Understanding skill demand and skill usage on the job: Evidence from job posts and individual-level data on skill usage*

Moira Daly† Fane Groes‡ Mathias Fjællegaard Jensen§

March 3, 2023

Abstract

Skill requirements in a job post reflect an employer's "wish list," but do they also reflect skills used on the job by the hired worker? We compare skill measures derived from the text of online job posts with individual-level data from the Danish Labour Force Survey (LFS) in which participants report their main skills used on the job as free text. By identifying individual workers from the LFS who can be matched to a job post, we validate that the extensive margin skills measures derived from job postings data reflect main skills used on the job. Thus, using job postings data to analyze skill usage on the job is generally a valid empirical strategy. However, we also show that heterogeneity in returns to skills is missed if only the extensive margin of skill demand is considered.

Keywords: skills, tasks, job postings, Labour Force Survey

JEL classifications: J24

---

*We are grateful for the help and comments of Lisa Kahn and Ning Zhang. We would also like to thank our research assistant at the time, Oliver-Alexander Press. This work was supported by the Novo Nordisk Foundation grant number NNF16OC0021056 and the Kraks Fond project number 301061. Jensen thanks the Independent Research Fund Denmark for financial support through grant number 1058-00011B.

†Corresponding author, Copenhagen Business School, Porcelænshaven 16A, Frederiksberg, Denmark (email: moda.eco@cbs.dk, phone: +45 38152351).

‡Copenhagen Business School, Porcelænshaven 16A, Frederiksberg, Denmark (email: fg.eco@cbs.dk).

§University of Oxford, 10 Manor Rd, Oxford, OX1 3UQ, UK (email: mathias.jensen@economics.ox.ac.uk).

1 Introduction

An extensive literature has studied the dynamics of task-specific skills in Europe, and particularly, in the United States (e.g., Autor et al., 2003; Spitz-Oener, 2006; Black & Spitz-Oener, 2010; Goos et al., 2014; Arntz et al., 2016; Beaudry et al., 2016). This early literature generally relies on measures of skills at the occupation level, but in the last decade, a new data source has emerged from the online posting of job vacancies. Machine reading of job posts has allowed researchers to explore the variation in the demand for task-specific skills within firms and occupations, as well as across time. Task-specific skills refer to skills related to certain tasks, such as social and cognitive skills, and not education levels. A new and rapidly expanding literature has used text from job posts to, for instance, understand the variation in demand for skills within occupations and the effect of this variation on workers' pay (e.g., Modestino et al., 2016; Deming & Kahn, 2018; Hershbein & Kahn, 2018; Marinescu & Rathelot, 2018; Grinis, 2019; Atalay et al., 2020; Blair & Deming, 2020; Deming & Noray, 2020; Modestino et al., 2020; Alekseeva et al., 2021; Daly et al., 2022; Braxton & Taska, 2023).[1]

This nascent literature relies on proprietary job vacancy data, which, by its nature, capture skill demand and not necessarily skill usage. Skills listed in a job vacancy may be an employer's "wish list," but not necessarily reflect skills used on the job by the hired worker. Although validation exercises have been performed at the occupational level (Hershbein & Kahn, 2018), the degree to which skill demand advertised by firms captures the skills used by workers at the individual level is still unclear, no doubt due to a lack of data. At the same time, recent papers match vacancy data and administrative data at increasingly granular levels (see e.g., Kettemann et al., 2018; Jensen, 2021; Bagger et al., 2022; Fluchtmann et al., 2022; see also Kircher, 2022, for a review of studies linking job seekers to vacancies). Thus, validations of the data at more granular levels are also warranted. By linking job posts at the pseudo-individual level with the self-reported skill usage of workers hired for the posted jobs, we are able to describe both the relationship between the skill demand as advertised by employers and the main skills workers report using on the job, and each of their effects on wages.

To capture skills usage on the job, we use the Danish Labour Force Survey (LFS), the country-specific version of the widely available non-proprietary European Union Labour Force Survey (EU LFS). Survey respondents are asked about the main tasks they perform on the job, and their free-text answers are recorded.[2] We extract these free-text answers from the LFS and, using a similar approach to that of Deming and Kahn (2018), categorize keywords from the text into 9 categories

[1] A number of other papers also analyze job vacancy data, but focus less on skill demand, see e.g. Adams et al. (2020); Azar et al. (2020); Clemens et al. (2020); Forsythe et al. (2020); Javorcik et al. (2020); Bagger et al. (2022).

[2] These data, collected from all EU member states, 4 candidate countries, and 3 countries of the European Free Trade Association, have the potential to serve as an important source of skills data.

of task-specific skills: Cognitive, Social, Management, Financial, Computer (general), Computer (specific), Writing/Language, Customer Service, Character.[3] We interpret a reported job task as the utilization of a task-specific skill, and thus, in order to be consistent in our terminology, we refer to the measures extracted from the LFS as the main skills used on the job.

Next, we apply the same categorization to the text of online job posts from Denmark. We compare the measures of task-specific skills extracted from job posts to the individual task-specific skills reported in the LFS.[4] To our knowledge, we are both the first to use the reported task-specific skills from any of the EU LFS surveys and to link these to skills demanded in job postings. As the current literature using job postings data focuses primarily on the US, we believe that, given that such an exercise is not possible in the US, Denmark provides an ideal environment to conduct this exercise as many features of the Danish labor market resemble those of the US labor market (see e.g., Botero et al., 2004; Groes et al., 2015; Heckman & Landersø, 2022). Particularly relevant for this analysis is the fact that Denmark and the US share very similar levels of labor market turnover rates, employment protection, and economic freedom (see e.g., Kreiner & Svarer, 2022). At the same time, data derived from online job postings are becoming increasingly available across countries, including many European and OECD countries, which makes a validation of the data increasingly relevant.

We find that a significant proportion of workers report using only a single main skill that falls under one of the commonly cited skill categories in the job postings literature.[5] On the other hand, employers demand skills from six different skill categories on average. Given that there is no limit on the heterogeneity of skills that employees report when describing their main skills, one interpretation of these findings is that workers are more specialized than what the skill indicators derived from job postings may suggest: whereas the job posting skills may capture the extensive margin of skills used on the job, the LFS measures capture the intensity of skills used on the job by only including the most important skills. We cannot separately distinguish between concepts of frequency of skill use and importance of skill use. Some workers may deem that their main skill is the task they perform most frequently, while others may consider it to be the task that they think is most important for their work. Nonetheless, we believe that understanding how a measure of individual-level, intensive skill usage relates to existing skills measures derived from job postings is a valuable addition to the literature.

Next, we show the degree to which measures of task-specific skills derived from job postings correspond with skills used on the job as reported by workers in the LFS. Workers who report a

[3]We list the most common keywords for each skill category in Tables A.1 and A.2.

[4]We focus on task-specific skills, meaning the type of skills that are associated with specific tasks, such as social skills, cognitive skills, and computer skills.

[5]Almost three-quarters of all workers report using a main skill that falls within the skill categories often used in the job posting literature.

particular main skill are extremely likely to be in a job that advertised for that particular skill. In addition, we find positive and significant correlations between job post skills and self-reported main skills from the LFS, the only exception being character skills. We also find that about one-tenth of workers report mainly using a skill type that their employer did not include in the job posting. An investigation of the relative employer-employee match quality of this group finds no evidence of shorter match durations or negative wage effects suggesting that mismatch among skill supply and demand is not substantial. Based on this evidence, we continue our analysis under the assumption that skill demand as captured by job posts and main skills reported on the LFS can be interpreted as extensive and intensive measures of skill usage, respectively.

