social=1	-0.00480*	-0.00843***	-0.00543*	0.00328	0.000142	-0.00343	-0.00649***
	[0.00234]	[0.00220]	[0.00249]	[0.00359]	[0.00225]	[0.00270]	[0.00194]
Character=1	0.0105***	-0.000340	0.0120***	-0.00500	0.00385	0.00985***	0.0102***
	[0.00163]	[0.00215]	[0.00173]	[0.00361]	[0.00243]	[0.00183]	[0.00178]
Writing/language=1	0.00216	0.00219	0.00254	-0.00103	0.00397	0.00247	0.00762***
	[0.00238]	[0.00231]	[0.00252]	[0.00387]	[0.00267]	[0.00270]	[0.00202]
Customer Service=1	0.00351	0.00995	0.00347	0.00311	-0.00665*	0.00600*	0.00443
	[0.00268]	[0.00559]	[0.00295]	[0.00412]	[0.00271]	[0.00303]	[0.00295]
Management=1	-0.00445*	0.00254	-0.00384*	-0.000546	0.000269	-0.00663***	-0.00405*
	[0.00176]	[0.00229]	[0.00184]	[0.00429]	[0.00262]	[0.00199]	[0.00192]
Financial=1	0.0159***	0.0166***	0.0158***	0.0276***	0.0214***	0.0186***	0.0165***
	[0.00164]	[0.00195]	[0.00171]	[0.00458]	[0.00248]	[0.00178]	[0.00175]
Computer (general)=1	0.00223	0.00719***	0.00251	0.00414	-0.00154	0.00286	0.00158
	[0.00198]	[0.00209]	[0.00206]	[0.00485]	[0.00293]	[0.00213]	[0.00195]
Computer (specific)=1	0.00313	0.00312	0.00273	0.00849	-0.00218	0.00297	0.00219
	[0.00285]	[0.00251]	[0.00302]	[0.00552]	[0.00307]	[0.00320]	[0.00241]
Female=1 × Cognitive=1	-0.00987***	-0.00865***	-0.00954***	-0.0135**	-0.00513	-0.00566*	-0.00752***
	[0.00237]	[0.00202]	[0.00251]	[0.00497]	[0.00267]	[0.00267]	[0.00192]
Female=1 × Social=1	0.00577**	0.00768***	0.00603**	0.000717	-0.00197	0.00760**	0.00754***
	[0.00214]	[0.00215]	[0.00229]	[0.00412]	[0.00233]	[0.00244]	[0.00195]
Female=1 × Character=1	-0.0160***	-0.00946***	-0.0172***	-0.00227	-0.00702**	-0.0178***	-0.0144***
	[0.00157]	[0.00225]	[0.00166]	[0.00422]	[0.00233]	[0.00184]	[0.00174]
Female=1 × Writing/language=1	0.000152	-0.00405	-0.000409	0.00110	0.00815**	0.000318	-0.00365
	[0.00208]	[0.00221]	[0.00221]	[0.00438]	[0.00257]	[0.00240]	[0.00187]
Female=1 × Customer Service=1	-0.0277***	-0.0499***	-0.0282***	-0.0176***	-0.00754**	-0.0408***	-0.0282***
	[0.00217]	[0.00486]	[0.00233]	[0.00494]	[0.00254]	[0.00251]	[0.00242]
Female=1 × Management=1	0.00370*	-0.00266	0.00306	-0.00602	-0.00842***	0.00617**	0.00378*
	[0.00166]	[0.00225]	[0.00175]	[0.00462]	[0.00250]	[0.00196]	[0.00182]
Female=1 × Financial=1	-0.0178***	-0.0202***	-0.0176***	-0.0242***	-0.0214***	-0.0200***	-0.0198***
	[0.00159]	[0.00192]	[0.00165]	[0.00551]	[0.00262]	[0.00175]	[0.00172]
Female=1 × Computer (general)=1	-0.00321	-0.00741***	-0.00322	-0.0152**	-0.00254	-0.00312	-0.00317
	[0.00183]	[0.00204]	[0.00191]	[0.00542]	[0.00265]	[0.00201]	[0.00187]
Female=1 × Computer (specific)=1	-0.00788**	-0.00579*	-0.00693*	-0.0124	-0.00489	-0.00809**	-0.00671**
	[0.00266]	[0.00246]	[0.00282]	[0.00650]	[0.00316]	[0.00302]	[0.00243]
Parent=1	0.0527***	0.0531***	0.0536***	0.0399***	0.0565***	0.0505***	0.0535***
	[0.00104]	[0.00167]	[0.00110]	[0.00278]	[0.00142]	[0.00117]	[0.00116]
Parent=1 × Female=1	-0.0554***	-0.0596***	-0.0564***	-0.0367***	-0.0496***	-0.0510***	-0.0542***
	[0.00129]	[0.00184]	[0.00136]	[0.00364]	[0.00172]	[0.00150]	[0.00137]
R^2	0.672	0.647	0.674	0.675	0.665	0.716	0.672
N	850063	294971	791761	58188	362655	479577	512184
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Individual controls and number of keywords: *parent dummy, parent*female interaction, age, age^2, years of experience, years of education, immigrant dummy, marriage dummy, part-time dummy, year FEs, start-month FEs, number of keywords.* Firm controls: *1-letter industry dummies, firm location, number of employees, a private sector dummy.* Occupation fixed effects: *See Appendix B.3.1 for details.* Firm fixed effects: *4682 firms in total.* Observations weighted by full-time equivalents. Cluster-robust standard errors in brackets, clustered at the firm*occupation*start-month level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. For version using continuous, standardised skill measures, see Table C.4

computer skills are interacted with the "cognitive" skill measure. In the earlier literature on the technological, cognitive and computer skills were emphasised as complementing technology during recent periods of technological change. The four sets of interaction terms are also interacted with the female dummy to check for possible gender differences in returns to these. The estimates are reported in Table 5.

The coefficients on the "social"*"cognitive" term and its interaction is negative and marginally significant, but its interaction with the female dummy is insignificant, and after including these terms, the coefficient on the "cognitive" skill measure and its interaction remain significant. Thus, at the individual level, I find no support for results in Deming and Kahn (2018). However, the interaction terms with the computer skill measures are also interesting. Returns to "computer (general)" skills do not only appear to be gender specific, but they also depend on the interaction with "cognitive" skills. When including the interaction term between "computer (general)" and "cognitive" skills, the coefficients on two skills alone are insignificant. Women also face lower returns to this interaction when compared to men.

The coefficient on the interaction between "computer (specific)" skills and "cognitive" skills is negative, but note that the coefficients on "computer (specific)" and "cognitive" skills alone are larger in this specification, and the overall effect of the combination of "computer (specific)" and "cognitive" skills remains positive. The gender difference in returns to this interaction is insignificant.

When including interaction terms between "cognitive" skills and both computer skills in the same specification (see column 5), the results from the two previous columns generally hold.

5.4.3. Continuous, standardised skill measures

The results using continuous, standardised skill measures are discussed in Appendix C.2. Figure C.4 and Table C.2 show the estimated coefficients on the continuous, standardised skill measures *without* gender interactions. Figure C.5 and Table C.3 show the analogous results for the model *including* gender interactions with the continuous skill measures. The standardised continuous skill measures generally confirm the previous results of women facing lower returns to most skills.

The negative coefficient on the binary "management" skill measure, see Figure 6, appears particularly controversial, but note that in the version using continuous, standardised skill measures, Figure C.5, the coefficient on the "management" skill measure is positive and significant for men. This difference can be explained by linear or increasing returns to management skills, rather than constant returns to management skills after surpassing a

Table 5: ln(hourly wage) regressed on skills, skill interactions, and gender interactions

	Dependent variable: ln(hourly wages)				
	(1) No interactions	(2) Cognitive and social	(3) Cognitive and computer(specific)	(4) Cognitive and computer(general)	(5) Cognitive and computer(both)
Female=1	-0.00866***	-0.00739**	-0.00821***	-0.0103***	-0.00990***
	[0.00225]	[0.00227]	[0.00226]	[0.00238]	[0.00237]
Cognitive=1	0.0112***	0.0190***	0.0135***	0.00451	0.00622*
	[0.00273]	[0.00495]	[0.00299]	[0.00270]	[0.00277]
Social=1	-0.00480*	-0.000226	-0.00489*	-0.00477*	-0.00488*
	[0.00234]	[0.00261]	[0.00232]	[0.00232]	[0.00230]
Character=1	0.0105***	0.0103***	0.0106***	0.0106***	0.0107***
	[0.00163]	[0.00163]	[0.00163]	[0.00163]	[0.00163]
Writing/language=1	0.00216	0.00218	0.00214	0.00212	0.00208
	[0.00238]	[0.00235]	[0.00238]	[0.00236]	[0.00234]
Customer Service=1	0.00351	0.00349	0.00331	0.00356	0.00332
	[0.00268]	[0.00268]	[0.00268]	[0.00268]	[0.00268]
Management=1	-0.00445*	-0.00469**	-0.00461**	-0.00436*	-0.00455**
	[0.00176]	[0.00175]	[0.00176]	[0.00177]	[0.00176]
Financial=1	0.0159***	0.0159***	0.0159***	0.0160***	0.0161***
	[0.00164]	[0.00164]	[0.00164]	[0.00165]	[0.00164]
Computer (general)=1	0.00223	0.00228	0.00215	-0.00310	-0.00416
	[0.00198]	[0.00197]	[0.00197]	[0.00268]	[0.00273]
Computer (specific)=1	0.00313	0.00295	0.00977*	0.00296	0.0115**
	[0.00285]	[0.00284]	[0.00443]	[0.00287]	[0.00447]
Female=1 × Cognitive=1	-0.00987***	-0.0130**	-0.0116***	-0.00337	-0.00464
	[0.00237]	[0.00430]	[0.00257]	[0.00244]	[0.00249]
Female=1 × Social=1	0.00577**	0.00334	0.00588**	0.00565**	0.00578**
	[0.00214]	[0.00240]	[0.00213]	[0.00211]	[0.00210]
Female=1 × Character=1	-0.0160***	-0.0159***	-0.0161***	-0.0162***	-0.0163***
	[0.00157]	[0.00157]	[0.00157]	[0.00157]	[0.00157]
Female=1 × Writing/language=1	0.000152	0.000158	0.000194	0.000196	0.000256
	[0.00208]	[0.00206]	[0.00208]	[0.00206]	[0.00205]
Female=1 × Customer Service=1	-0.0277***	-0.0278***	-0.0276***	-0.0278***	-0.0278***
	[0.00217]	[0.00217]	[0.00217]	[0.00217]	[0.00217]
Female=1 × Management=1	0.00370*	0.00385*	0.00387*	0.00365*	0.00386*
	[0.00166]	[0.00166]	[0.00166]	[0.00166]	[0.00166]
Female=1 × Financial=1	-0.0178***	-0.0179***	-0.0179***	-0.0180***	-0.0181***
	[0.00159]	[0.00158]	[0.00158]	[0.00159]	[0.00159]
Female=1 × Computer (general)=1	-0.00321	-0.00320	-0.00313	0.00182	0.00277
	[0.00183]	[0.00183]	[0.00182]	[0.00251]	[0.00254]
Female=1 × Computer (specific)=1	-0.00788**	-0.00777**	-0.0126**	-0.00755**	-0.0143**
	[0.00266]	[0.00266]	[0.00448]	[0.00271]	[0.00451]
Cognitive=1 × Social=1		-0.0113*			
		[0.00500]			
Female=1 × Cognitive=1 × Social=1		0.00527			
		[0.00454]			
Cognitive=1 × Computer (specific)=1			-0.0122*		-0.0157**
			[0.00501]		[0.00529]
Female=1 × Cognitive=1 × Computer (specific)=1			0.00812		0.0118*
			[0.00518]		[0.00543]
Cognitive=1 × Computer (general)=1				0.0128**	0.0151**
				[0.00452]	[0.00472]
Female=1 × Cognitive=1 × Computer (general)=1				-0.0121**	-0.0142**
				[0.00418]	[0.00435]
R^2	0.672	0.672	0.672	0.672	0.672
N	850063	850063	850063	850063	850063
Individual controls	Yes	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes	Yes
Occupation FEs	Yes	Yes	Yes	Yes	Yes
Firm FEs	Yes	Yes	Yes	Yes	Yes

Individual controls and number of keywords: *parent dummy, parent*female interaction, age, age^2, years of experience, years of education, immigrant dummy, marriage dummy, part-time dummy, year FEs, start-month FEs, number of keywords.* Firm controls: *1-letter industry dummies, firm location, number of employees, a private sector dummy.* Occupation fixed effects: *See Appendix B.3.1 for details.* Firm fixed effects: *4682 firms in total.* Observations weighted by full-time equivalents. Cluster-robust standard errors in brackets, clustered at the firm*occupation*start-month level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
For version using continuous, standardised skill measures, see Table C.5

certain skill level threshold, which is the underlying assumption for the binary skill measures. When using the continuous, standardised skill measures, the coefficient on the interaction between the female indicator and the "management" skill measure is negative and significant.

