sociological science

Are Occupations "Bundles of Skills"? Identifying Latent Skill Profiles in the Labor Market Using Topic Modeling

Marie Labussière,[a] Thijs Bol[b]

a) Sciences Po; b) University of Amsterdam

Abstract: Skills are considered a key determinant of workers' labor market opportunities, especially in times of rapid technological change. However, existing research rarely conceptualizes and measures skills in their own right, instead relying on occupations as a proxy. How does this limit our understanding of the labor market structure and of wage inequality? In this article, we leverage a unique dataset of millions of online job postings in the United Kingdom to measure the skill profiles of jobs and analyze their similarity within and between occupational categories. Our data-driven approach reveals substantial discrepancies between occupational classifications and the actual skill content of jobs. We further demonstrate that job-level variation in skill content constitutes an independent source of wage inequality—one that is obscured by analyses at the occupational level. These findings challenge the conventional view of occupations as coherent bundles of skills, offering new avenues for analyzing labor market stratification.

Keywords: occupations; skills; tasks; social stratification; topic modeling; online job ads

Reproducibility Package: All code necessary to reproduce the results reported in this article is publicly available in a replication package hosted on GitHub (https://github.com/mlabussiere/Occupations-bundles-of-skills.git). The online supplement also contains additional information on the data, methods, and robustness checks. The data are subject to access restrictions and cannot be shared publicly.

Citation: Labussière, Marie, Thijs Bol. 2026. "Are Occupations "Bundles of Skills"? Identifying Latent Skill Profiles in the Labor Market Using Topic Modeling" Sociological Science 13: 362-407.

Received: December 9, 2025

Accepted: March 2, 2026

Published: April 13, 2026

Editor(s): Ari Adut, Vincent Buskens

DOI: 10.15195/v13.a16

IN the context of rapid technological change, the notion of skill has gained renewed prominence in public and policy debates. A widely accepted view holds that workers' vulnerability to job displacement and labor market precarity is shaped primarily by the nature of their skills and the tasks they are able to carry out on the job (Autor, Levy, and Murnane 2003; Acemoglu and Autor 2011). For example, workers who cannot meet the growing demand for analytical, complex, and non-routine tasks are expected to face diminishing career opportunities and wage penalties (e.g., Autor and Handel 2013; Hanushek et al. 2017; Haslberger 2021). This understanding has motivated the expansion of skill-oriented policies aimed at better equipping present and future workers with the skills deemed necessary in an evolving labor market (OECD 2019, 2024).

Paradoxically, despite their alleged growing importance, the ways in which sociologists and economists conceptualize and measure skills have remained relatively crude (Liu and Grusky 2013). Conceptually, skills or tasks[1] are often reduced to a small set of independent dimensions (e.g., Deming 2017; Buchmann, Buchs, and Gnehm 2020; Haslberger 2021), overlooking the possibility that technological change affects not only the prevalence of particular skills but also the ways workers must combine them. Such a narrow view of workers' skills stands in tension with a growing body of research showing that employers increasingly require workers to have broad, diverse, and complementary skill sets (Anderson 2017; Stephany and Teutloff 2024; Liu et al. 2025). Yet, it is still unclear to what extent more comprehensive conceptualizations of skills would enhance our understanding of labor market inequality.

The conceptual crudeness of "skill" is mirrored methodologically. In most studies, skills are measured using occupational classifications, which are designed to group jobs that perform similar types of work (see, e.g., International Labour Office 2012). Economists typically operationalize skill or task dimensions at the occupational level based on expert coding, such as the O*NET expert-based database (e.g., Goos, Manning, and Salomons 2014; Deming 2017; Autor and Thompson 2025). Although less prevalent, stratification sociologists tend to use a similar strategy (e.g., Liu and Grusky 2013; Lin and Hung 2022; Mouw, Kalleberg, and Schultz 2024). Other studies, implicitly or explicitly, rely on occupations as an empirical proxy for skills rather than measuring them directly. For example, stratification scholars explain part of the wage inequality across occupations by differences in occupational skill requirements (e.g., Mouw and Kalleberg 2010; Williams 2013; Janietz and Bol 2020). Similarly, research on career mobility emphasizes that the ability of workers to change occupations over the course of their careers depends heavily on the (lack of) transferability of their skills across occupations (e.g., Sacchi, Kriesi, and Buchmann 2016; Kalleberg and Mouw 2018). Overall, when it comes to operationalization, skills are thus predominantly *measured through* or *inferred from* occupational categories, rather than observed directly.