We then estimate standard wage regressions and explore the returns to skills on the extensive margin with and without including on-the-job measures of skill intensity (as captured by the LFS). Including measures of on-the-job skill usage greatly increases the model's ability to explain the variation in wages when individual controls are not included. The inclusion of skill intensity measures does not qualitatively affect the estimates of the extensive margin return to advertised skills. On the other hand, much of the variation in wages explained by the intensity of skill usage is absorbed once individual controls are included in the regression. Taken together, these results suggest that the precision of extensive margin skill return estimates can increase noticeably if intensive skill measures are included, but their inclusion is less necessary when sufficient individual controls are available.

For several skills, we find large differences in the estimated returns to skills derived from job posts and from the LFS. In particular, individuals in jobs that advertise for cognitive and management skills are rewarded substantially more if they use these skills intensively as measured by the LFS. On the other hand, workers in jobs that advertise for writing/language and customer service are severely penalized for using these skills intensively. Our results highlight the fact that although estimates of the return to skill on the extensive margin can accurately describe average returns to skills, workers who intensively use these skills can substantially benefit or suffer, depending on the skill considered.

While much of the literature has focused on the returns to skill demand on the extensive margin (e.g., Deming & Kahn, 2018), few studies have tried to quantify the intensive margin of skill demand, and fewer still have looked at both the extensive and intensive margin of skills. We believe our ability to study both the intensive and extensive margin of task-specific skills at the job level is new to the literature. We proceed by providing details on the data in Section 2, the results in Section 3, and conclude in Section 4.

2 Data

2.1 Danish Job Postings

The online job postings data (JP) from 2007–2017 are supplied by the consultancy firm HBS Economics (HBS) and cover the near universe of publicly accessible online job posts in Denmark. The Danish JP are generally analogous to the equivalent US data supplied by Burning Glass Technologies.[6] However, relative to the US, Denmark has a large public sector that is legally obligated to post all jobs online.[7] To facilitate comparison with the US job postings literature, we consider only positions advertised by private firms, and we concentrate on job posts for occupations that are well represented in both the job posts and the LFS data: professionals, technicians and associate professionals, clerical workers, and service and sales workers.[8]

The JP include keywords from each job post, a posting date, and an occupational code. In addition, the JP contain a firm identifier that allows us to match the data with Danish employer-employee matched registers provided by Statistics Denmark. By matching a job post to a firm, we can further match a specific job post with employees who recently started working within the same firm and within the corresponding occupation.

Like Deming and Kahn (2018), we focus on the extensive margin in the JP: indicators are created at the job-posting level capturing whether or not a posting contains a keyword in a particular skill category. In Panel A of Table 1, Column 1 presents the fraction of job posts with each of the categorized task-specific skills from all of the private-sector JP from the period. Cognitive skills are one of the least-occurring skills, included in 61% of the job posts, whereas 87–97% of the job posts include customer service, social, or character skills.

2.2 Labour Force Survey Text Data

Since the 1980s, EU member states have administered the Labour Force Survey based on common survey guidelines to enable cross-country comparisons. About 1.5 million people were surveyed quarterly in 2018.[9] One of the variables in the EU LFS is an occupational code, based on the ISCO standards. To classify individuals into the correct occupation, the national statistical institutes collect information on job titles, and more importantly, job tasks. More specifically, in the Danish

[6] We have purchased the data through the Danish consultancy firm, HBS Economics. See appendix for more details.

[7] Approximately 30% of workers are in the public sector.

[8] We exclude blue collar occupations as in Deming & Kahn (2018). We further exclude managers (ISCO occupation 1) because these jobs are not well represented in the LFS data. We include 1-digit ISCO occupation codes of 2-5.

[9] The EU LFS is found here: `https://ec.europa.eu/eurostat/web/microdata/european-union-labour-force-survey`

LFS, respondents are asked to "Describe the specific main tasks in your job."[10] Although the national statistical institutes collect and process data on job titles and job tasks, these free-text data tend not to be available to researchers. Uniquely, we have access to the Danish text data from the LFS from 2007–2017. Importantly, Statistics Denmark supplies this data with personal identifiers so it can be easily linked to Danish register data, allowing us to determine the firm for which a surveyed individual works, the wage they earn, and many other worker-firm characteristics that are not available in the LFS.

In order to compare how skills from the JP reflect on-the-job skill usage as captured by the LFS, we consider workers in the LFS who started their job within the last year, as skill usage may change with tenure. Column 2 of Table 1, Panel A, presents summary statistics of the categorized task-specific skills from the LFS data. The lower incidence of task-specific skills is immediately clear—a consequence of the LFS capturing a measure of main skill(s) only. In contrast, the measures of task-specific skills derived from the job postings data demonstrate that employers almost always list at least one "character" and one "social" keyword.

2.3 Linking the JP and LFS

Each job posting in the JP is matched with one or more individuals in the registers, following the matching procedure described by Jensen (2021) and Daly et al. (2022). We identify individuals in the employer-employee linked register data who have recently started a new job (either in a new occupation or in a new firm). We match individuals to a job post in the same firm-occupation cell if the job was posted in the month in which they started their new job or a maximum of four months prior. We call such a match a pseudo-individual match. This means that a person starting in a given firm-occupation cell can also be linked to multiple job posts if more than one job post has been posted in the same firm-occupation cell within the five-month window. If a person is matched to more than one job post, the indicator equals 1 if a skill is mentioned in any of the job posts.

We are able to match about one-fifth of all new jobs recorded in the Danish register data to job posts in the relevant occupation, reflecting the fact that many private companies do not advertise all new jobs, especially new jobs resulting from occupational changes within a firm. Our final sample links individuals with the firm and job posting to which they responded, and includes individual responses to the LFS. From the job postings, we derive skill indicators associated with

[10] In Danish, the specific questions on the survey include: 1) B2Stil: "Hvad er din stillingsbetegnelse/titel?" and 2) B2StilA: "Beskriv de konkrete hoved-arbejdsopgaver i din stilling." The question included in the German Mikrozensus 2021 is: "Please describe your current work in keywords." The question included in the UK LFS is: "What did you mainly do in your job?", and in the Swedish LFS: "What are your main tasks?" See national LFS questionnaires by year here: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=EU_labour_force_survey_-_documentation

each job match, and from the register data, we obtain other characteristics of the job during the respondent's first year in the job (e.g., wages and hours worked).

To understand the representativeness of the JP and LFS data, we report various summary statistics in Appendix Table B.1. We find that age and location vary slightly between the samples, so we control for these factors in our empirical specification. We find that the matched JP-LFS (henceforth the Estimation Sample) is representative of the overall population of new jobs, of the JP-population matched data, and of the full set of new jobs identified in the LFS. As we would expect given that the LFS is a representative sample of Danish workers, the match rate between the full population of new jobs to the JP is similar to the match rate between the subset of new jobs that are sampled in the LFS and the JP. See Appendix A for more details on the data.

2.4 Interpretation of Task-Specific Skill Measures

It is important to recognize how the measures of task-specific skills differ depending on whether they are derived from the LFS or from the job postings data. To fix ideas, we consider a very simplified setting in which some workers are hired into jobs that demand a particular skill and, for a subset of these workers, that skill is intensively used. Let s^{JP} and s^{LFS} be the observed skill measures captured by job postings data and the LFS, respectively:

$$s^{JP} = \mathbb{1}(s^D + e^{JP} > a)$$
$$s^{LFS} = \mathbb{1}(s^S + e^{LFS} > b)$$

where s^D corresponds to the true unobserved employer skill demand, s^S corresponds to the true unobserved skill supply of workers, and e^{JP} and e^{LFS} correspond to measurement error in both of these observed variables, respectively. This formulation highlights that there are generally three ways a potential misalignment of these two skill measures could occur.