For the remaining skills measures, however, the results using the binary skill measures align well with those using continuous, standardised skill measures, and the coefficients on the binary measures remain easier to interpret.

6. Conclusions

In this paper, I return to the question of whether or not women and men face equal pay for equal work, which is defined by US Equal Pay Act of 1963 as requiring *"substantially equal skill ... within the same establishment"* (U.S. Equal Employment Opportunity Commission, 1997). Previously, data on task-specific skills have either been derived from small surveys or from occupation-level databases such as O*NET, and thus, estimating gender differences in returns to skills was unfeasible while also controlling for occupation and firm FEs.

In order to determine whether or not women and men face different returns to the same task-specific skills, a combination of Danish job vacancy and Danish matched employer-employee data is operationalised for the first time. Internationally, only few studies has merged job vacancy data with individual-level data, but with much lower match rates (e.g. Kettemann et al., 2018). Thus, this paper is one of the first to utilise individual-level variation in characteristics and wages together with skills data from job vacancies. I derive nine task-specific skill measures from job posts by a reading of keywords. The vacancy and register data are matched on the firm*occupation*start-month-level, which involves some aggregation. However, a high degree of gender segregation at these levels preserves variation in skills across genders. With the matched data it is possible to show that variation in skills cannot be entirely explained by an extensive set of control variables, occupation FEs, and firm FEs. Keeping this in mind, the skills measures are included in wage regressions. With control variables, and firm and occupation FEs included in wage regressions, the coefficients on the nine skill measures are largely insignificant, but only when gender interactions with the skill measures are excluded. When including interactions between gender and skills, a different story emerges.

Even after including occupation FEs, firm FEs, and the extensive set of control variables, the coefficients on the female interactions with "cognitive", "character", "customer service", "financial", and "computer(specific)" skills are negative and significant. "Social" skills are associated with lower wages for men, but not for women. Thus, ignoring the gendered dimension of returns to skills would generally lead to (gender-)biased results and conclusions, overestimating the returns to skills for women and underestimating them for men.

For the subsample of workers in the private sector, I find no significant gender differences in returns to "social" and "cognitive" skills, but instead, the female interaction with "management" skills is negative and significant.

Additionally, interactions between the skill measures are considered. After including controls and FEs, the interaction between the "cognitive" and "social" skill measures is not associated with higher wages at the individual level, which contrasts a number of studies on returns to skills, which highlight the interaction between social and cognitive skills as particularly important (Deming, 2017; Deming and Kahn, 2018; Weinberger, 2014). This may be due to differences between the US and Danish labour market, but it may also be due to the fact that individual-level data on job skill requirements and wages have not been utilised in a US context. Particularly, Deming and Kahn (2018) find positive correlations between the dual requirement of social and cognitive skills in job posts and MSA-occupation level wages as well as firm performance. A possible explanation of the difference between my findings and those of Deming and Kahn (2018) is that the individual team member providing social skills may not be fully rewarded for the positive spillovers on team members. An individual's social skills may increase own marginal productivity slightly, but also increase the overall productivity of a team or firm (cf. Deming, 2017). Thus, individual returns to social skills (and the interaction with cognitive skills) may be low, while the total returns to teams and firms may be high.

Furthermore, the interaction between the skill measures "cognitive" and "computer (general)" is associated with higher wages for men, but not for women. The interaction between cognitive and computer skills was highlighted as particularly important in the early literature on technological change. My results indicate that this skill interaction remains important for wage formation as it is associated with higher wages for men, but not for women when controlling for occupation and firm FEs.

How do the findings of gender differences in returns to task-specific skills relate to the existing literature on gender differences in earnings? Firstly, the aim of this paper is to compare the earnings of women and men that, by a very narrow definition, undertake "equal work" and not to provide a decomposition of the gender pay gap. Therefore, I control for factors that may affect earnings either because they have an effect on the nature of work itself, or because they are related to non-pecuniary benefits and cost at the job. Thus, when estimating gender differences in returns to skills, I control for sorting into firms and occupations, parental status, years of education, part-time status etc.; factors that are all highlighted as drivers of gender differences in earnings (e.g. Blau and Kahn, 2017). While controlling for those factors, I compare the earnings of women and men who are employed in jobs that require the same task-specific skills, and I find that women face lower return to most of these skills. In this setting, women and men do not face equal pay for equal skills, and thus, not for equal work. Because of the relatively short sample period, I cannot confirm whether or not changing skills prices has caused a narrowing of the gender pay gap (as pointed out by Bacolod and Blum, 2010; Beaudry and Lewis, 2014; Rendall, 2010; Yamaguchi, 2018). However, my results confirm that differences in returns to skills contribute to the gender pay gap (cf. Lindley, 2012).

How can the findings of gender differences in returns to skills be explained? Although I control for many contributors to gender differences in pay, some of the factors that are highlighted in the existing literature as contributing to the general gender pay gap may also explain gender differences in returns to skills.

For example, negotiation of wages at the firm-level contribute to gender differences in pay (Card et al., 2016; Bertrand, 2011). Similarly, negotiation of wages may also cause differences in returns to skills if women and men with same skills negotiate different wages within the same firm. Blau and Kahn (2017) emphasise that negotiation of wages is a form of bargaining, and if women face discrimination in the rest of the labour market, their outside options are also relatively worse, and thus, their bargaining power is – on average – also lower than that of men. This points to another explanation of gender differences in returns to skills, namely taste-based and statistical discrimination (Guryan and Charles, 2013). Employers may require the same task-specific skills of their female and male workers, but incorrectly assume that one of them is more productive than the other.

Finally, in line with Babcock, Recalde, and Vesterlund (2017a), Babcock, Recalde, Vesterlund, and Weingart (2017b), and Niederle and Vesterlund (2007), it may be the case that women are more likely to be assigned tasks with low rewards or are more likely to avoid competition to get out of these tasks. Despite the detailed task-specific skill measures used in this paper, even within the nine categories of task-specific skills, it may be case that women end up undertaking the less rewarding tasks.

Bargaining, discrimination, and gender-specific task assignments are all potential explanations of gender differences in returns to task-specific skills. However, gender-specific task assignment within the nine categories of task-specific skills is the only explanation that points to the fact that women and men may be paid differently due to differences in the work they undertake. If this is the dominant explanation, could it still be argued that women and men face equal pay for equal work? Recall that the US Equal Pay Act of 1963 defines equal work as requiring *"substantially equal skill"* and not necessarily as an identical set of tasks (U.S. Equal Employment Opportunity Commission, 1997). In this paper, I have utilised data on required task-specific skills in job posts to show that women face lower returns to five out of nine skills, and thus, by the definition in the US Equal Pay Act of 1963, the ambition of equal pay for equal work appears to remain far from reality, at least in Denmark.

References

Allen, J., Levels, M., Van Der Velden, R., 2013. Skill mismatch and skill use in developed countries: Evidence from the PIAAC study. ROA Research Memorandum 17, 45.

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

Babcock, L., Recalde, M. P., Vesterlund, L., 2017a. Gender Differences in the Allocation of Low-Promotability Tasks: The Role of Backlash. American Economic Review 107, 131–135.

Babcock, L., Recalde, M. P., Vesterlund, L., Weingart, L., 2017b. Gender differences in accepting and receiving requests for tasks with low promotability. American Economic Review 107, 714–747.

Bacolod, M. P., Blum, B. S., 2010. Two Sides of the Same Coin: U.S. "Residual" Inequality and the Gender Gap. The Journal of Human Resources 45, 197–242.

Bagger, J., Fontaine, F., Galenianos, M., Trapeznikova, I., 2021. Vacancies, Employment Outcomes and Firm Growth: Evidence from Denmark. Working Paper .

Beaudry, P., Lewis, E., 2014. Do male-female wage differentials reflect differences in the return to skill? Cross-city evidence from 1980-2000. American Economic Journal: Applied Economics 6, 178–194.

Bertrand, M., 2011. New perspectives on gender. In: *Handbook of Labor Economics*, Elsevier B.V., vol. 4, pp. 1543–1590.

Bertrand, M., Goldin, C., Katz, L. F., 2010. Dynamics of the gender gap for young professionals in the financial and corporate sectors. American Economic Journal: Applied Economics 2, 228–255.

Black, S. E., Spitz-Oener, A., 2010. Explaining Women's Success: Technological Change and the Skill Content of Women's Work. The Review of Economics and Statistics 120, 187–194.

Blau, F. D., Kahn, L. M., 2017. The Gender Wage Gap: Extent, Trends, and Explanations. Journal of Economic Literature 55, 178–865.

Cameron, A. C., Miller, D. L., 2015. A Practitioner ' s Guide to Cluster- Robust Inference. Journal of Human Resources 50, 317–372.

Card, D., Cardoso, A. R., Kline, P., 2016. Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. Quarterly Journal of Economics 131, 633–686.

Cerina, F., Moro, A., Rendall, M., 2020. The Role of Gender in Employment Polarization. Working Paper .

Cortes, G. M., Jaimovich, N., Siu, H., 2018. The "End of Men" and Rise of Women in the High-Skilled Labor Market. NBER Working Paper Series 24274.

Daly, M., Jensen, M. F., le Maire, D., 2021. University admission and preferred field of study. Working Paper pp. 1–73.

Deming, D., Kahn, L. B., 2018. Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals. Journal of Labor Economics 36, S337–S369.

Deming, D. J., 2017. The growing importance of social skills in the labor market. Quarterly Journal of Economics 132, 1593–1640.

Edgeworth, F. Y., 1922. Equal Pay to Men and Women for Equal Work. The Economic Journal 32, 431–457.

Fawcett, M. G., 1918. Equal Pay for Equal Work. The Economic Journal 28, 1–6.

Goldin, C., 2014. A grand gender convergence: Its last chapter. American Economic Review 104, 1091–1119.

Gupta, N. D., Smith, N., 2002. Children and Career Interruptions: The Family Gap in Denmark. Economica 69, 609–629.

Guryan, J., Charles, K. K., 2013. Taste-Based or Statistical Discrimination: The Economics of Discrimination Returns to its Roots. The Economic Journal 123, F417–F432.

Hersch, J., 1998. Compensating Differentials for Gender-Specific Job Injury Risks. American Economic Review 88, 598–607.

Hershbein, B., Kahn, L. B., 2018. Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings. American Economic Review 108, 1737–1772.

Jarman, J., Blackburn, R. M., Racko, G., 2012. The Dimensions of Occupational Gender Segregation in Industrial Countries. Sociology 46, 1003–1019.