We argue that while skills have become increasingly important in public and policy debates, in our scholarship, we may lack the conceptual and methodological tools to capture their full effects. Currently, the influence of skills is conflated with that of occupations, reflecting the implicit assumption that occupations can be treated as bundles of skills—that is, that workers within the same occupation share a characteristic set of skills that explains their labor market outcomes. In this article, we develop a direct and comprehensive measure of skills to examine this assumption, pursuing two main goals. First, we analyze whether a more detailed and realistic operationalization of skills adds descriptive value beyond what occupations capture, by mapping job skills across the occupational spectrum. Second, we assess whether this operationalization has explanatory value: to what extent do the skills required on the job predict wages independently of occupation?

We are not the first to challenge the idea that occupations can be treated as bundles of skills. First, there has long been evidence that workers transfer a substantial part of their skills portfolios when changing occupations (Poletaev and Robinson 2008; Gathmann and Schönberg 2010; Cheng and Park 2020), suggesting overlaps in the skills required across *different* occupations. Second, the content of jobs has been shown to vary widely *within* occupations (Yamaguchi 2012), even when using detailed occupational categories (Autor and Handel 2013; Cassidy 2017; Freeman, Ganguli, and Handel 2020; Martin-Caughey 2021).

What remains unclear is the extent to which these discrepancies between occupations and skills matter for our understanding of labor market inequality. Does skill heterogeneity within occupations hide unexplored forms of stratification, or does it mainly capture irrelevant, idiosyncratic differences between jobs? Although not a new question (e.g., Jackson 2007; Autor and Handel 2013), recent research using job-level data and computational methods shed new light on within-occupation heterogeneity (Marinescu and Wolthoff 2020; Martin-Caughey 2021). In particular, Martin-Caughey (2021) shows that heterogeneity in job content is associated with wage inequality patterns that are obscured at the occupation level. This raises the question of whether additional patterns emerge when the skill content of a job is explicitly modeled.

If the skill content of jobs shapes workers' labor market outcomes beyond what is captured by occupational categories, this would have implications not only for how we measure skills but also for how we conceptualize the role of occupations. Indeed, occupations are a central unit for understanding social inequality more broadly (Blau and Duncan 1967; Parkin 1971; Weber 1979). They have been shown to underlie various forms of inequality, such as social closure (Weeden 2002), unequal positions in status hierarchies (Treiman 1977; Ganzeboom, De Graaf, and Treiman 1992), class divisions (Goldthorpe 2002), or differences in work precarity (Kalleberg 2009). For many sociologists, occupations are therefore much more than mere bundles of skills—even if it remains difficult to specify precisely what this "more" entails (Williams 2013; Haupt and Ebner 2020; Leicht 2020). By measuring skills and occupations independently, this article makes it possible to assess the extent to which the explanatory power of occupations can be attributed to their underlying skill composition or other mechanisms posited by sociologists. Until recently, a direct and granular operationalization was difficult to achieve because of scarce detailed job-level skill data and the lack of methods capable of capturing the multidimensionality of skills in a parsimonious way.

In particular, previous studies measuring the task or skill content of jobs have faced at least three challenges, which we address in this article. First, existing data provide limited leverage to analyze variation in skill content within and between occupations. Expert-informed sources, such as the O*NET, assumes within-occupation homogeneity of skills, while employee surveys typically provide only very generic skill information. To overcome these limitations, we rely on real job postings, where employers define the skill requirements and wages for the jobs they are hiring for. Besides providing rich job-level information, analyzing the relationship between skills and wages from the employer's perspective is conceptually meaningful: as Jackson (2007:370) and Bills, Di Stasio, and Gërxhani (2017:292) argue, the demand side may be a more fundamental determinant of inequality than the supply side, as employers ultimately decide which workers' skills are rewarded in the labor market.

Second, rather than using preconceived categories of skills that may not reflect the rapidly evolving skill structure of the labor market, our approach, based on natural language processing, allows us to derive data-driven, job-level measures of skills while maintaining interpretability. Third, previous research on skills has tended to focus on narrow sets of skills and to analyze them in isolation, in contrast to the idea that skills are "coherently bundled into jobs" (Fernández-Macías and Bisello 2020:6). We develop the notion of skill profiles to reflect the combination of skills required on the job, and draw on granular data from millions of job postings to identify them.