First, there may be labor market frictions generating mismatch between skill supply and demand: firms may not be able to hire workers possessing the skills they demand, and instead hire workers with a different skill profile, i.e. $s^S \neq s^D$. Second, it is likely that $a < b$: workers are specifically asked to report only their main skills; in contrast, employers may list skills that are desirable but not necessary in addition to those that are necessary, given their small cost of doing so. Third, even in the case that there are no labor market frictions ($s^S = s^D = s^*$), and if the skill indicators capture the same intensity of the underlying skill ($a = b$), measurement error in either or both of the observed skill measures could lead to a misalignment between s^{JP} and s^{LFS}. Measurement error could arise on the demand side because certain skill categories are so fundamentally part of a job that explicitly stating the requirement may be unnecessary. On the supply side, measurement error could come from how a worker perceives a main task, or from the distinction between tasks and skills. Although this distinction is not often emphasized in the literature, in our context it may be more of an issue. For example, an employer may list "detail-oriented" as a skill required for the job, but when asked about skills usage (tasks) on the job, an employee would likely not list "detail-oriented".

In the absence of mismatch and measurement error, we can write our skill measures as:

$$\begin{aligned} s^{JP} &= \mathbb{1}(s^* > a) = s \\ s^{LFS} &= \mathbb{1}(s^* > a) \cdot \mathbb{1}(s^* > b) = s \cdot m \end{aligned}$$

where s is an indicator equal to 1 if a job requires a skill of at least level a, and m is an indicator taking the value of 1 if that skill is particularly important, above level b. If we believe a is close to 0, we can interpret s as the extensive margin of the skill and m as a measure of skill intensity. In this case, m corresponds to the fraction of workers hired in jobs requiring skill s who use that skill intensively. If we regress s^{LFS} on s^{JP}, the slope should recover this fraction.

The availability of a measure of on-the-job skill intensity allows us to investigate how sensitive wage regressions are to using just extensive margins defined from skills listed in job postings rather than, or in addition to, regressions that also include a measure of skill intensity. In this simplified environment, we can write the following population model:

$$w = \beta_0 + \beta_1 s + \beta_2 (s \cdot m) + u \tag{1}$$

where w is log wages, and under our simplifying assumptions, u captures all other unobserved determinants of wage such that $E(u \mid s^*, m^*) = 0$. β_1 is the return to the extensive margin of skill s, and β_2 is the additional return for using that skill intensively. Note that in this framework, there is no need to include m separately in the regression as individuals only intensively use the skill if they also use the skill at the extenisive margin.

In the data, there are individuals who use a skill intensively and work in a job that did not advertise for that skill ($s^{JP} = 0$ and $s^{LFS} = 1$). We conjecture that if this misalignment is mainly being driven mismatch, we should see evidence of negative effects of this misalignment on wages and on tenure as firms find workers with whom they are a better match. We explore this possibility further in the results section and find no evidence of either. Thus, we conclude that the relatively small misalignment is due to measurement error,

We then proceed to estimate a more general form of Equation 1 that allows each of the 9 possible main skills to be correlated with multiple skills advertised in the JP. Specifically, in order to understand the relationship between wage, indicators of advertised skill, and main skill, we estimate:

$$wage_i = \gamma_0 + \sum_{k=1}^{9}(\gamma_k s_{ki}^{JP} + \theta_k s_{ki}^{LFS}) + x_i\delta + \epsilon_i \tag{2}$$

where $wage_i$ is the natural logarithm of the hourly wage of worker i, γ_0 captures average earnings for those who work in jobs that require skills not captured by our categorization and who mainly use skills not captured by our categorization.[11] γ_k captures the average wage premium (or penalty) of those who work in jobs that advertised for skill k, whereas θ_k captures the average wage premium (or penalty) of those who mainly use skill k. Finally, x_i always contains year and municipality indicators. We also estimate specifications in which individual controls are included: age, age^2, experience, experience^2, years of education fixed effects, and a gender dummy.

3 Results

3.1 Validation of Skill Measures

Column 1 of Panel B, Table 1, reports the frequency of having only one, two, and three different types of skills mentioned in a job post. The same frequencies, but now referring to the main skills reported in the LFS, are shown in Column 2. Almost all job posts contain at least one skill that falls within the categories we consider. On the other hand, in the LFS, almost three-quarters of the workers report at least one main skill that fall within one of our 9 skill categories. As workers are free to report whatever they consider their main skill used a certain degree of measurement error is expected. For example, a phlebotomist may report that their main skill is to "take blood samples," something not captured by our skills measures. Upon closer inspection, this type of measurement error appears to be driving the issue: a frequent, non-categorized reported main skills is, for example, "cash register" (translated to English).

Of the workers reporting at least one main skill in the LFS, 72% of them report only one of the of nine skills we consider. About 22% report skills that fall across two main skill categories, and fewer than 8% of workers report skills that fall across three or more skill categories. In job posts, employers tend to mention skills that fall across more skill categories. Only 13% of employees work in jobs for which three or fewer skills were listed in the corresponding job post. On average, employers demand skills that fall across six different skill categories. Because employees are free to list many main skills in the LFS, a possible interpretation of these facts is that workers are more specialized than what the skill indicators derived from job postings may suggest. In other words, the skill intensities within each of the employer-demanded skill categories may vary substantially.

[11]We follow Deming & Kahn (2018) in our skill categorization. See sections 2.1, 2.2 and the Appendix for more discussion of our skill categorization.

The job posting skills then capture the extensive margin of skills used on the job, and the LFS main skills capture the intensity of skills used on the job.

Next, we verify that if an individual lists a particular main skill in the LFS survey, then the job for which they are hired also requires that task-specific skill. Table 2, Panel A, presents the probability that a job posting requires a particular skill, conditional on the LFS respondent reporting that task-specific skill. For instance, in Column 2, we see that almost 85% of those LFS respondents who stated that one of their main skills was cognitive skills hold a job that advertised for cognitive skills. Regardless of the skill category, we can clearly see that for the vast majority of individuals, the main skill they use on the job corresponds, at the individual level, to a skill listed in the job posting to which they applied. This is our first piece of evidence that skill measures derived from job posts in fact capture skills used on the job.[12] Yet, as these conditional probabilities are not 1, there are employees reporting in the LFS that they use a main skill which their employer did not explicitly include in the job posting. Pooling across LFS main skill types, we find that this is true for 10% of workers.

Table 2, Panel B, presents the reverse conditional probability: the probability that an LFS respondent lists a task-specific skill conditional on the job post requiring that skill. For instance, about 8.5% of those who have a job that sought cognitive skills primarily use cognitive skills on the job. On the other hand, almost 40% of those who work in a job that advertised for customer service skills state that they mainly handle customer service. Comparing these conditional probabilities to the unconditional probabilities in Table 1, Column 2, we see that the former is greater than the latter for all skills but character skills. This is our second piece of evidence that measures of skill demand derived from job postings reflect skills used on the job.

Table 3 continues this exercise by presenting the results of regressing an indicator of the LFS task-specific main skill measures on the skill indicators derived from the job posting data. If LFS respondents had been asked to report all of the skills used on the job (as opposed to just main skills), we would expect that coefficients along the diagonal would be close to 1, in the absence of measurement error and mismatch. However, given that the LFS captures an indicator for (only) the most important skills, this need not be the case.[13] For instance, we learn from Column

[12] Another approach would be to calculate a "most important" skill from the JP and use it for comparison purposes. The issue with this approach is how exactly to construct such a variable. In the LFS data, the importance is implied in the question. In the JP, one might be tempted to assume that word frequency corresponds to importance; however, if this were true, the most important skill across occupations would be character—not necessarily because this skill is relatively more important, but because this skill category includes relatively more words.