Joshi, H., Paci, P., Waldfogel, J., 1999. The Wages of Motherhood: Better or Worse? Cambridge Journal of Economics 23, 543–564.

Kettemann, A., Mueller, A. I., Zweimüller, J., 2018. Vacancy Durations and Entry Wages: Evidence from Linked Vacancy-Employer- Employee Data. IZA Discussion Paper Series .

Kleven, H., Landais, C., Søgaard, J. E., 2019. Children and gender inequality: Evidence from Denmark. American Economic Journal: Applied Economics 11, 181–209.

Levanon, A., Grusky, D. B., 2016. The Persistence of Extreme Gender Segregation in the Twenty-first Century. American Journal of Sociology 122, 573–619.

Lindley, J., Machin, S., 2012. The Quest for More and More Education: Implications for Social Mobility. Fiscal Studies 33, 265–286.

Lindley, J. K., 2012. The gender dimension of technical change and the role of task inputs. Labour Economics 19, 516–526.

Lindley, J. K., 2016. Lousy pay with lousy conditions: The role of occupational desegregation in explaining the UK gender pay and work intensity gaps. Oxford Economic Papers 68, 152–173.

Modestino, A. S., Shoag, D., Ballance, J., 2016a. Downskilling: changes in employer skill requirements over the business cycle. Labour Economics 41, 333–347.

Modestino, A. S., Shoag, D., Ballance, J., 2016b. Upskilling: do employers demand greater skill when workers are plentiful? Working Paper pp. 1–46.

Niederle, M., Vesterlund, L., 2007. Do women shy away from competition? Do men compete too much? Quarterly Journal of Economics 122, 1067–1101.

Nielsen, H. S., Simonsen, M., Verner, M., 2004. Does the gap in family-friendly policies drive the family gap? Scandinavian Journal of Economics 106, 721–744.

Olivetti, C., Petrongolo, B., 2014. Gender gaps across countries and skills: Demand, supply and the industry structure. Review of Economic Dynamics 17, 842–859.

Olivetti, C., Petrongolo, B., 2016. The Evolution of Gender Gaps in Industrialized Countries. Annual Review of Economics pp. 405–434.

Rendall, M., 2010. Brain Versus Brawn: The Realization of Women's Comparative Advantage. SSRN Electronic Journal pp. 1–26.

U.S. Equal Employment Opportunity Commission, 1997. Facts About Equal Pay and Compensation Discrimination.

Weinberger, C. J., 2014. The Increasing Complementarity between Cognitive and Social Skills. The Review of Economics and Statistics 96, 849–861.

Yamaguchi, S., 2018. Changes in Returns to Task-Specific Skills and Gender Wage Gap. Journal of Human Resources 53, 32–70.

Appendix A. Additional results

Table A.1: ln(hourly wage) regressed on skills *without* gender interaction

	Dependent variable: ln(hourly wages)			
	(1)	(2)	(3)	(4)
	Individual controls	Occupation FEs	Firm FEs	Both FEs
Female=1	-0.0236***	-0.0277***	-0.0339***	-0.0273***
	[0.00424]	[0.00111]	[0.00122]	[0.000979]
Cognitive=1	0.0581***	0.0115***	0.0430***	0.00540***
	[0.00555]	[0.00185]	[0.00269]	[0.00161]
Social=1	0.0243***	0.00000872	0.00391	-0.00151
	[0.00601]	[0.00177]	[0.00275]	[0.00148]
Character=1	-0.0236***	-0.00350**	-0.0243***	0.000757
	[0.00213]	[0.00126]	[0.00132]	[0.00105]
Writing/language=1	0.01000**	0.00639***	0.00275	0.00234
	[0.00318]	[0.00166]	[0.00180]	[0.00143]
Customer Service=1	-0.0393***	-0.0150***	-0.0199***	-0.00967***
	[0.00295]	[0.00285]	[0.00241]	[0.00237]
Management=1	0.0194***	0.00108	0.0184***	-0.00205
	[0.00230]	[0.00152]	[0.00165]	[0.00116]
Financial=1	-0.00593*	0.00343*	-0.00450**	0.00485***
	[0.00235]	[0.00141]	[0.00162]	[0.00111]
Computer (general)=1	-0.00572	0.00208	-0.00134	0.000500
	[0.00345]	[0.00146]	[0.00189]	[0.00118]
Computer (specific)=1	0.0112	0.00653*	-0.00354	-0.00112
	[0.00635]	[0.00318]	[0.00277]	[0.00196]
Parent=1	0.0781***	0.0574***	0.0654***	0.0526***
	[0.00192]	[0.00110]	[0.00121]	[0.00103]
Parent=1 × Female=1	-0.0975***	-0.0602***	-0.0775***	-0.0550***
	[0.00378]	[0.00134]	[0.00157]	[0.00127]
R^2	0.478	0.635	0.565	0.672
N	850068	850068	850063	850063
Individual controls	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes
Occupation FEs	No	Yes	No	Yes
Firm FEs	No	No	Yes	Yes

Individual controls and number of keywords: *parent dummy, parent*female interaction, age, age^2, years of experience, years of education, immigrant dummy, marriage dummy, part-time dummy, year FEs, start-month FEs, number of keywords.* Firm controls: *1-letter industry dummies, firm location, number of employees, a private sector dummy.* Occupation fixed effects: *See Appendix B.3.1 for details.* Firm fixed effects: *4682 firms in total.* Observations weighted by full-time equivalents. Cluster-robust standard errors in brackets, clustered at the firm*occupation*start-month level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. For version using continuous, standardised skill measures, see Table C.2

Table A.2: ln(hourly wage) regressed on skills *with* gender interaction

	Dependent variable: ln(hourly wages)			
	(1)	(2)	(3)	(4)
	Individual controls	Occupation FEs	Firm FEs	Both FEs
Female=1	-0.00219	-0.00867***	-0.000758	-0.00866***
	[0.00605]	[0.00254]	[0.00297]	[0.00225]
Cognitive=1	0.0766***	0.0181***	0.0573***	0.0112***
	[0.0102]	[0.00311]	[0.00466]	[0.00273]
Social=1	0.0262**	-0.00316	0.00301	-0.00480*
	[0.00964]	[0.00279]	[0.00437]	[0.00234]
Character=1	-0.00349	0.00413*	-0.00855***	0.0105***
	[0.00386]	[0.00204]	[0.00194]	[0.00163]
Writing/language=1	0.00476	0.00579*	0.00160	0.00216
	[0.00640]	[0.00292]	[0.00293]	[0.00238]
Customer Service=1	-0.0130***	-0.00392	-0.00214	0.00351
	[0.00393]	[0.00334]	[0.00276]	[0.00268]
Management=1	0.00780*	-0.00119	0.0158***	-0.00445*
	[0.00357]	[0.00232]	[0.00224]	[0.00176]
Financial=1	-0.00300	0.0159***	0.0111***	0.0159***
	[0.00477]	[0.00233]	[0.00224]	[0.00164]
Computer (general)=1	-0.0157*	0.00385	-0.00201	0.00223
	[0.00616]	[0.00244]	[0.00281]	[0.00198]
Computer (specific)=1	0.0277**	0.0117**	0.00172	0.00313
	[0.00949]	[0.00439]	[0.00381]	[0.00285]
Female=1 × Cognitive=1	-0.0328***	-0.0110***	-0.0242***	-0.00987***
	[0.00854]	[0.00270]	[0.00375]	[0.00237]
Female=1 × Social=1	-0.00560	0.00596*	0.00134	0.00577**
	[0.00760]	[0.00244]	[0.00356]	[0.00214]
Female=1 × Character=1	-0.0341***	-0.0131***	-0.0257***	-0.0160***
	[0.00358]	[0.00189]	[0.00191]	[0.00157]
Female=1 × Writing/language=1	0.00820	0.000766	0.00199	0.000152
	[0.00590]	[0.00259]	[0.00257]	[0.00208]
Female=1 × Customer Service=1	-0.0579***	-0.0243***	-0.0371***	-0.0277***
	[0.00397]	[0.00284]	[0.00245]	[0.00217]
Female=1 × Management=1	0.0190***	0.00355	0.00375	0.00370*
	[0.00369]	[0.00203]	[0.00199]	[0.00166]
Female=1 × Financial=1	-0.00409	-0.0206***	-0.0251***	-0.0178***
	[0.00486]	[0.00216]	[0.00202]	[0.00159]
Female=1 × Computer (general)=1	0.0162**	-0.00346	0.000714	-0.00321
	[0.00520]	[0.00218]	[0.00238]	[0.00183]
Female=1 × Computer (specific)=1	-0.0309***	-0.0100**	-0.0102**	-0.00788**
	[0.00753]	[0.00344]	[0.00320]	[0.00266]
Parent=1	0.0787***	0.0574***	0.0653***	0.0527***
	[0.00196]	[0.00111]	[0.00122]	[0.00104]
Parent=1 × Female=1	-0.0987***	-0.0604***	-0.0776***	-0.0554***
	[0.00376]	[0.00136]	[0.00160]	[0.00129]
R^2	0.481	0.636	0.566	0.672
N	850068	850068	850063	850063
Individual controls	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes
Occupation FEs	No	Yes	No	Yes
Firm FEs	No	No	Yes	Yes

Individual controls and number of keywords: *parent dummy, parent*female interaction, age, age^2, years of experience, years of education, immigrant dummy, marriage dummy, part-time dummy, year FEs, start-month FEs, number of keywords.* Firm controls: *1-letter industry dummies, firm location, number of employees, a private sector dummy.* Occupation fixed effects: *See Appendix B.3.1 for details.* Firm fixed effects: *4682 firms in total.* Observations weighted by full-time equivalents. Cluster-robust standard errors in brackets, clustered at the firm*occupation*start-month level. ** $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.*
For version using continuous, standardised skill measures, see Table C.3

Appendix B. Data details

Some of the data details are also described in Daly, Jensen, and le Maire (2021), where we utilise a subset of the data described here.

B.1. Register data

Observations with the following characteristics are sequentially excluded from the sample:

a) With a missing DISCO-code (or a DISCO-code that cannot be converted to one of the 228 occupational groups, see Appendix D.2) or firm identifier
b) With a total number of hours for a given year below the equivalent of a full-time month (1923.96/12 as defined by Statistics Denmark) or above 3,500 hours. Part-time workers remain included.
c) Aged under 20, or over 65 when their job spell commences.
d) With an hourly wage below 30 DKK or above 5,000 DKK (in 2016-levels)
e) With total wages exceeding 10,000,000 DKK (in 2016-levels)
f) Enrolled at an educational institution when their job spell commences.
g) At firms with less than 5 full-time equivalents in the matched sample (see below).
h) In an occupational group with less than 50 full-time equivalents in matched sample (see below).

Criterion a) and b) are the most restrictive. Criterion a) is necessary to construct job spells at the firm*occupation level, missing DISCO-codes are mostly observed in the private sector. I describe the DISCO-coding in detail below. Criterion b) is imposed to avoid observations where hours of work may be misreported, e.g. freelance work. In addition, I believe that jobs spells with fewer hours than the equivalent of a full-time month are less likely to appear in the job vacancy data due to fixed costs of hiring.

B.2. Vacancy data

As pointed out in the section 3.2, all keywords are assigned either a skill category or a noise tag. This is done the following way: 1) The most frequent keywords (approx. 2000) are assigned a skill category or noise tag manually. These words amount to the vast majority of keyword-observations. 2) Using online dictionary APIs each word's synonyms are obtained. Each word's synonyms are assigned the same category. 3) Using online dictionary APIs each word's definition is obtained. 4) Using the definition of the words, the remaining non-categorised words are assigned a category using machine-learning methods. The machine-learning methods are described in more detail here.