More specifically, we draw on a unique dataset of online job postings collected in 2019 by Lightcast, a labor market analytics company (formerly known as Burning Glass Technologies). The data cover the near universe of job ads published online that year and contain highly detailed information about the set of skill requirements listed by employers. The context for our study is the United Kingdom, a country where occupational wage inequality is large (Bol and Weeden 2015) and has increased in recent decades (Williams 2013). The United Kingdom has a relatively deregulated labor market (Hall and Soskice 2001b) and a generalist educational system, which is only loosely linked to the labor market. As a result, many jobs do not require specific educational degrees or qualifications (Bol and Weeden 2015), and employers often need to list skills explicitly rather than relying on assumed qualifications (e.g., boiler repair for plumbers).[2] The United Kingdom thus provides an especially suitable context for examining how occupational wage differences relate to skill similarity using job postings data.

Using 6.9 million job postings and a topic modeling approach, we generate skill profiles for a representative sample of 600,000 jobs in the United Kingdom. We find that jobs require much more diverse skill profiles than is usually assumed, highlighting the need to account for both the multidimensionality of skills and their interdependence at the job level. Moreover, we find that most of the skill similarity across jobs occurs between rather than within occupational boundaries, challenging the common conceptual and empirical conflation between occupations and skills. We also show that skill content measured at the job-level matters: our comprehensive measure of skill profiles is an important predictor of the wage offered in job postings, complementing rather than supplanting the effect of occupations. These findings open new avenues for understanding the nature of occupations and the distinct role that skills play in shaping labor market inequality, which we discuss in the final section.

Background

What is an occupation?

Occupations have long been central to sociological research on inequality. Since the discipline's inception (Durkheim 1893; Weber 1979), scholars have studied the types of work that people perform and their implications for social stratification. This interest coincided with those of governments, which began systematically measuring and categorizing jobs at the end of the nineteenth century (Leicht 2020).

Occupations are typically defined as sets of jobs requiring similar skills (Hauser and Warren 1997:180), where skills refer to the ability to carry out the tasks and duties of a given job. The most widely used occupational classification, the International Standard Classification of Occupations (ISCO), is based on both the type of work performed by workers and the skill level of the job, which reflects the complexity and range of the tasks involved (International Labour Office 2012). National occupational classifications often follow this line of reasoning (see, e.g., Office for National Statistics 2010:2).

Occupations are thus explicitly defined as bundles of skills used on the job. This approach dominates in economics, where occupations serve as proxies for job task content (e.g., Autor et al. 2003) or for task-specific human capital (e.g., Gathmann and Schönberg 2010). In sociology, by contrast, occupations are often conceptualized more broadly—as status groupings, social classes, or institutional units—depending on the theoretical tradition (Haupt and Ebner 2020). Across these traditions, occupations play a central role in structuring wage inequality (Mouw and Kalleberg 2010; Williams 2013; Janietz and Bol 2020) and intragenerational mobility (Lin and Hung 2022).

However, the precise mechanisms linking occupations and inequality are often left implicit, with occupations functioning as "a placeholder for complex social processes" (Haupt and Ebner 2020:21). This ambiguity reflects a broader lack of conceptual clarity about what an occupation actually entails. With a few rare exceptions (Sakamoto and Wang 2020; Kim, Kim, and Ban 2020; Martin-Caughey 2021), scholars have taken the definition and measures of occupational categories for granted, without trying to "pin down what exactly occupations proxy" (Williams 2013:853).

This ambiguity is often resolved implicitly by treating occupations as expressions of a natural or objective ordering of skills. However, as Braverman (1998) notes, "classifications of workers (...) are neither 'natural' nor self-evident" (p. 296). Historically, assessments of job skill levels were derived from normative conceptions rather than empirical investigation (Braverman 1998:297). As early as the 1950s, Caplow (1954) showed that occupational rankings often reflect cultural norms, such as the superiority of white-collar over manual work, rather than actual skill requirements (p. 45). This implicit hierarchy between occupations remains influential today, with professional occupations generally seen as requiring the most skill and elementary occupations the least.

Given these conceptual ambiguities and the increasing availability of rich job-level skill data, it is timely to open the black box of occupations and measure the skill content of jobs directly (Martin-Caughey 2021). Doing so allows researchers to assess whether occupations primarily reflect skill bundles or broader institutional and status mechanisms. However, this requires conceptualizing and measuring skills separately from occupations, capturing their multidimensional and interdependent nature, to explore the extent to which they may play a distinct role in generating inequality.