[13] Table 3 is a correlation of the two skill measures that can give different results at the individual and occupational level. For instance, at the individual level, the correlation between skill measures would be low if few workers have a given skill as their main skill while the skill is relatively common in the JP. However, at the occupational level, the correlation between the same skill measures can be high if the occupational fractions of the few workers with the skill from the LFS data correlate across occupations with the higher occupational fraction of workers with the skill in the JP. One example of this could be if a main skill (from LFS) was used in two occupations with the probability

1, Table 3, that workers who intensively use cognitive skills represent about 4.5% (significantly different from 0) of the employees working in jobs that required cognitive skills, holding other extensive skill requirements constant. As all coefficients along the diagonal are positive, except for character skills, we take this as our third piece of evidence that skills advertised by employers are actually being used on the job.

Table 3 also illustrates how skill bundles sought by employers vary according to the main skills performed on the job. In Column 1, we see that workers who mainly use cognitive skills are more likely to hold a job that advertised for management skills, but less likely to be in a job that advertised for customer service skills. From Column 2, we see that workers who intensively use social skills are significantly more likely (about 2.6%) to be working for an employer who stated that their employees should possess cognitive skills and significantly less likely (3%) to be working for an employer seeking employees with financial skills.[14]

Next, we seek to better understand why 10% of employees report using a main skill not advertised by their employer in their corresponding job post. We hypothesize that if there is skill mismatch, this misalignment would likely imply a lower match quality: such matches would have shorter durations and/or employees would receive lower wages.[15] On the other hand, if workers correctly report the skills they intensively use on the job, but employers implicitly, rather than explicitly, state that a skill is demanded in the job post, we expect no negative effects on match duration or wages.

Table 4 presents the results of regressing various measures of match quality on an indicator of skill misalignment: whether or not a worker is mainly using a skill type on the job that was not advertised by their employer. We look at the effect of this misalignment on the probability of the match lasting more than 1, 2 and 3 years in Columns 1, 2 and 3 respectively and find small, statistically insignificant effects. In Column 4, we regress wages on the same indicator of skill misalignment and again find no significant difference. We conclude from this exercise that there is no evidence of skill mismatch among main skills, but rather that skills extracted from job posts involve some measurement error.[16]

---

of 0.01 and 0.02. If this same skill was used at the extensive margin (from the JP) in the same two occupations with probability 0.4 and 0.8, then the occupational correlation would be 1. However, the individual-level correlation could at most be 0.025, but it could also be zero or negative.

[14]By simulating an upper bound of the regression coefficients if all LFS skills were a subset of JP skills, i.e., with no measurement error or mismatch, we have a comparison for the size of the regression coefficients. We provide these simulated correlations in Appendix Table B.3. Compared to the simulated upper bound, we see that the unconditional regression coefficients represent around 45% of the maximum unconditional correlation.

[15]If the choice to report a skill as a main skill is driven only by the desire to be perceived in a particular way and is not correlated at all to an actual deprioritization of the true main skill, then we would not expect to see tenure or wage effects.

[16]We might be worried that attenuation due to measurement error is masking evidence of negative wage and tenure effects. But given the significant wage effects we find in Table 5 that are generally of the same magnitudes reported in the literature, we do not believe this is the case.

3.2 Skills and Wages

Given that we find no evidence of mismatch, we move on to further understand skill premia or penalties by considering both the extensive and intensive skill margins. Columns 1 and 3 of Table 5 present the results from estimating Equation 2 when restricting the δ_k to zero, whereas Columns 2 and 4 present estimates from unrestricted models. The first two columns present the results when year and municipality indicators, but not individual controls, are included whereas Columns 3 and 4 show the results when individual controls are also included.[17]

From Columns 1 and 3 in Table 5, which list the estimated wage skill premia at the extensive margin, we see that cognitive and management skills are associated with positive and significant wage returns whereas character skills are associated with significantly negative returns. Specifically, Column 1 shows that cognitive and management skills, on average, are associated with 11.4 and 5.5 log wage points (lwp) higher wages when we do not control for individual characteristics. In contrast, having character skills are associated with a 12.8 lwp lower wage. After controlling for individual characteristics in Column 3, the estimated returns decrease in absolute value (but remains statistically significant) demonstrating the importance of controlling for selection when estimating skill premia.[18] If we assume that firms advertise only the skills that will be used on the job, these coefficients are the weighted average of those who use these skills intensively and those who do not. In this sense, these estimates correspond to what is often reported in the literature.

In Columns 2 and 4, we include the main skill indicators from the LFS. Generally speaking, the inclusion of the main skill indicators has relatively little effect on the estimated coefficients on the JP skill measures, although the coefficients' absolute values decrease in all cases. In addition, the inclusion of the main skill indicators also greatly increases the explanatory power of the model when no individual controls are included. However, much of the variation in wages explained by the intensity of skill usage is absorbed once individual controls are included in the regression. This suggests that extensive margin skill returns can be more precisely estimated if intensive skill measures and individual controls are included; something that is rarely done in the existing literature using skill measures from job posts.

Finally, we highlight three ways in which the returns to extensive and intensive skills vary according to the skill considered. First, we see additional positive wage returns to cognitive and management skills when these skills are used intensively. When individual controls are not included, employees with jobs that advertised for, and mainly use, cognitive skills earn wages that are 17 lwp higher compared to employees who work in jobs that did not advertise for cognitive

[17]The individual controls include: age, age^2, experience, experience^2, years of education fixed effects, and a gender dummy.

[18]An interesting exercise to perform in this context would be to add occupation indicators. Jensen (2021) is able to further explore the effects of occupation and firm fixed effects on the returns to task-specific skills due to his larger estimation sample.

skills and who did not report using cognitive skills as a main skill. Employees who work in jobs that advertise for cognitive skills, but do not include cognitive skill as a main skill, earn wages that are only 10 lwp higher than those who work in jobs that do not advertise for or use cognitive skills. When individual controls are added to the regression, these returns more than halve to about 7 lwp and 4 lwp, respectively. We also see that individuals whose main skill is management receive higher wages, a wage premium at around 12 lwp. Recall that managers, as defined from occupation codes, are not included in this analysis due to poor representation in the LFS, suggesting that there are high returns to management skills prior to entering a management occupation.

Second, we find that those who work in a job that advertised for character skills, but who do not mainly use character skills, face wage penalties, but those in jobs that both advertise for character skills and who use character skills as a main skill face no such penalties. These results reflect the fact that, uniquely, character skills are negatively correlated in the LFS and JP (see Table 3).

Third, those working in jobs in which writing/language and customer service skills are used intensively recieve lower wages. However, employees holding jobs that advertised for these skills, but who do not report them as main skills, see no negative effects on wages. After controlling for individual characteristics, mainly using writing/language skills is associated with 7.7 lwp lower wages, and workers who mainly use customer service have 2.4 lwp lower wages on average.

4 Conclusion

The aim of this paper is to better understand both how well skill measures derived from job postings data capture skills used on the job, and next, the extent to which the availability of an intensive measure of skills, as opposed to just an extensive measure of skills, improve the estimates of returns to task-specific skills. We pursue this by comparing the demand of task-specific skills from job posts to individual self-reported main skill usage extracted from the Danish Labour Force Survey (LFS) for private employees over the 10-year period beginning in 2007. To our knowledge, we are the first to compare skills demanded in job posts with self-reported skills usage at such a granular level.