The training set consists of both the more than 2000 manually categorised words and their categorised synonyms. In order to categorise the remaining words, the dictionary definition of each keyword obtained from two dictionaries, one Danish dictionary and one English dictionary. To use the English dictionary, the keywords are translated beforehand. Although the translation step may seem tedious, it involves some regularisation of the keywords, which again helps when looking up definitions of the words. Next, the classification exercise is undertaken.

Two approaches to the classification problem is repeated for both Danish and English versions of the keywords' definitions. The first approach is a one-step categorisation, where each keywords is assigned one of 10 categories, i.e. either one of the nine skills or a noise tag. A Random Forest Predictor is used for this exercise. The second approach is a two-step categorisation. In the first step, each keyword is classified as either noise or non-noise. In the second step each non-noise word is assigned to one of the nine skill categories. For both steps a Random Forest Predictor is applied.

Thus, four predicted categorisations are available for each keyword that was not a part of the training set: a one-step and a two-step version for both the Danish and English definitions. If predictions from all four approaches agree on a category, the keyword is assigned to this category. The same step is undertaken if predictions from three out of four approaches agree. Some words' definitions are only available in either the Danish or English dictionary. These words are categorised if the two approaches in the same language agree and if the probability of the predicted class is relatively high. For the few words that have not been categorised after these steps, English predictions with very high probabilities are considered and assigned to keywords. The predictions based on the English definitions are typically more reliable due to longer definitions of the keywords. If keywords are not categorised after this procedure, they are assigned a noise tag.

B.3. Data match

B.3.1. DISCO-codes

Statistics Denmark has adopted 6-digit versions of the International Labour Organization's ISCO88 and ISCO08 occupational codes, namely DISCO88 and DISCO08. The two extra digits are added to provide additional detail. Prior to 2010, all occupation codes in Statistics Denmark's datasets are coded using the 6-digit DISCO88 codes (although for some early years, only 4-digit codes are reported). From 2010 and onwards, all occupation codes in Statistics Denmark's datasets are coded using the 6-digit DISCO08 codes. Unfortunately, the (D)ISCO88 and (D)ISCO08 codes do not map consistently one-to-one, one-to-many or many-to-one, and thus, a crosswalk cannot be straightforwardly produced, and crosswalks are not provided by either ILO or Statistics Denmark. Since I consider labor market outcomes from 2008-2017, I need consistent occupational codes over this period. Therefore, I produce a revealed crosswalk, using occupational information on people that remain in the same job during the break in occupational codes from December 2009 to January 2010. I aggregate 4-digit DISCO88 and 6-digit DISCO08 codes into 228 mutually exclusive occupational groups. In comparison, there are 145 unique 3-digit and 479 unique 4-digit valid DISCO88 codes reported in the monthly employment register (BFL) from 2008. Thus, my grouping gives a level of detail somewhere between the 3- and 4-digit level. The crosswalk is included in Online Appendix D.2.

Although variables on wages and hours in the employment register are automatically imputed from the Danish tax authorities' data, the DISCO-codes are not. As they require some "manual" coding, i.e. placing a worker in a category, they do not appear in the Danish tax authorities' data. Hence, Statistics Denmark collects the DISCO-codes in a separate procedure. For public employees, Statistics Denmark impute DISCO-codes directly from the public wage data where every employee's job title/position is recorded. In the private sector, Statistics Denmark collect data on employees from firms with 100 or more employees every year.[14] Smaller firms are sampled to report DISCO-codes on their employees from year to year. Private employers are supplied with a correspondence table between job titles/positions and DISCO-codes in order to secure consistent reporting.[15] If a private firm is not sampled, Statistics Denmark impute an individual's DISCO-code from the previous year given that changes no in the individual's employment are observed. Otherwise, they estimate a DISCO-code from register data on each individual's education, the industry of the individual's employer, and the individual's membership of an unemployment insurance fund (these funds are often occupation-specific).[16]

In the case of the job vacancy data, HBS first extract a job title from each job post. Using a correspondence table between job titles/positions and DISCO-codes similar to that supplied by Statistics Denmark to DISCO-reporting firms, HBS can then identify the *6-digit* DISCO-code which corresponds to the extracted job title.[17] Thus, both the register data's and the

[14] For more details, see:
https://www.dst.dk/ext/loen/Vejl_Lon_ligeaar–pdf

[15] For more details, see:
https://www.dst.dk/da/Indberet/oplysningssider/loenstatistik/stillingsbetegnelser-disco-08-i-loenstatistikken

[16] For more details, see:
https://www.dst.dk/Site/Dst/SingleFiles/hojkvalbilag.aspx?varid=107187&bilagid=183191

[17] For more details, see:
http://www.hbseconomics.dk/wp-content/uploads/2017/09/Eftersp%C3%B8rgslen-efter-sproglige-kompetencer.pdf

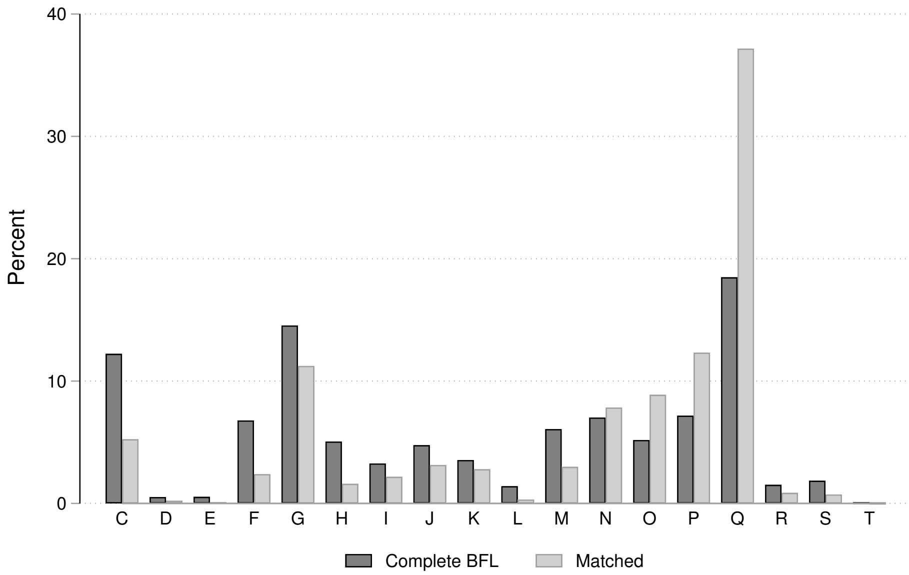
job vacancy data's *6-digit* DISCO-codes are imputed from detailed job titles/positions.

Though DISCO-codes are generally imputed in a similar manner in both the register data and the job vacancy data, some inconsistencies are to be expected at the very detailed *6-digit* level. For example, there are three subdivisions of school teachers at the *6-digit* level and only one at the *4-digit* and *3-digit* levels. For more details, see: https://www.dst.dk/da/Statistik/dokumentation/nomenklaturer/disco-08 However, due the 2009/2010 break in occupational codes, I convert the DISCO-codes from both data sources into the 228 consistent occupational groups. Another advantage of this strategy is that the aggregation resulting from the conversion also eliminates these potential inconsistencies at the very detailed 6-digit level.

B.3.2. Occupational and industrial distributions

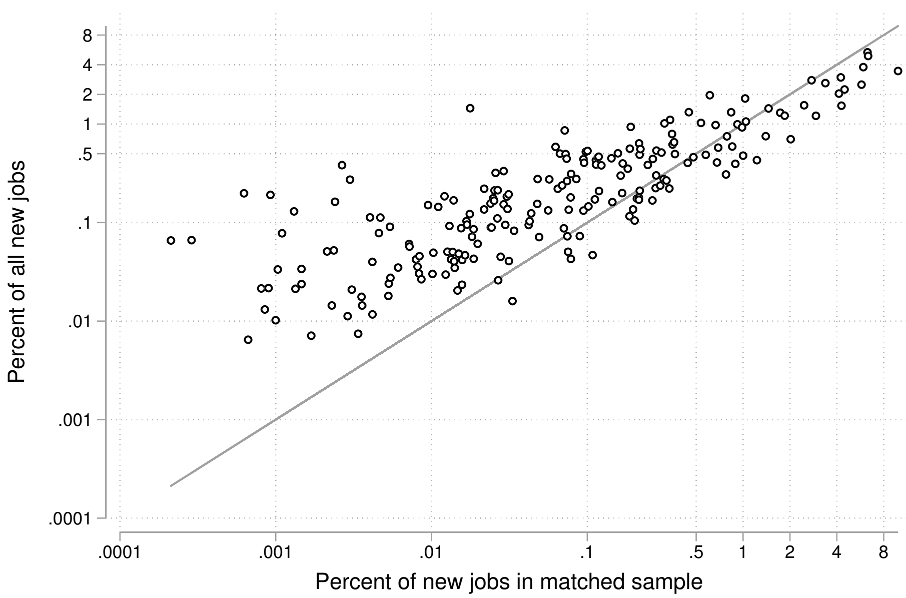
A lists of d 1-letter industries with their titles are included in Table B.1. Figure B.2 and Figure B.1 shows the occupational and industrial distribution of workers in the matched and the full sample respectively. Notice the overrepresentation of the industries "O Public administration and defence; compulsory social security", "P Education", and "Q Human health and social work activities". These occupations are dominated by large groups of public employees, namely teachers, nurses and care assistant. Figure B.2 shows the representation of the 228 occupational groups in the matched data. If a data point lies to the left of the 45 degree line, it indicates that an occupation is underrepresented in the matched data, and if to the right of the 45 degree line, it is overrepresented. Thus, the figure shows that smaller occupations generally are underrepresented and that larger occupations generally are overrepresented in the matched sample. Some occupations are dropped entirely as they are very small, and in the final estimation sample 160 occupations are represented, see exclusion criteria in Appendix B.1.

Figure B.1: Distribution of industries in the monthly employment register (BFL) and matched data



Source: BFL, 2008-2018, HBS-Jobindex 2007-2017.
Note: Observations weighted by full-time equivalents.

Figure B.2: Distribution of occupations in the monthly employment register (BFL) and matched data



Source: BFL, 2008-2018, HBS-Jobindex 2007-2017.
Note: Observations weighted by full-time equivalents.

Table B.1: 1-letter industries

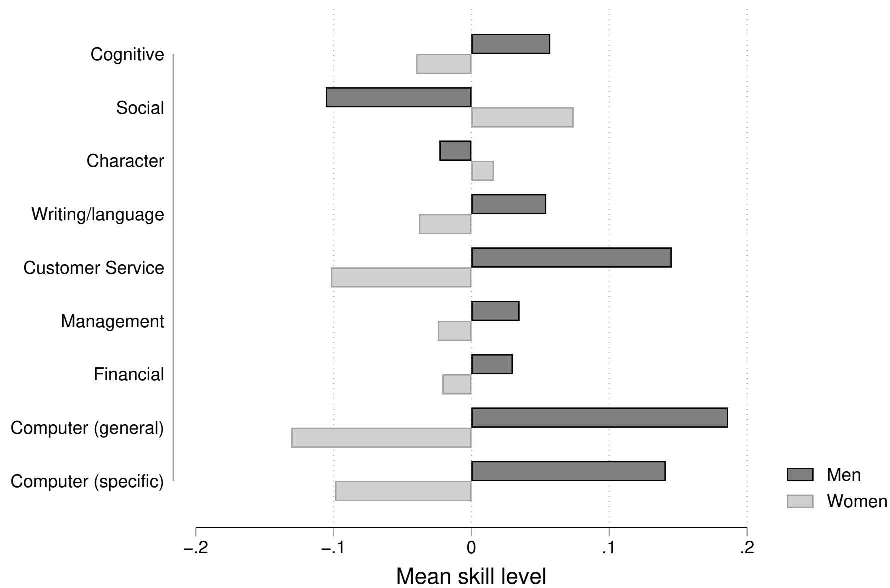
1-letter code	Industry title
A	Agriculture, forestry and fishing
B	Mining and quarrying
C	Manufacturing
D	Electricity, gas, steam and air conditioning supply
E	Water supply; sewerage; waste management and remediation activities
F	Construction
G	Wholesale and retail trade; repair of motor vehicles and motorcycles
H	Transporting and storage
I	Accommodation and food service activities
J	Information and communication
K	Financial and insurance activities
L	Real estate activities
M	Professional, scientific and technical activities
N	Administrative and support service activities
O	Public administration and defence; compulsory social security
P	Education
Q	Human health and social work activities
R	Arts, entertainment and recreation
S	Other services activities

Source: http://ec.europa.eu/competition/mergers/cases/index/nace_all.html

Appendix C. Results using continuous skills measures

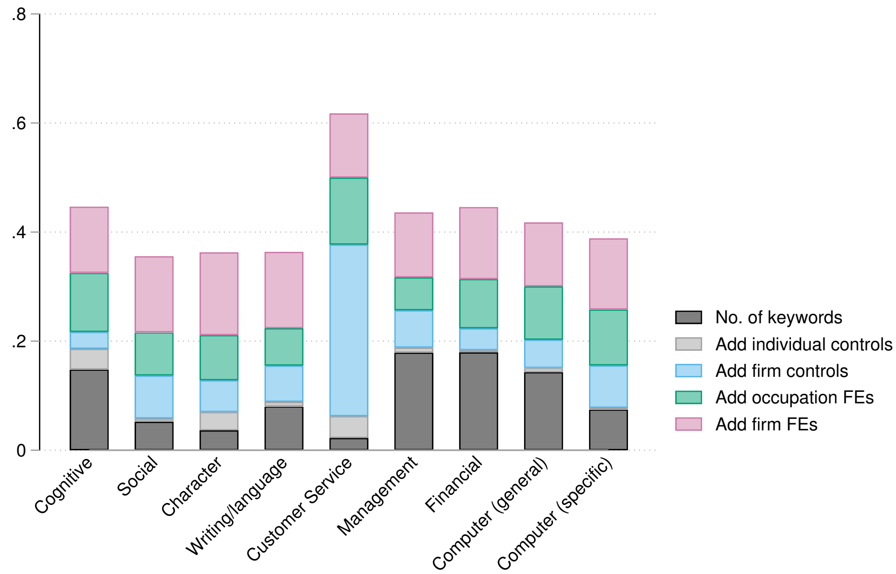
C.1. Descriptive results

Figure C.1: Continuous, standardised skill measures: Mean skill levels by gender



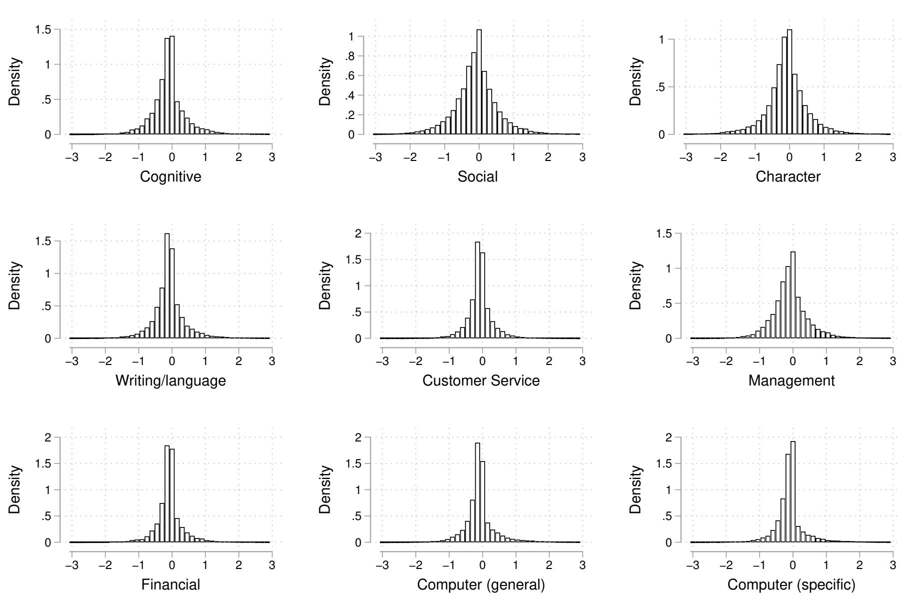
Source: BFL, 2008-2018, HBS-Jobindex 2007-2017.
Note: Observations weighted by full-time equivalents.

Figure C.2: Continuous, standardised skill measures: Adjusted R^2 from regressions of skills level on various controls



Notes: Skills are regressed on various sets of controls. Individual controls and number of keywords: *parent dummy, parent*female interaction, age, age*^{2}*, years of experience, years of education, immigrant dummy, marriage dummy, part-time dummy, year FEs, start-month FEs, number of keywords.* Firm controls: *1-letter industry dummies, firm location, number of employees, a private sector dummy.* Occupation fixed effects: *See Appendix B.3.1 for details.* Firm fixed effects: *4682 firms in total.* Observations weighted by full-time equivalents.

Figure C.3: Continuous, standardised skill measures: Distributions of de-meaned skill requirements within firm*occupation cells



Note: The figure shows individual skill level minus the mean skill level within each firm*occupation cells.

Table C.1: Continuous, standardised skill measures: Correlation tables of skills, wages and gender

	ln(wage)	Female	Cognitive	Social	Character	Writing/language	Customer Service	Management	Financial	Computer (general)	Computer (specific)
ln(wage)	1										
Female	-0.150***	1									
Cognitive	0.248***	-0.0479***	1								
Social	0.0760***	0.0885***	0.0447***	1							
Character	-0.155***	0.0194***	-0.181***	0.0752***	1						
Writing/language	0.0411***	-0.0455***	0.0380***	-0.0843***	-0.0819***	1					
Customer Service	-0.0866***	-0.122***	-0.0906***	-0.0658***	0.148***	0.0383***	1				
Management	0.103***	-0.0292***	0.187***	0.158***	-0.112***	-0.0588***	-0.0721***	1			
Financial	0.0936***	-0.0251***	0.143***	0.0169***	-0.0711***	-0.00244*	0.0375***	0.203***	1		
Computer (general)	0.153***	-0.156***	0.181***	-0.0455***	-0.109***	0.0404***	0.0547***	0.0672***	0.0407***	1	
Computer (specific)	0.152***	-0.118***	0.171***	-0.00310**	-0.0942***	0.0506***	0.00715***	0.0989***	0.0365***	0.265***	1

Notes: Observations weighted by full-time equivalents.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

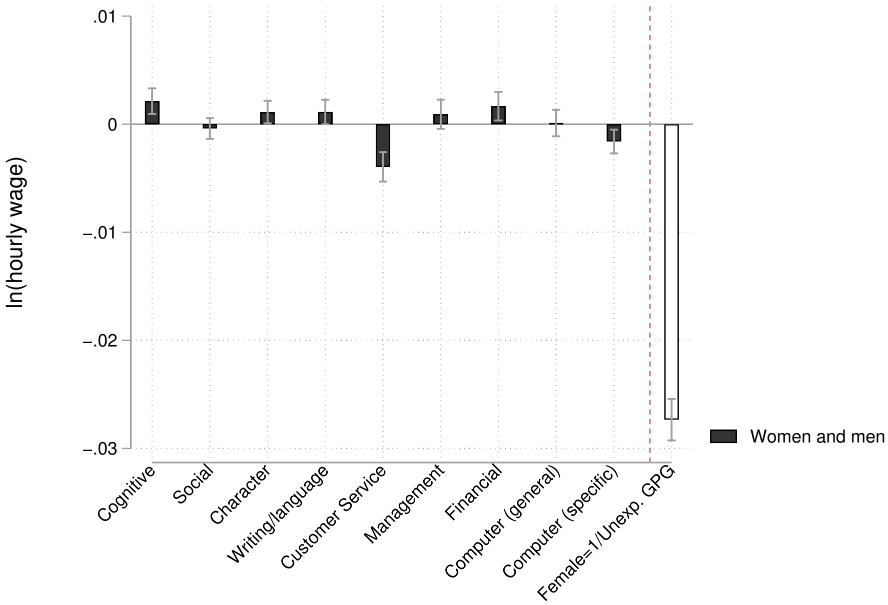
C.2. Regression estimates

Figure C.4 and Table C.2 show the estimated coefficients on the continuous, standardised skill measures *without* gender interactions. The estimates show that when both occupation and firm FEs are included in the model, the coefficients on the skill measures tend to be insignificant or only marginally significant. Exceptions are jobs with higher levels of "customer service" skills, which appear is associated with lower wages. Jobs categorised with higher level of "cognitive" skill are also found to be associated with higher wages here, and the effect is highly significant. Small, positive effects of "character", "writing/language" and "financial" skills are found, but the effects are only marginally significant. Somewhat surprisingly, "computer (specific)" are correlated with lower wages within occupation*firm cells.

Figure C.5 and Table C.3 show the analogous results for the model *including* gender interactions with the continuous skill measures. In the model including both occupation and firm FEs, consider the coefficients on the skill measures that are not interacted with the female dummy, γ_1. As proviously mentioned, these coefficients can be interpreted as *within*-firm*occupation returns to skills for men. The coefficients on the "cognitive", "character" and "financial" skill measures are large, positive and significant even after introducing occupation and firm FEs, which confirms the findings from the analysis using the binary skill measures. In addition, when using the continuous skill measures, higher level of "writing/language" skills are also correlated with higher wages. The coefficients on the "social" measure remains negative and is highly significant. The main contrast to the findings using the binary skills measures is the changed sign for the coefficient on "management" skills, which is positive for the standardised, continuous skills measures. This difference can be explained by linear or increasing returns to management skills, rather than constant returns to management skills after surpassing a certain skill level threshold, which is the underlying assumption for the binary skill measures.

In the model including both occupation and firm FEs, the coefficients on the female*skill interactions, γ_2, can be interpreted as *within*-firm*occupation gender differences in returns. Notice the coefficient on the interaction terms with all skill measures, but "social" skills, are negative, and for most of the skills also highly significant. Thus, the standardised continuous skill measures confirm the previous results of women facing lower returns to most skills. The interaction term with the "social" skill measure is positive and significant at the 0.01-level. Also for the interaction term, a change of sign on the "management" skill measure is observed when using the standardised continuous skill measure. This again implies linear or increasing returns to management skill as well as gender differences in those.

Figure C.4: Continuous, standardised skill measures: Coefficients on skills *without* gender interaction



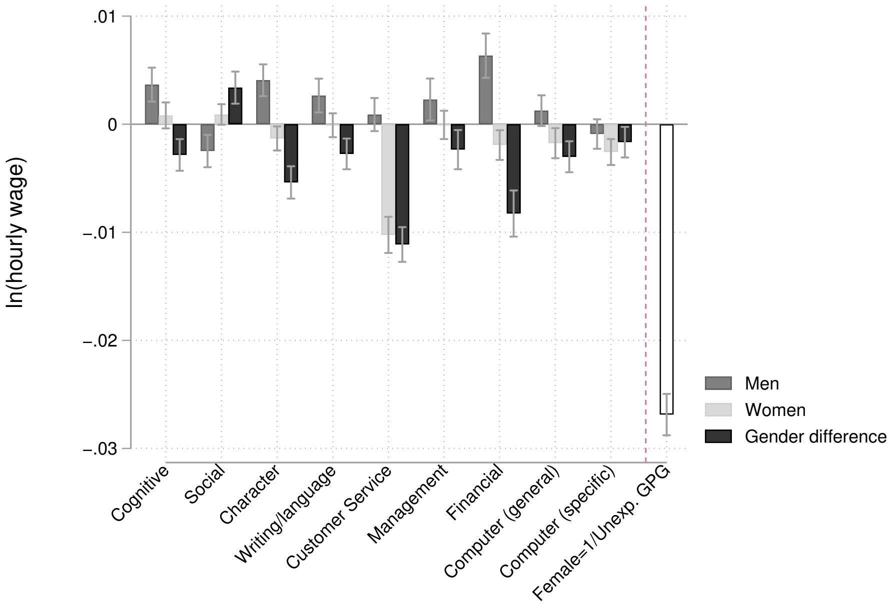
Notes: Individual controls and number of keywords: *parent dummy, parent*female interaction, age, age^2, years of experience, years of education, immigrant dummy, marriage dummy, part-time dummy, year FEs, start-month FEs, number of keywords.* Firm controls: *1-letter industry dummies, firm location, number of employees, a private sector dummy.* Occupation fixed effects: *See Appendix B.3.1 for details.* Firm fixed effects: *4682 firms in total.* Observations weighted by full-time equivalents. Standard errors clustered at the firm*occupation*start-month level, 95 % confidence intervals indicated.