We explore the degree to which measures of task-specific skills derived from job postings correspond with skills used on the job as reported by workers in the LFS. We find that workers who report a particular main skill are extremely likely to be in a job for which that particular skill was demanded in the corresponding job post. Moreover, we generally find positive and significant correlations between skills derived from job posts and self-reported main skills from the LFS. We move on to investigate the relative employer-employee match quality and find no evidence of substantial mismatch between skill supply and demand, allowing us to propose a framework in which to understand the two sets of skill measures: we interpret skill demand as captured by job posts and main skills reported on the LFS as extensive and intensive measures of skill usage, respectively.

Our results from estimating standard wage regressions suggest that analyses estimating the extensive margin skill returns can be more precisely estimated if intensive skill measures and individual controls are included. We find several large differences between the extensive margin returns to advertised skills and the returns to skills mainly used on the job as captured by the LFS. For example, returns to cognitive and management skills increase with the intensity of their use, whereas writing/language and customer service skills are associated with lower wages when these skills are used more intensively. Our finding of consistently positive returns to cognitive skills is in line with the existing literature, e.g., Spitz-Oener (2006); Black & Spitz-Oener (2010); Beaudry et al. (2016); Deming & Kahn (2018); Atalay et al. (2020).[19]

Our findings suggest that the skills measures derived from job postings data typically used in the literature capture main skills used on the job, and thus, using job postings data to analyze skill usage on the job is generally a valid empirical strategy. However, a rich dimension of heterogeneity in skill returns may be missed if only the extensive margin of skill demand is considered. Our data allows us to study how skills measures from job posts and self-reported skills covary at the individual level, and thus, we avoid issues of interpretation that arise when studying correlations between more aggregated variables. While much of the existing literature has focused on the returns to skill demand on the extensive margin, we believe our ability to study both the intensive and extensive margins of task-specific skills at the job level is new to the literature.

[19] Although some of these paper consider routine and non-routine cognitive skills separately. Beaudry et al. (2016) describe a decline in the in the demand for cognitive skills after 2000, but they still find that cognitive skills are associated with higher wages.

References

Adams, A., Balgova, M., & Qian, M. (2020). Flexible Work Arrangements in Low Wage Jobs: Evidence from Job Vacancy Data. *CEPR Discussion Paper Series*, (DP15263).

Alekseeva, L., Azar, J., Giné, M., Samila, S., & Taska, B. (2021). The demand for AI skills in the labor market. *Labour Economics*, 71(May), 102002.

Arntz, M., Gregory, T., & Zierahn, U. (2016). The risk of automation for jobs in OECD countries: A comparative analysis. *OECD Social, Employment and Migration Working Papers*, 189.

Atalay, E., Phongthiengtham, P., Sotelo, S., & Tannenbaum, D. (2020). The Evolution of Work in the United States. *American Economic Journal: Applied Economics*, 12(2), 1–36.

Autor, D. H., Levy, F., & Murnane, R. J. (2003). Skill demand, inequality, and computerization: Connecting the dots. In D. K. Ginther & M. Zavodny (Eds.), *Technology, Growth, and the Labor Market*, number December chapter 6, (pp. 107–129). Netherlands: Kluwer Academic Publishers.

Azar, J., Marinescu, I., Steinbaum, M., & Taska, B. (2020). Concentration in US labor markets: Evidence from online vacancy data. *Labour Economics*, 66(101886).

Bagger, J., Fontaine, F., Galenianos, M., & Trapeznikova, I. (2022). Vacancies, employment outcomes and firm growth: Evidence from Denmark. *Labour Economics*, 75, 102103.

Beaudry, P., Green, D. A., & Sand, B. M. (2016). The Great Reversal in the Demand for Skill and Cognitive Tasks. *Journal of Labor Economics*, 34(1), S199–S247.

Black, S. E. & Spitz-Oener, A. (2010). Explaining Women' S Success : Technological Change and the Skill Content of Women' S Work. *Review of Economics and Statistics*, 92(120), 187–194.

Blair, P. Q. & Deming, D. J. (2020). Structural Increases in Skill Demand After the Great Recession. *AEA Papers and Proceedings*, 110(May), 362–65.

Botero, J. C., Djankov, S., La Porta, R., Lopez-De-Silanes, F., & Shleifer, A. (2004). The regulation of labor. *Quarterly Journal of Economics*, 119(4), 1339–1382.

Braxton, J. C. & Taska, B. (2023). Technological Change and the Consequences of Job Loss. *American Economic Review*, 113(2), 279–316.

Clemens, J., Kahn, L. B., & Meer, J. (2020). Dropouts Need Not Apply? The Minimum Wage and Skill Upgrading. *Journal of Labor Economics*, 39(S1), S107–S149.

Daly, M., Jensen, M. F., & le Maire, D. (2022). University Admission and the Similarity of Fields of Study: Effects on Earnings and Skill Usage. *Labour Economics*, 75, 102118.

Deming, D. & Kahn, L. B. (2018). Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals. *Journal of Labor Economics*, 36(S1), S337–S369.

Deming, D. J. & Noray, K. (2020). Earnings Dynamics, Changing Job Skills, and STEM Careers. *The Quarterly Journal of Economics*, 135(4), 1965–2005.

Fluchtmann, J., Glenny, A. M., Harmon, N., & Maibom, J. (2022). Unemployed Job Search Across People and Over Time: Evidence from Applied-For Jobs. *Working Paper*.

Forsythe, E., Kahn, L. B., Lange, F., & Wiczer, D. (2020). Labor demand in the time of COVID-19: Evidence from vacancy postings and UI claims. *Journal of Public Economics*, 189.

Goos, M., Manning, A., & Salomons, A. (2014). Explaining Job Polarization: Routine-Biased Technological Change and Offshoring. *American Economic Review*, 104(8), 2509–2526.

Grinis, I. (2019). The STEM requirements of "Non-STEM" jobs: Evidence from UK online vacancy postings. *Economics of Education Review*, 70, 144–158.

Groes, F., Kircher, P., & Manovskii, I. (2015). The U-shapes of occupational mobility. *Review of Economic Studies*, 82(2), 659–692.

Heckman, J. & Landersø, R. (2022). Lessons for Americans from Denmark about inequality and social mobility. *Labour Economics*, 77(August), 101999.

Hershbein, B. & Kahn, L. B. (2018). Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings. *American Economic Review*, 108(7), 1737–1772.

Javorcik, B., Stapleton, K., Kett, B., & O'kane, L. (2020). Unravelling Deep Integration: Local Labour Market Effects of the Brexit Vote. *Oxford Discussion Paper*, (November), 5–6.

Jensen, M. F. (2021). Gender Differences in Returns to Skills. *Working Paper*.

Kettemann, A., Mueller, A. I., & Zweimüller, J. (2018). Vacancy Durations and Entry Wages: Evidence from Linked Vacancy-Employer- Employee Data. *IZA Discussion Paper Series*, (11852).

Kircher, P. (2022). Job Search in the 21St Century. *Journal of the European Economic Association*, 20(6), 2317–2352.

Kreiner, C. T. & Svarer, M. (2022). Danish Flexicurity: Rights and Duties. *Journal of Economic Perspectives*, 36(4), 81–102.

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

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

Modestino, A. S., Shoag, D., & Ballance, J. (2020). Upskilling: Do Employers Demand Greater Skill When Workers Are Plentiful? *The Review of Economics and Statistics*, 102(4), 793–805.

Spitz-Oener, A. (2006). Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure. *Journal of Labor Economics*, 24(2), 235–270.

5 Tables

Table 1: Summary statistics of skill categories, estimation sample

	(1) Derived from JP	(2) Derived from LFS
Panel (A)		
Skill Category		
Cognitive	0.61	0.06
Social	0.90	0.05
Management	0.74	0.22
Financial	0.63	0.10
Computer, General	0.59	0.06
Computer, Specific	0.39	0.05
Writing/Language	0.71	0.02
Customer Service	0.87	0.34
Character	0.97	0.05
Panel (B)		
At least 1 skill in a category	0.99	0.72
Conditional on at least 1 skill in a category, fraction with skills falling across:		
Only one skill category	0.03	0.72
Two or fewer different skill categories	0.06	0.95
Three or fewer different skill categories	0.13	0.99
Observations	**2750**	

Notes: In Column 1, we report the fraction of workers in our estimation sample that are in jobs that require one of the 9 skills as captured by the job posting for that job. In Column 2, among the same group of workers, we report the fraction who report that they use a main skill in the 9 categories. See Appendix Table B.1 for more details on the estimation sample.

Table 2: Conditional probabilities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Conditional on LFS (main) skill, probability that job post has the same skill									
	Cognitive	Social	Management	Financial	Computer, General	Computer, Specific	Writing/ Language	Customer Service	Character
Mean	0.849	0.915	0.848	0.853	0.872	0.683	0.808	0.932	0.976
SD	(0.359)	(0.280)	(0.360)	(0.355)	(0.335)	(0.467)	(0.398)	(0.251)	(0.153)
N	166	141	617	278	172	142	52	948	126
Panel B: Conditional on job post (required) skill, probability that LFS has the same skill									
	Cognitive	Social	Management	Financial	Computer, General	Computer, Specific	Writing/ Language	Customer Service	Character
Mean	0.085	0.052	0.258	0.137	0.093	0.090	0.022	0.371	0.046
SD	(0.278)	(0.222)	(0.437)	(0.344)	(0.291)	(0.287)	(0.146)	(0.483)	(0.210)
N	1668	2488	2031	1734	1613	1073	1942	2383	2658

Notes: Panel A presents the probability that a job posting requires a particular skill, conditional on the LFS respondent reporting that task-specific skill. Panel B presents the opposite conditional probability of Panel A; the probability that a LFS respondent is listing that task-specific skill conditional on the job posting requires the skill. The probabilities in Panel B are smaller than Panel A because most workers only report using one main skill. The probabilities in Panel B are, however, larger than the unconditional probabilities in Table 1, Column 2, Panel A, which confirms the positive correlation between individual skill measures in the JP and LFS data. The Estimation Sample, $n = 2,750$, is used to calculate these probabilities – see Appendix Table B.1 for more details on the sample.

Table 3: Regression of LFS (main) skill on job posting skills

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	LFS (main) skill								
Skill present on job posting:	Cognitive	Social	Manage-ment	Financial	Computer, General	Computer, Specific	Writing/ Language	Customer Service	Character
Cognitive	0.045***	0.026***	0.018	-0.004	0.034***	0.023**	0.004	-0.096***	-0.006
	(0.011)	(0.010)	(0.021)	(0.016)	(0.012)	(0.012)	(0.008)	(0.029)	(0.011)
Social	-0.009	0.024	-0.038	-0.052*	-0.025	0.001	-0.007	0.068	0.014
	(0.020)	(0.019)	(0.034)	(0.032)	(0.017)	(0.015)	(0.015)	(0.045)	(0.016)
Management	0.029**	0.015	0.079***	0.008	0.001	0.019*	-0.008	-0.019	0.019
	(0.012)	(0.013)	(0.023)	(0.020)	(0.012)	(0.011)	(0.009)	(0.034)	(0.018)
Financial	-0.001	-0.030***	0.010	0.131***	-0.010	-0.026**	-0.003	-0.059**	0.005
	(0.013)	(0.012)	(0.020)	(0.018)	(0.012)	(0.011)	(0.007)	(0.026)	(0.013)
Computer, General	0.018	-0.003	0.028	-0.055**	0.046***	0.025**	0.013*	-0.067**	0.026**
	(0.011)	(0.012)	(0.026)	(0.023)	(0.013)	(0.011)	(0.008)	(0.034)	(0.013)
Computer, Specific	0.017	-0.009	0.062***	-0.008	0.057***	0.050***	-0.015***	-0.060*	-0.015
	(0.020)	(0.010)	(0.023)	(0.019)	(0.015)	(0.014)	(0.006)	(0.033)	(0.010)
Writing/Language	0.013	-0.003	0.034	-0.016	-0.003	-0.006	0.015**	-0.056**	-0.007
	(0.009)	(0.011)	(0.021)	(0.018)	(0.011)	(0.012)	(0.007)	(0.028)	(0.012)
Customer Service	-0.062***	-0.008	0.027	-0.034*	-0.014	-0.015	-0.023**	0.285***	0.005
	(0.023)	(0.015)	(0.027)	(0.020)	(0.016)	(0.016)	(0.011)	(0.031)	(0.014)
Character	-0.010	-0.042	-0.061	0.073*	-0.009	0.011	0.013	-0.018	-0.016
	(0.039)	(0.043)	(0.054)	(0.038)	(0.030)	(0.019)	(0.021)	(0.074)	(0.024)
Observations	2,750	2,750	2,750	2,750	2,750	2,750	2,750	2,750	2,750
R-squared	0.029	0.007	0.031	0.039	0.039	0.028	0.009	0.072	0.006
Clusters	893	893	893	893	893	893	893	893	893

Notes: Columns 1 to 9 present the results of regressing each of the LFS task-specific main skill measures on job posting skills categories. The Estimation Sample is used to calculate these probabilities – see Appendix Table B.1 for more details on the sample. Note that the skill categories are not mutually exclusive; an individual can have more than one skill. Standard errors, in parentheses, clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4: Effects of skill discrepancy on match duration and wages

	(1)	(2)	(3)	(4)
	Indicator for length of time in job:			Average wage in job
	1(years>=1)	1(years>=2)	1(years>=3)	
	Panel A: Without Individual Controls			
Indicator for skill discrepancy	0.015	0.027	-0.002	0.010
	(0.033)	(0.037)	(0.039)	(0.031)
R-squared	0.061	0.079	0.076	0.205
	Panel B: With Individual Controls			
Indicator for skill discrepancy	0.016	0.027	-0.001	0.000
	(0.032)	(0.036)	(0.037)	(0.017)
R-squared	0.088	0.114	0.114	0.601
Observations	2,733	2,474	2,148	2,750
Year Indicators	YES	YES	YES	YES
Municipality Indicators	YES	YES	YES	YES
Individual Controls	YES	YES	YES	YES
Clusters	887	809	710	893

Notes: Columns 1, 2 and 3 show the estimates of regressing an indicator of whether or not the worker-firm-occupation match last for at least 1, 2 and 3 years respectively on an indicator of skill discrepancy. The skill discrepancy indicator takes the value of 1 if a worker reports mainly using a skill category that the firm did not include in their job post. Panel A presents the estimates of these regressions when just year and municipality fixed effects are included in the regressions. Panel B presents the results if individual controls are included: age, age^2, experience, experience^2, years of education fixed effects, and a gender dummy. Note that for each of the regressions shown in Columns 1-3, we require that the individual started the job at least 1, 2 and 3 years prior to the end of our sample, respectively. This is why the sample sizes are smaller than the Estimation Sample, $n = 2,750$. In Column 4, we regress the indicator of skill discrepancy on average wages on the job. All standard errors are clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 5: Wage regressions