Table C.2: Continuous, standardised skill measures: ln(hourly wage) regressed on skills *without* gender interaction

	Dependent variable: ln(hourly wages)			
	(1)	(2)	(3)	(4)
	Individual controls	Occupation FEs	Firm FEs	Both FEs
Female=1	-0.0208***	-0.0276***	-0.0324***	-0.0273***
	[0.00444]	[0.00110]	[0.00123]	[0.000978]
Cognitive	0.0210***	0.00578***	0.0170***	0.00213***
	[0.00135]	[0.000816]	[0.000874]	[0.000606]
Social	0.00943***	0.00136	0.00408***	-0.000397
	[0.00143]	[0.000723]	[0.000773]	[0.000492]
Character	-0.0133***	-0.00244***	-0.0177***	0.00112*
	[0.00120]	[0.000641]	[0.000794]	[0.000531]
Writing/language	0.00218*	0.00256***	0.000571	0.00113*
	[0.00110]	[0.000624]	[0.000788]	[0.000576]
Customer Service	-0.0137***	-0.00390***	-0.0105***	-0.00395***
	[0.00129]	[0.000954]	[0.000971]	[0.000696]
Management	0.0111***	0.00467***	0.0169***	0.000918
	[0.00254]	[0.000994]	[0.00118]	[0.000689]
Financial	0.00126	0.00316***	-0.00322***	0.00166*
	[0.00120]	[0.000816]	[0.000965]	[0.000673]
Computer (general)	-0.000832	0.00117	-0.00282***	0.000106
	[0.00121]	[0.00115]	[0.000740]	[0.000624]
Computer (specific)	0.00410**	0.00136	-0.00126	-0.00160**
	[0.00131]	[0.00101]	[0.000693]	[0.000560]
Parent=1	0.0783***	0.0574***	0.0652***	0.0526***
	[0.00203]	[0.00110]	[0.00121]	[0.00103]
Parent=1 × Female=1	-0.0986***	-0.0603***	-0.0774***	-0.0550***
	[0.00399]	[0.00134]	[0.00157]	[0.00127]
R^2	0.476	0.635	0.566	0.672
N	850068	850068	850063	850063
Individual controls	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes
Occupation FEs	No	Yes	No	Yes
Firm FEs	No	No	Yes	Yes

Notes: Individual controls and number of keywords: *parent dummy, parent*female interaction, age, age^2, years of experience, years of education, immigrant dummy, marriage dummy, part-time dummy, year FEs, start-month FEs, number of keywords.* Firm controls: *1-letter industry dummies, firm location, number of employees, a private sector dummy.* Occupation fixed effects: *See Appendix B.3.1 for details.* Firm fixed effects: *4682 firms in total.* Observations weighted by full-time equivalents. Cluster-robust standard errors in brackets, clustered at the firm*occupation*start-month level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Figure C.5: Continuous, standardised skill measures: Coefficients on skills for women and men, and their gender difference



Notes: Individual controls and number of keywords: *parent dummy, parent*female interaction, age, age^2, years of experience, years of education, immigrant dummy, marriage dummy, part-time dummy, year FEs, start-month FEs, number of keywords.* Firm controls: *1-letter industry dummies, firm location, number of employees, a private sector dummy.* Occupation fixed effects: *See Appendix B.3.1 for details.* Firm fixed effects: *4682 firms in total.* Observations weighted by full-time equivalents. Standard errors clustered at the firm*occupation*start-month level, 95 % confidence intervals indicated.

Table C.3: Continuous, standardised skill measures: ln(hourly wage) regressed on skills *with* gender interaction

	Dependent variable: ln(hourly wages)			
	(1)	(2)	(3)	(4)
	Individual controls	Occupation FEs	Firm FEs	Both FEs
Female=1	-0.0183***	-0.0268***	-0.0314***	-0.0269***
	[0.00420]	[0.00108]	[0.00121]	[0.000976]
Cognitive	0.0250***	0.00715***	0.0200***	0.00366***
	[0.00196]	[0.00105]	[0.00119]	[0.000802]
Social	0.00909***	-0.000133	0.00172	-0.00249**
	[0.00237]	[0.00115]	[0.00123]	[0.000764]
Character	-0.00442*	-0.000743	-0.0106***	0.00407***
	[0.00185]	[0.000974]	[0.00102]	[0.000750]
Writing/language	0.00502**	0.00419***	0.00332**	0.00265***
	[0.00163]	[0.000871]	[0.00113]	[0.000796]
Customer Service	-0.00258	0.000512	-0.00220*	0.000889
	[0.00139]	[0.00113]	[0.000971]	[0.000781]
Management	0.000825	0.00684***	0.0152***	0.00229*
	[0.00440]	[0.00140]	[0.00156]	[0.000988]
Financial	0.00978***	0.00869***	0.00701***	0.00634***
	[0.00206]	[0.00128]	[0.00145]	[0.00105]
Computer (general)	0.00182	0.00173	-0.000354	0.00125
	[0.00135]	[0.00125]	[0.000847]	[0.000725]
Computer (specific)	0.00634***	0.00200	0.000222	-0.000911
	[0.00157]	[0.00123]	[0.000828]	[0.000697]
Female=1 × Cognitive	-0.00751***	-0.00254**	-0.00555***	-0.00284***
	[0.00156]	[0.000895]	[0.00104]	[0.000747]
Female=1 × Social	-0.000515	0.00247*	0.00346**	0.00338***
	[0.00205]	[0.00101]	[0.00112]	[0.000756]
Female=1 × Character	-0.0171***	-0.00327***	-0.0127***	-0.00539***
	[0.00160]	[0.000917]	[0.000950]	[0.000763]
Female=1 × Writing/language	-0.00502***	-0.00294***	-0.00453***	-0.00275***
	[0.00140]	[0.000829]	[0.000972]	[0.000725]
Female=1 × Customer Service	-0.0271***	-0.0103***	-0.0185***	-0.0111***
	[0.00144]	[0.00106]	[0.000986]	[0.000820]
Female=1 × Management	0.0207***	-0.00403***	0.00320*	-0.00235*
	[0.00421]	[0.00117]	[0.00136]	[0.000924]
Female=1 × Financial	-0.0142***	-0.00991***	-0.0168***	-0.00827***
	[0.00206]	[0.00134]	[0.00195]	[0.00109]
Female=1 × Computer (general)	-0.00760***	-0.00181	-0.00607***	-0.00301***
	[0.00121]	[0.000928]	[0.000934]	[0.000731]
Female=1 × Computer (specific)	-0.00521***	-0.00149	-0.00365***	-0.00166*
	[0.00123]	[0.000935]	[0.000838]	[0.000722]
Parent=1	0.0791***	0.0571***	0.0651***	0.0525***
	[0.00207]	[0.00111]	[0.00122]	[0.00104]
Parent=1 × Female=1	-0.100***	-0.0601***	-0.0773***	-0.0550***
	[0.00395]	[0.00136]	[0.00158]	[0.00129]
R^2	0.480	0.636	0.568	0.672
N	850068	850068	850063	850063
Individual controls	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes
Occupation FEs	No	Yes	No	Yes
Firm FEs	No	No	Yes	Yes

Notes: Individual controls and number of keywords: *parent dummy, parent*female interaction, age, age*^{2}*, years of experience, years of education, immigrant dummy, marriage dummy, part-time dummy, year FEs, start-month FEs, number of keywords.* Firm controls: *1-letter industry dummies, firm location, number of employees, a private sector dummy.* Occupation fixed effects: *See Appendix B.3.1 for details.* Firm fixed effects: *4682 firms in total.* Observations weighted by full-time equivalents. Cluster-robust standard errors in brackets, clustered at the firm*occupation*start-month level.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table C.4: Continuous, standardised skill measures: ln(hourly wage) regressed on skills *with* gender interaction for subpopulations

	Dependent variable: ln(hourly wages)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	All	Professionals	Large firms	Small firms	Private sector	Full-time	Whole year
Female=1	-0.0269***	-0.00995***	-0.0257***	-0.0513***	-0.0508***	-0.0324***	-0.0280***
	[0.000976]	[0.00139]	[0.00102]	[0.00316]	[0.00156]	[0.00118]	[0.00114]
Cognitive	0.00366***	0.00394***	0.00354***	0.00534***	0.00385***	0.00349***	0.00313***
	[0.000802]	[0.000962]	[0.000866]	[0.00137]	[0.00106]	[0.000872]	[0.000829]
Social	-0.00249**	-0.00369***	-0.00338***	0.00222*	0.000979	-0.00227**	-0.00248**
	[0.000764]	[0.00107]	[0.000857]	[0.00106]	[0.000963]	[0.000876]	[0.000803]
Character	0.00407***	-0.00173	0.00513***	-0.00129	0.00235*	0.00452***	0.00435***
	[0.000750]	[0.00117]	[0.000861]	[0.000998]	[0.000919]	[0.000831]	[0.000825]
Writing/language	0.00265***	0.00131	0.00305***	-0.00177	0.00313***	0.00335***	0.00412***
	[0.000796]	[0.00103]	[0.000888]	[0.00107]	[0.000908]	[0.000919]	[0.000873]
Customer Service	0.000889	0.00339*	0.000712	0.00211	-0.00104	0.00221**	0.000889
	[0.000781]	[0.00168]	[0.000858]	[0.00139]	[0.000804]	[0.000851]	[0.000853]
Management	0.00229*	0.00381***	0.00232*	0.00675***	0.00401***	0.00237*	0.00252**
	[0.000988]	[0.00103]	[0.00106]	[0.00180]	[0.00111]	[0.00107]	[0.000901]
Financial	0.00634***	0.00378**	0.00662***	0.00568*	0.00596***	0.00640***	0.00498***
	[0.00105]	[0.00116]	[0.00113]	[0.00228]	[0.00106]	[0.00107]	[0.000941]
Computer (general)	0.00125	0.00147	0.00123	0.00170	0.00102	0.00123	0.000740
	[0.000725]	[0.000926]	[0.000835]	[0.000876]	[0.000725]	[0.000758]	[0.000771]
Computer (specific)	-0.000911	0.000241	-0.00104	0.00142	-0.00165*	-0.000983	-0.000619
	[0.000697]	[0.000838]	[0.000768]	[0.00121]	[0.000800]	[0.000743]	[0.000705]
Female=1 × Cognitive	-0.00284***	-0.00465***	-0.00289***	-0.00395*	-0.00109	-0.00160	-0.00255**
	[0.000747]	[0.000891]	[0.000804]	[0.00160]	[0.00104]	[0.000818]	[0.000807]
Female=1 × Social	0.00338***	0.00460***	0.00395***	0.000749	-0.00237*	0.00477***	0.00331***
	[0.000756]	[0.00105]	[0.000843]	[0.00138]	[0.00101]	[0.000878]	[0.000819]
Female=1 × Character	-0.00539***	-0.000486	-0.00614***	-0.00118	-0.00128	-0.00693***	-0.00500***
	[0.000763]	[0.00119]	[0.000861]	[0.00129]	[0.000975]	[0.000908]	[0.000874]
Female=1 × Writing/language	-0.00275***	-0.00194*	-0.00292***	0.000787	-0.000324	-0.00360***	-0.00368***
	[0.000725]	[0.000940]	[0.000807]	[0.00123]	[0.000888]	[0.000891]	[0.000828]
Female=1 × Customer Service	-0.0111***	-0.0203***	-0.0110***	-0.00881***	-0.00584***	-0.0158***	-0.0120***
	[0.000820]	[0.00168]	[0.000891]	[0.00160]	[0.00100]	[0.000985]	[0.000944]
Female=1 × Management	-0.00235*	-0.00379***	-0.00225*	-0.00973***	-0.00628***	-0.00210*	-0.00265**
	[0.000924]	[0.00102]	[0.000988]	[0.00214]	[0.00114]	[0.00105]	[0.000949]
Female=1 × Financial	-0.00827***	-0.00709***	-0.00847***	-0.00691**	-0.00771***	-0.00811***	-0.00687***
	[0.00109]	[0.00109]	[0.00117]	[0.00254]	[0.00140]	[0.00113]	[0.00105]
Female=1 × Computer (general)	-0.00301***	-0.00376***	-0.00277***	-0.00564**	-0.00207*	-0.00329***	-0.00304***
	[0.000731]	[0.000958]	[0.000797]	[0.00192]	[0.000985]	[0.000801]	[0.000817]
Female=1 × Computer (specific)	-0.00166*	-0.00218*	-0.00148	-0.00225	-0.00180	-0.00155	-0.00158*
	[0.000722]	[0.000986]	[0.000788]	[0.00164]	[0.000928]	[0.000790]	[0.000785]
Parent=1	0.0525***	0.0529***	0.0533***	0.0403***	0.0565***	0.0503***	0.0533***
	[0.00104]	[0.00165]	[0.00110]	[0.00278]	[0.00142]	[0.00117]	[0.00116]
Parent=1 × Female=1	-0.0550***	-0.0592***	-0.0560***	-0.0373***	-0.0495***	-0.0507***	-0.0539***
	[0.00129]	[0.00182]	[0.00136]	[0.00365]	[0.00172]	[0.00150]	[0.00136]
R^2	0.672	0.647	0.674	0.675	0.665	0.716	0.672
N	850063	294971	791761	58188	362655	479577	512184
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Individual controls and number of keywords: *parent dummy, parent*female interaction, age, age^2, years of experience, years of education, immigrant dummy, marriage dummy, part-time dummy, year FEs, start-month FEs, number of keywords.* Firm controls: *1-letter industry dummies, firm location, number of employees, a private sector dummy.* Occupation fixed effects: *See Appendix B.3.1 for details.* Firm fixed effects: *4682 firms in total.* Observations weighted by full-time equivalents. Cluster-robust standard errors in brackets, clustered at the firm*occupation*start-month level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table C.5: Continuous, standardised skill measures: ln(hourly wage) regressed on skills, skill interactions, and gender interactions