	(1)	(2)	(3)	(4)
Job posting skill indicators:				
Cognitive	0.114***	0.100***	0.041**	0.043***
	(0.026)	(0.023)	(0.017)	(0.017)
Social	-0.006	0.005	0.007	0.008
	(0.031)	(0.028)	(0.025)	(0.025)
Management	0.055*	0.03	0.032*	0.022
	(0.028)	(0.027)	(0.018)	(0.018)
Financial	0.028	0.025	0.011	0.012
	(0.024)	(0.023)	(0.017)	(0.017)
Computer, General	0.022	-0.001	0.006	-0.001
	(0.035)	(0.031)	(0.018)	(0.018)
Computer, Specific	0.055*	0.031	0.023	0.018
	(0.032)	(0.028)	(0.016)	(0.016)
Writing/Language	0.007	-0.001	-0.009	-0.009
	(0.025)	(0.023)	(0.016)	(0.015)
Customer Service	-0.100***	-0.067***	-0.008	-0.01
	(0.028)	(0.025)	(0.020)	(0.020)
Character	-0.128**	-0.111**	-0.101***	-0.093***
	(0.051)	(0.045)	(0.037)	(0.034)
LFS (main) skill indicators:				
Cognitive		0.071**		0.026
		(0.030)		(0.022)
Social		0.067**		0.038
		(0.030)		(0.023)
Management		0.213***		0.095***
		(0.019)		(0.014)
Financial		0.005		0.001
		(0.039)		(0.023)
Computer, General		0.098***		0.027
		(0.035)		(0.022)
Computer, Specific		0.111***		0.041*
		(0.033)		(0.021)
Writing/Language		-0.104*		-0.077**
		(0.055)		(0.032)
Customer Service		-0.123***		-0.024*
		(0.020)		(0.013)
Character		0.134***		0.091***
Observations	2,750	2,750	2,750	2,750
Clusters	893	893	893	893
R-squared	0.252	0.347	0.608	0.623
F-statistic	–	21.99***	–	8.57***
Individual Controls	NO	NO	YES	YES

Notes: Columns 1 and 3 presents the results from estimating Equation 2 when restricting θ_k to zero, and Columns 2 and 4 presents the results from estimating Equation 2 when allowing θ_k to vary. The F-statistic shown is the result of a joint hypothesis test with that null that all of the θ_k are 0. The Estimation Sample is used to calculate these probabilities – see Appendix Table B.I for more details on the sample. Individual controls include: age, age^2, experience, experience^2, years of education fixed effects, and a gender dummy. Note that the skill categories are not mutually exclusive; an individual can have more than one skill. Standard errors, in parentheses, clustered at the firm level. All regressions include year and municipality fixed effects. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Appendix A: Data

A.1 Job postings data

The JP are supplied by HBS Economics (HBS). The data are provided after an inital cleaning procedure has been performed. HBS asserts that their data contain the near universe of publicly accessible Danish online job posts from 2007 to 2017. Duplicates are removed and the data cleaned before machine reading the job posts. HBS extracts the date on which a given job post was posted online, a firm ID, and an occupation code. 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 extracts keywords from the raw text in the job post. In many ways, the resulting data are similar to the US job postings data supplied by Burning Glass Technologies.

We group individual keywords from the job posts into 9 different skills categories, using the categories in (2018). We do this by manually assigning the most frequently occurring keywords (around 2,000 terms) to a skill category or noise tag. The remaining words are categorized using synonyms or machine-learning methods based on each word's dictionary definition (see Jensen, 2021). The most frequent keywords for each skill category are reported in Table A.1

A.2 Danish Labour Force Survey

Similar to the job postings data, we also group individual keywords from the self-reported skills usage into 9 different skills categories. Free-text answers to the task/skill usage question are cleaned by removing stop words (e.g., "and," "or") and are spell checked. Next, the same mapping of keywords to skill groups used with the JP are used for the remaining LFS words, categorizing the majority of them. We then manually categorize approximately 800 of the most frequent additional keywords from the LFS text data, such that slightly more than 75% of all keyword observations from the LFS are categorized. As mentioned above, we interpret a reported task as the utilization of a task-specific skill, and therefore prefer to refer to the measures extracted from the LFS as the main skills used on the job. The most frequent keywords for each skill category are reported in Table A.2.

A.3 Data match

To understand the representativeness of our samples, we compare the population of professionals, technicians, and associate professionals, clerical workers, and service and sales workers who are in the first year of a new private-sector job (either in a new occupation or in a new firm) as captured by Danish register data, shown in Column 1 of Appendix Table B.1, to the JP-Population Matched sample, shown in Column 2. Compared to the population as a whole, individuals in the

JP-Population Matched sample are slightly younger and less experienced. This stems from the fact that entry-level jobs are more often posted on an online platform relative to more senior jobs, which are often filled either internally without advertisement or via established networks, explaining the relatively lower levels of experience shown in Column 2.

Next, we look at the subset of individuals from the register data who have answered the LFS in order to compare their self-reported main skills to the skills advertised in the JP. Column 3 in Appendix Table B.1 presents the results from merging the LFS with register data that capture first-year individual-level employment spells in the private sector so that firm identifiers can be appended. In the LFS sample, workers are less likely to be students and more likely to be older and live outside of Copenhagen. In our regression specifications, we include controls for age, education, experience, and municipality in order to account for this finding. Column 4 presents the subsample for whom information is available on both the job posting skills and self-reported main skills (i.e., the result of merging the JP and LFS samples), and as expected, the resulting sample looks quite similar to the JP-Population Matched sample shown in Column 2. The take-away from this exercise is that the matched JP-LFS (the estimation sample) is representative of the JP-Population Matched sample.

A.4 Keywords

Below we show the keywords of each skill category for the LFS and JP skills. As a robustness check, we have performed the wage regressions where the skill categories from the JP skills only include key words that are also in the top 50 percent of the LFS skills. The coefficients in wage regression do not significantly change and we therefore conclude that differences across skill measures in the individual keywords included in the skill are not driving the results.

Table A.1: Keywords accounting for top 50% of character keyword observations, JP

Keywords - English	Keywords - Danish	Skill	Frequency
Committed	Engageret	Character	309701
Responsibility	Ansvar	Character	299171
Self Employed	Selvstændig	Character	298694
Professional	Faglig	Character	191637
Friendly	Venlig	Character	186248
Active	Aktiv	Character	184816
Flexible	Fleksibel	Character	181899
Honest	Ærlig	Character	132007
Nature	Natur	Character	128892
Dynamic	Dynamisk	Character	121374
Positive	Positiv	Character	118939
Open	Åben	Character	108771
Personal	Personlig	Character	99536
Joy	Glæde	Character	98588
Professional	Professionel	Character	97590
Structured	Struktureret	Character	96414
Good Mood	Godt Humør	Character	91386
Humor	Humor	Character	80531
Targeted	Målrettet	Character	79881
Burner	Brænder	Character	79728
Drive	Drive	Character	78783
Informal	Uformel	Character	75214
Overview	Overblik	Character	71501
Busy	Travl	Character	68029
Order	Orden	Character	67111
Stable	Stabil	Character	66880
Values	Værdier	Character	65210
Respect	Respekt	Character	62315
Solution	Løsning	Cognitive	117829
Logical	Logisk	Cognitive	75394
Research	Forskning	Cognitive	59794
Optimization	Optimering	Cognitive	48250
Analysis	Analyse	Cognitive	29163
Issues	Problemstillinger	Cognitive	24744
Technical	Teknisk	Computer, General	132799
SUPPRESS*	SUPPRESS*	Computer, General	105486
System	System	Computer, General	58856
Data	Data	Computer, General	40914
IDENT*	Ident	Computer, Specific	65442
Program	Program	Computer, Specific	28246
Padding	Padding	Computer, Specific	27177

Keywords - English	Keywords - Danish	Skill	Frequency
Platform	Platform	Computer, Specific	21438
E Security	E Security	Computer, Specific	7048
Server	Server	Computer, Specific	7031
Hardware	Hardware	Computer, Specific	6279
Databases	Databaser	Computer, Specific	6043
Service	Service	Customer Service	237279
Sell	Sælge	Customer Service	199260
Customer	Kunde	Customer Service	198198
Orders	Ordrer	Customer Service	73799
Guide	Vejlede	Customer Service	37558
Serves	Betjener	Customer Service	34490
Economy	Økonomi	Financial	60573
Budget	Budget	Financial	42324
Financial Accounting	Regnskab	Financial	40103
Purchase	Indkøb	Financial	36767
Resources	Ressourcer	Financial	31373
Turnover	Omsætning	Financial	31139
Margin	Margin	Financial	24407
Accounting	Bogføring	Financial	20284
Import	Import	Financial	19540
Bookkeeping	Bogholderi	Financial	18373
Reconciliation	Afstemning	Financial	17999
Balance	Balance	Financial	17929
Management	Ledelse	Management	105495
Plan	Planlægge	Management	96861
Operation	Drift	Management	92595
Implement	Implementere	Management	66631
Administration	Administration	Management	64596
Coordinate	Koordinere	Management	63942
Supervision	Supervision	Management	45611
Control	Styring	Management	33495
Management	Forvaltning	Management	32727
Organize	Organisere	Management	29302
Cooperation	Samarbejde	Social	440544
Team	Team	Social	363728
Communication	Kommunikation	Social	149524
Extroverted	Udadvendt	Social	126718
Social	Social	Social	98160
Dialog	Dialog	Social	70567
Danish	Dansk	Writing/Language	160608
English	Engelsk	Writing/Language	109639

Keywords - English	Keywords - Danish	Skill	Frequency
Write	Skrive	Writing/Language	105023
Language	Sprog	Writing/Language	68069

Notes: * IDENT refers to words that have been anonymized as they could otherwise potentially identify a firm. SUPPRESS refers to the group of words of sufficiently low frequency. Keeping the actual low frequency words would make it possible to identify individual observations from the raw data.

Table A.2: Keywords accounting for top 50% of character keyword observations, LFS

Keywords - English	Keywords - Danish	Skill	Frequency
Responsibility	Ansvar	Character	5003
Personal	Personlig	Character	3471
Research	Forskning	Cognitive	3578
Analysis	Analyse	Cognitive	2330
Mathematics	Matematik	Cognitive	1690
Technical	Teknisk	Computer, General	3265
Software	Software	Computer, General	2023
System	System	Computer, General	1919
Program	Program	Computer, Specific	1608
Server	Server	Computer, Specific	1045
Graphic	Grafisk	Computer, Specific	920
Edb	Edb	Computer, Specific	845
System Developer	Systemudvikler	Computer, Specific	555
Sell	Sælge	Customer Service	21094
Customer	Kunde	Customer Service	18319
Serves	Betjener	Customer Service	6846
Service	Service	Customer Service	4838
Expediting	Ekspederer	Customer Service	2613
Financial Accounting	Regnskab	Financial	7995
Bookkeeping	Bogholderi	Financial	5833
Purchase	Indkøb	Financial	3996
Economy	Økonomi	Financial	3293
Accounting	Bogføring	Financial	2811
Management	Ledelse	Management	23071
Administration	Administration	Management	6560
Manager	Manager	Management	3816
Plan	Planlægge	Management	3340
Manager	Direktør	Management	3048
Operation	Drift	Management	2389
Department Manager	Afdelingsleder	Management	2376
Social	Social	Social	12957
People	Folk	Social	3459
Dansk	Dansk	Writing/Language	2993
Write	Skrive	Writing/Language	2728
English	Engelsk	Writing/Language	1337

Appendix B: Additional results

Table B.1: Comparison of matched samples

	(1)		(2)		(3)		(4)	
	Population		JP-Population Matched		LFS		JP-LFS Matched *Estimation Sample*	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Female	0.519		0.523		0.497		0.494	
Age	35.790	(11.621)	34.420	(11.288)	36.390	(11.501)	35.325	(10.890)
Immigrant	0.078		0.083		0.061		0.063	
Married	0.361		0.327		0.406		0.390	
Number of children under 18	0.677	(0.975)	0.639	(0.954)	0.740	(0.997)	0.729	(0.975)
Registered experience	11.731	(10.848)	10.704	(10.405)	12.060	(10.905)	11.132	(10.237)
Potential experience (years since ended education)	12.470	(15.867)	11.344	(15.091)	11.974	(14.247)	10.699	(10.584)
Years of education (monthly)	14.314	(2.270)	14.180	(2.290)	14.738	(2.334)	14.766	(2.277)
Student at any point in month	0.168		0.171		0.101		0.096	
Home region indicators:								
Northern Jutland	0.083		0.083		0.092		0.089	
Mid-Jutland	0.215		0.222		0.232		0.241	
Southern Denmark	0.174		0.170		0.184		0.167	
Capital Region	0.405		0.401		0.373		0.387	
Zealand	0.124		0.124		0.119		0.117	
Work region indicators:								
Northern Jutland	0.078		0.077		0.084		0.078	
Mid-Jutland	0.208		0.216		0.228		0.233	
Southern Denmark	0.167		0.162		0.179		0.169	
Capital Region	0.454		0.459		0.425		0.446	
Zealand	0.091		0.086		0.084		0.075	
Person-year observations	2,821,996		499,645		13,138		2,750	
Share of new jobs matched to job post			*17.71%*				*20.93%*	

Notes: "Population" refers to the full population of new individual-level job spells in the private sector starting in January 2008 to July 2017 in ISCO-groups 2 to 5. "JP-Population Matched" refers to the subsample of new job spells that can be linked to corresponding job post(s). "LFS" refers to the subsample of new job spells that can be linked to an observation in the LFS within the first year of commencing the job spell. "JP-LFS Matched / Estimation Sample" refers to the subsample of new job spells that can be linked to both corresponding job post(s) and an observation in the LFS within the first year of commencing the job spell.

Table B.2: Summary statistics of keywords and skill categories by sample

	(1)	(2)
	JP	**LFS**
Cognitive	0.563	0.051
Social	0.902	0.053
Management	0.697	0.230
Financial	0.574	0.119
Computer, General	0.554	0.062
Computer, Specific	0.349	0.041
Writing/Language	0.674	0.025
Customer Service	0.876	0.313
Character	0.968	0.040
Fraction with at least 1 skill keyword	0.990	0.691
Fraction with 1 skill keyword conditional on having at least 1 skill keyword	0.009	0.467
Observations	**499,645**	**13,138**

Notes: In Column 1, we report the fraction of jobs categorized as requiring one of the 9 skills in their corresponding job post(s). In Column 2, we report the fraction of workers who report that they use a main skill in the 9 categories. The JP sample is used to calculate the fractions in Column 1 – see Appendix Table B.1, Column 2, for more details on the sample. The LFS sample is used to calculate the fractions in Column 2 – see Appendix Table B.1, Column 3 for more details on the sample.

Table B.3: Simulated regression coefficients

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Simulated maximum coefficient									
	Cognitive	Social	Manage-ment	Financial	Computer, General	Computer, Specific	Writing/ Language	Customer Service	Character
Simulated Maximum	0.098	0.056	0.30	0.158	0.102	0.127	0.027	0.393	0.051
Panel B: Unconditional coefficients									
	Cognitive	Social	Manage-ment	Financial	Computer, General	Computer, Specific	Writing/ Language	Customer Service	Character
Unconditional coefficient	0.062	0.0063	0.124	0.094	0.073	0.064	0.010	0.196	0.014

Notes: Panel A presents the unconditional regression coefficient from LFS skills regressed on JP skills in simulated data where individuals with LFS skills is a strict subset of individuals with JP skill and where the probability of having a given skill is from Table 1. Panel B represents the unconditional regression coefficients similar to Table 3 to be compared with the unconditional coefficients in Panel A.