	Dependent variable: ln(hourly wages)				
	(1)	(2)	(3)	(4)	(5)
	No interactions	Cognitive and social	Cognitive and computer(specific)	Cognitive and computer(general)	Cognitive and computer(both)
Female=1	-0.0269***	-0.0268***	-0.0268***	-0.0267***	-0.0267***
	[0.000976]	[0.000980]	[0.000981]	[0.000984]	[0.000986]
Cognitive	0.00366***	0.00379***	0.00370***	0.00388***	0.00386***
	[0.000802]	[0.000838]	[0.000827]	[0.000835]	[0.000846]
Social	-0.00249**	-0.00246**	-0.00248**	-0.00243**	-0.00243**
	[0.000764]	[0.000768]	[0.000764]	[0.000763]	[0.000763]
Character	0.00407***	0.00406***	0.00407***	0.00411***	0.00411***
	[0.000750]	[0.000751]	[0.000751]	[0.000751]	[0.000751]
Writing/language	0.00265***	0.00263***	0.00266***	0.00267***	0.00267***
	[0.000796]	[0.000794]	[0.000796]	[0.000797]	[0.000797]
Customer Service	0.000889	0.000885	0.000886	0.000899	0.000900
	[0.000781]	[0.000780]	[0.000781]	[0.000780]	[0.000780]
Management	0.00229*	0.00223*	0.00229*	0.00229*	0.00229*
	[0.000988]	[0.000988]	[0.000988]	[0.000988]	[0.000988]
Financial	0.00634***	0.00632***	0.00634***	0.00634***	0.00634***
	[0.00105]	[0.00105]	[0.00105]	[0.00105]	[0.00105]
Computer (general)	0.00125	0.00123	0.00126	0.00115	0.00115
	[0.000725]	[0.000725]	[0.000725]	[0.000724]	[0.000723]
Computer (specific)	-0.000911	-0.000915	-0.000885	-0.000848	-0.000864
	[0.000697]	[0.000697]	[0.000703]	[0.000700]	[0.000705]
Female=1 × Cognitive	-0.00284***	-0.00302***	-0.00284***	-0.00292***	-0.00291***
	[0.000747]	[0.000764]	[0.000760]	[0.000771]	[0.000777]
Female=1 × Social	0.00338***	0.00332***	0.00338***	0.00340***	0.00340***
	[0.000756]	[0.000761]	[0.000756]	[0.000756]	[0.000757]
Female=1 × Character	-0.00539***	-0.00537***	-0.00540***	-0.00540***	-0.00540***
	[0.000763]	[0.000763]	[0.000762]	[0.000763]	[0.000763]
Female=1 × Writing/language	-0.00275***	-0.00273***	-0.00275***	-0.00276***	-0.00276***
	[0.000725]	[0.000724]	[0.000725]	[0.000726]	[0.000726]
Female=1 × Customer Service	-0.0111***	-0.0111***	-0.0111***	-0.0111***	-0.0111***
	[0.000820]	[0.000820]	[0.000820]	[0.000819]	[0.000819]
Female=1 × Management	-0.00235*	-0.00227*	-0.00235*	-0.00233*	-0.00233*
	[0.000924]	[0.000927]	[0.000924]	[0.000925]	[0.000925]
Female=1 × Financial	-0.00827***	-0.00823***	-0.00827***	-0.00828***	-0.00828***
	[0.00109]	[0.00109]	[0.00109]	[0.00109]	[0.00109]
Female=1 × Computer (general)	-0.00301***	-0.00297***	-0.00298***	-0.00291***	-0.00291***
	[0.000731]	[0.000732]	[0.000731]	[0.000725]	[0.000727]
Female=1 × Computer (specific)	-0.00166*	-0.00166*	-0.00161*	-0.00153*	-0.00152*
	[0.000722]	[0.000722]	[0.000731]	[0.000728]	[0.000734]
Cognitive × Social		0.000509			
		[0.000505]			
Female=1 × Cognitive × Social		-0.000797			
		[0.000510]			
Cognitive × Computer (specific)			-0.000130		0.0000672
			[0.000320]		[0.000334]
Female=1 × Cognitive × Computer (specific)			-0.000252		-0.0000250
			[0.000413]		[0.000425]
Cognitive× Computer (general)				-0.000745	-0.000769
				[0.000396]	[0.000413]
Female=1 × Cognitive × Computer (general)				-0.00111*	-0.00110
				[0.000553]	[0.000577]
R^2	0.672	0.672	0.672	0.672	0.672
N	850063	850063	850063	850063	850063
Individual controls	Yes	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes	Yes
Occupation FEs	Yes	Yes	Yes	Yes	Yes
Firm FEs	Yes	Yes	Yes	Yes	Yes

Notes: Individual controls and number of keywords: *parent dummy, parent*female interaction, age, age*^{2}*, years of experience, years of education, immigrant dummy, marriage dummy, part-time dummy, year FEs, start-month FEs, number of keywords.* Firm controls: *1-letter industry dummies, firm location, number of employees, a private sector dummy.* Occupation fixed effects: *See Appendix B.3.1 for details.* Firm fixed effects: *4682 firms in total.* Observations weighted by full-time equivalents. Cluster-robust standard errors in brackets, clustered at the firm*occupation*start-month level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Appendix D. Online Appendix

D.1. Validation of skill measures

PIAAC skill use data from Denmark is used to validate the skill measures derived from the job vacancy data.[18] However, PIAAC data is only available from interviews conducted in 2011 and 2012, and with 4-digit DISCO08 codes. Since 6-digit DISCO08-codes are not available, the above-mentioned conversion to 228 time consistent occupational groups cannot be undertaken. Thus, for the validation exercise described below, I first instead limit my sample to job spells commencing in January 2010 or thereafter.

In PIAAC, skill use is measured in five discrete values, for example: Never=1, Less than once a month=2, Less than once a week but at least once a month=3, At least once a week but not every day=4, Every day=5. I rescale this to a measure between 0 and 1, and collapse to get mean skill use at the 4-digit occupational level.

Next, the PIAAC skill use measures are merged onto the estimation sample from above, and the correlations with the binary skill measures and standardised, continuous skill measures are reported in Table D.1 and Table D.2 respectively.

Generally, the correlation coefficients confirm that the skills measures derived from the job postings data are indicative of reported skill usage in PIAAC at the occupational level. For example, see the large correlation coefficient between "Problem solving - Complex problems" skill use in PIAAC and the "cognitive" skill measures derived from keywords in job posts. However, notice that there are some inconsistencies between Table D.1 and Table D.2. For example, for "writing/language" skills, the sign of the correlation coefficients with the various "Literacy" skill use measures from PIAAC is generally positive when consider the binary skill measures (Table D.1), but not when considering the continuous skill measures (Table D.2). Thus, the binary skill measures seem to capture skill use better, possibly because of the skill measures are suffer from noise when the fraction of keywords indicative of certain skills is very low. At the same time, coefficients on the binary skill measures are also easier to interpret in regression analyses, and thus, the main parts of this paper report results using the binary skill measures, although the results generally are robust to the use of the continuous skills measures, see Appendix C

[18] For more details on PIAAC, see https://www.oecd.org/skills/piaac/ or, for example, Allen, Levels, and Van Der Velden (2013)

Table D.1: Correlations between binary skill measures and PIAAC skill use measures

	(1) Cognitive	(2) Social	(3) Character	(4) Writing/language	(5) Customer Service	(6) Management	(7) Financial	(8) Computer (general)	(9) Computer (specific)
Time cooperating with co-workers	0.0262***	0.0725***	-0.000626	-0.0311***	-0.111***	0.0859***	0.00400**	0.0571***	-0.0248***
How often - Sharing work-related info	0.139***	0.149***	-0.113***	0.0331***	-0.136***	0.137***	0.0769***	0.156***	0.0415***
How often - Teaching people	0.216***	0.180***	-0.190***	0.0903***	-0.229***	0.188***	0.0420***	0.135***	0.0614***
How often - Selling	-0.0487***	-0.0800***	0.159***	-0.149***	0.579***	-0.0682***	0.0339***	-0.00533***	0.00256*
How often - Advising people	0.157***	0.207***	-0.112***	-0.0117***	0.0999***	0.176***	0.101***	0.0908***	-0.00202
How often - Planning own activities	0.229***	0.224***	-0.201***	0.146***	-0.231***	0.224***	0.179***	0.121***	0.130***
How often - Planning others activities	0.120***	0.103***	-0.144***	0.0599***	-0.171***	0.181***	0.0751***	0.0999***	0.0252***
How often - Organising own time	0.282***	0.171***	-0.199***	0.123***	-0.153***	0.266***	0.270***	0.190***	0.181***
How often - Influencing people	0.165***	0.207***	-0.112***	0.0230***	0.0223***	0.166***	0.0805***	0.0974***	0.0326***
How often - Negotiating with people	0.188***	0.161***	-0.0978***	0.0614***	0.106***	0.190***	0.193***	0.103***	0.0713***
Problem solving - Simple problems	0.212***	0.195***	-0.168***	0.0419***	-0.106***	0.174***	0.111***	0.193***	0.102***
Problem solving - Complex problems	0.343***	0.196***	-0.240***	0.141***	-0.170***	0.271***	0.242***	0.272***	0.232***
How often - Working physically for long	-0.333***	-0.119***	0.199***	-0.190***	0.0592***	-0.271***	-0.348***	-0.292***	-0.252***
How often - Using hands or fingers	-0.165***	-0.0128***	0.0802***	-0.115***	-0.106***	-0.105***	-0.203***	-0.171***	-0.171***
Literacy - Read directions or instructions	0.238***	0.165***	-0.212***	0.0466***	-0.152***	0.178***	0.146***	0.182***	0.0674***
Literacy - Read letters memos or mails	0.329***	0.243***	-0.253***	0.154***	-0.166***	0.297***	0.269***	0.254***	0.155***
Literacy - Read newspapers or magazines	0.312***	0.164***	-0.206***	0.116***	-0.0308***	0.248***	0.251***	0.261***	0.182***
Literacy - Read professional journals or publications	0.353***	0.157***	-0.264***	0.140***	-0.187***	0.282***	0.243***	0.244***	0.143***
Literacy - Read books	0.214***	0.120***	-0.163***	0.128***	-0.283***	0.107***	0.0126***	0.127***	0.0560***
Literacy - Read manuals or reference materials	0.288***	0.121***	-0.217***	0.0427***	-0.115***	0.209***	0.154***	0.260***	0.129***
Literacy - Read financial statements	0.0870***	0.0155***	0.0449***	-0.00703***	0.288***	0.0998***	0.254***	0.0921***	0.0925***
Literacy - Read diagrams maps or schematics	0.407***	0.0436***	-0.236***	0.113***	-0.0403***	0.269***	0.324***	0.370***	0.266***
Literacy - Write letters memos or mails	0.316***	0.246***	-0.264***	0.165***	-0.207***	0.299***	0.285***	0.254***	0.158***
Literacy - Write articles	0.209***	0.0737***	-0.179***	0.137***	-0.100***	0.203***	0.199***	0.207***	0.154***
Literacy - Write reports	0.198***	0.225***	-0.205***	0.0394***	-0.246***	0.182***	0.0803***	0.0738***	0.0161***
Literacy - Fill in forms	0.159***	0.0940***	-0.118***	-0.0363***	0.0247***	0.188***	0.166***	0.119***	-0.0208***
Numeracy - How often - Calculating costs or budgets	0.0610***	-0.0502***	0.103***	-0.0537***	0.439***	0.0266***	0.187***	0.0934***	0.0896***
Numeracy - How often - Use or calculate fractions or percentage	0.320***	-0.00366**	-0.0907***	0.0334***	0.201***	0.189***	0.309***	0.294***	0.220***
Numeracy - How often - Use a calculator	0.267***	-0.0155***	-0.0575***	0.0000838	0.276***	0.176***	0.313***	0.278***	0.190***
Numeracy - How often - Prepare charts graphs or tables	0.307***	0.0660***	-0.184***	0.138***	-0.113***	0.235***	0.250***	0.269***	0.256***
Numeracy - How often - Use simple algebra or formulas	0.368***	-0.0102***	-0.157***	0.0688***	0.0209***	0.213***	0.289***	0.329***	0.266***
Numeracy - How often - Use advanced math or statistics	0.326***	-0.0604***	-0.165***	0.120***	-0.0870***	0.165***	0.234***	0.323***	0.280***
ICT - Internet - How often - For mail	0.351***	0.224***	-0.257***	0.180***	-0.172***	0.324***	0.303***	0.278***	0.187***
ICT - Internet - How often - Work related info	0.376***	0.211***	-0.255***	0.167***	-0.131***	0.324***	0.298***	0.291***	0.208***
ICT - Internet - How often - Conduct transactions	0.218***	0.0452***	-0.0663***	0.0659***	0.152***	0.157***	0.279***	0.226***	0.223***
ICT - Computer - How often - Spreadsheets	0.293***	0.0710***	-0.146***	0.160***	0.0153***	0.247***	0.345***	0.292***	0.281***
ICT - Computer - How often - Word	0.343***	0.202***	-0.258***	0.202***	-0.177***	0.314***	0.305***	0.266***	0.196***
ICT - Computer - How often - Programming language	0.257***	0.00397**	-0.112***	0.0952***	-0.0272***	0.123***	0.164***	0.301***	0.404***
ICT - Computer - How often - Real-time discussions	0.289***	0.0611***	-0.150***	0.163***	-0.0106***	0.168***	0.249***	0.313***	0.398***
Observations	636845	636845	636845	636845	636845	636845	636845	636845	636845

Note: Observations weighted by full-time equivalents.

Table D.2: Correlations between standardised, continuous skill measures and PIAAC skill use measures

	(1) Cognitive	(2) Social	(3) Character	(4) Writing/language	(5) Customer Service	(6) Management	(7) Financial	(8) Computer (general)	(9) Computer (specific)
Time cooperating with co-workers	-0.0107***	0.0305***	0.00824***	-0.133***	-0.138***	0.0921***	-0.0690***	0.0176***	0.0145***
How often - Sharing work-related info	0.117***	0.125***	-0.106***	-0.124***	-0.179***	0.139***	-0.00218	0.0718***	0.0749***
How often - Teaching people	0.197***	0.178***	-0.192***	-0.0826***	-0.286***	0.131***	-0.0240***	0.0372***	0.0762***
How often - Selling	-0.0820***	-0.0308***	0.125***	-0.0434***	0.549***	-0.0725***	0.0621***	0.0242***	-0.0204***
How often - Advising people	0.106***	0.217***	-0.111***	-0.152***	0.0224***	0.143***	0.0465***	0.000614	0.0103***
How often - Planning own activities	0.226***	0.252***	-0.164***	0.000796	-0.215***	0.199***	0.171***	0.0707***	0.117***
How often - Planning others activities	0.110***	0.124***	-0.140***	-0.0472***	-0.192***	0.175***	0.00614***	0.0212***	0.0278***
How often - Organising own time	0.280***	0.191***	-0.183***	-0.0316***	-0.112***	0.262***	0.239***	0.138***	0.168***
How often - Influencing people	0.126***	0.216***	-0.122***	-0.142***	-0.0598***	0.127***	0.0122***	0.00980***	0.0427***
How often - Negotiating with people	0.161***	0.191***	-0.0978***	-0.0595***	0.0869***	0.186***	0.150***	0.0250***	0.0581***
Problem solving - Simple problems	0.181***	0.188***	-0.182***	-0.136***	-0.166***	0.162***	0.0493***	0.114***	0.121***
Problem solving - Complex problems	0.342***	0.208***	-0.243***	-0.0525***	-0.170***	0.272***	0.166***	0.194***	0.227***
How often - Working physically for long	-0.341***	-0.145***	0.212***	-0.0749***	0.00254*	-0.297***	-0.334***	-0.191***	-0.232***
How often - Using hands or fingers	-0.202***	-0.0536***	0.111***	-0.0911***	-0.128***	-0.160***	-0.197***	-0.100***	-0.137***
Literacy - Read directions or instructions	0.218***	0.145***	-0.193***	-0.114***	-0.172***	0.162***	0.0629***	0.0965***	0.0943***
Literacy - Read letters memos or mails	0.307***	0.251***	-0.236***	-0.0436***	-0.166***	0.273***	0.218***	0.138***	0.159***
Literacy - Read newspapers or magazines	0.294***	0.185***	-0.219***	-0.0377***	-0.0430***	0.247***	0.213***	0.148***	0.173***
Literacy - Read professional journals or publications	0.347***	0.173***	-0.292***	-0.0704***	-0.213***	0.249***	0.167***	0.108***	0.134***
Literacy - Read books	0.213***	0.104***	-0.209***	-0.0484***	-0.353***	0.0576***	-0.0322***	0.00568***	0.0623***
Literacy - Read manuals or reference materials	0.273***	0.105***	-0.219***	-0.105***	-0.136***	0.181***	0.0419***	0.187***	0.155***
Literacy - Read financial statements	0.0550***	0.0600***	0.0390***	0.0317***	0.338***	0.135***	0.346***	0.0763***	0.0568***
Literacy - Read diagrams maps or schematics	0.415***	0.0542***	-0.266***	0.00388**	-0.00636***	0.285***	0.254***	0.287***	0.249***
Literacy - Write letters memos or mails	0.299***	0.252***	-0.245***	-0.0377***	-0.201***	0.287***	0.230***	0.134***	0.162***
Literacy - Write articles	0.207***	0.105***	-0.208***	0.0122***	-0.0982***	0.224***	0.134***	0.117***	0.135***
Literacy - Write reports	0.170***	0.205***	-0.186***	-0.157***	-0.284***	0.130***	0.00431***	0.0226***	0.0329***
Literacy - Fill in forms	0.111***	0.0725***	-0.107***	-0.138***	0.0190***	0.157***	0.105***	0.0506***	0.00549***
Numeracy - How often - Calculating costs or budgets	0.0415***	0.00255*	0.0750***	0.0262***	0.458***	0.0544***	0.256***	0.0936***	0.0536***
Numeracy - How often - Use or calculate fractions or percentage	0.319***	0.0137***	-0.120***	-0.00712***	0.228***	0.193***	0.320***	0.226***	0.197***
Numeracy - How often - Use a calculator	0.259***	0.0160***	-0.0798***	-0.00900***	0.309***	0.197***	0.327***	0.233***	0.170***
Numeracy - How often - Prepare charts graphs or tables	0.332***	0.0936***	-0.199***	0.0288***	-0.0783***	0.227***	0.202***	0.223***	0.234***
Numeracy - How often - Use simple algebra or formulas	0.381***	-0.000416	-0.185***	0.00376**	0.0605***	0.210***	0.272***	0.270***	0.250***
Numeracy - How often - Use advanced math or statistics	0.390***	-0.0315***	-0.220***	0.0375***	-0.0645***	0.147***	0.168***	0.257***	0.251***
ICT - Internet - How often - For mail	0.332***	0.233***	-0.239***	-0.00506***	-0.150***	0.306***	0.254***	0.161***	0.187***
ICT - Internet - How often - Work related info	0.348***	0.215***	-0.247***	-0.0185***	-0.115***	0.298***	0.239***	0.172***	0.206***
ICT - Internet - How often - Conduct transactions	0.188***	0.0699***	-0.0697***	0.0548***	0.207***	0.183***	0.331***	0.192***	0.199***
ICT - Computer - How often - Spreadsheets	0.302***	0.106***	-0.148***	0.0898***	0.0859***	0.273***	0.330***	0.244***	0.254***
ICT - Computer - How often - Word	0.337***	0.226***	-0.249***	0.0241***	-0.153***	0.307***	0.265***	0.149***	0.187***
ICT - Computer - How often - Programming language	0.283***	0.00551***	-0.121***	0.0594***	0.0273***	0.124***	0.105***	0.334***	0.401***
ICT - Computer - How often - Real-time discussions	0.308***	0.0866***	-0.162***	0.105***	0.0410***	0.198***	0.220***	0.300***	0.375***
Observations	636845	636845	636845	636845	636845	636845	636845	636845	636845

Note: Observations weighted by full-time equivalents.

D.2. DISCO-crosswalk

The crosswalk between 4-digit DISCO88 codes and 6-digit DISCO08 codes can be found here: https://www.dropbox.com/s/xmvzwpmwvqnntft/disco-crosswalk.pdf?dl=0. The results of applying the crosswalk is 228 time consistent occupational groups.