From skills to skill profiles

Despite growing attention to changes in the skills and task composition of jobs, the nature of skills themselves remains largely opaque. There is little discussion—and even less consensus—about which skills matter most for workers' labor market outcomes. Since the seminal work of Autor et al. (2003), research has overwhelmingly relied on crude dichotomies, such as routine/non-routine or abstract/manual tasks (for recent examples in sociology, see Rohrbach-Schmidt 2019; Schulz, Solga, and Pollak 2023). A smaller body of literature has employed more detailed skill classifications (e.g., Liu and Grusky 2013; Cheng and Park 2020) or examined additional job characteristics, such as task complexity (Caines, Hoffmann, and Kambourov 2017) or social skills (Deming 2017). Yet these approaches share an underlying assumption: that the skill composition of jobs can be captured by one or a few independent dimensions, abstracted from the way skills are combined and embedded within jobs.

Although Liu and Grusky (2013) already noted this, recent studies have persisted in treating specific skill types as independent factors, ignoring that their meaning and relevance depend on the other skills required for the job. For example, problem-solving denotes different competencies for a welder than for a manager. Jobs typically combine tasks of different natures (e.g., cognitive, interpersonal, physical, creative) that can only be performed competently if workers develop and use the right mix of skills. This creates interactions and interdependencies among skills that factor into workers' productivity and wages. As Anderson (2017) shows, wage differentials are better explained by skill combinations than by isolated skills, as some skills are synergistic and enhance the effectiveness of others (Anderson 2017). For example, programming languages or generalist legal skills can be combined with a wide range of high-value skills, giving workers who master them a comparative wage advantage (Stephany and Teutloff 2024).

Building on this insight, we introduce the concept of *skill profiles*, defined as the combination of skills required to perform a given job. This concept captures both the interdependence and complementarity of skills, as well as their contextual meaning, recognizing that the same skill may be used and valued differently depending on its configuration with others (e.g., Hurrell, Scholarios, and Thompson 2013). A skill-profile perspective requires considering the full spectrum of skills that workers develop and use, from the very general or transversal skills that are relevant in many work contexts (e.g., communication and organizational skills) to the highly specialized knowledge and technical know-how.

A skill-profile approach reveals important dynamics in the relationship between skills and inequality that remain invisible when skills are analyzed in isolation. Changes in labor market demand do not simply raise or lower the value of individual skills; they also reshape the complementarities among workers' skill sets (Stephany and Teutloff 2024). For example, Deming (2017) finds that the puzzling stagnation in the demand for high-skilled labor since 2000 masks a growing complementarity between cognitive and social skills: the return to math skills has increased mainly for workers with high levels of interpersonal skills. A skill-profile approach is increasingly relevant given evidence that new forms of work organization require workers to develop skills that were traditionally associated with different occupations (Hénaut, Lena, and Accominotti 2023; Liu et al. 2025). Consistently, workers who combine skills from different domains are found to enjoy the strongest labor market advantage (Anderson 2017; Stephany and Teutloff 2024).

Occupations, skills, and inequality in the labor market

The value of introducing a new skill measure lies not in its complexity per se, but in whether it captures patterns that occupation-level measures may miss. To assess the contribution of our skill-profile approach, we return to the long-standing question of wage inequality and evaluate whether job-level skill profiles provide explanatory leverage beyond standard occupational categories.

Occupations are a well-known determinant of wage disparities between workers. Across highly different countries, studies have shown that a large share of the total wage inequality is found between occupations (Weeden et al. 2007; Mouw and Kalleberg 2010; Williams 2013; Bol and Weeden 2015). These studies estimate that roughly 40–50 percent of all variance in wages is between occupations, and that this share has increased over time (see Williams (2013) for the United Kingdom, Mouw and Kalleberg (2010) and Weeden et al. (2007) for the United States). While the empirical relationship is clear, the mechanisms underlying this association are less well-specified, mirroring the conceptual ambiguity noted above.

For some scholars, occupations are primarily relevant to wages because they capture skill differences between workers—especially skill differentials not reflected in standard measures of education (e.g., Juhn, Murphy, and Pierce 1993; Böhm, Von Gaudecker, and Schran 2024). For others, occupations are much more than bundles of skills (e.g., Weeden 2002; Kalleberg, Mouw, and Schultz 2025). They can be defined as institutional units that erect and defend barriers to entry into labor market positions (Bol and Weeden 2015; Damelang, Haupt, and Abraham 2018; Haupt 2023). These institutional barriers, such as occupational licenses or educational credentials, generate economic rents for occupational incumbents, producing wages in excess of what would prevail in a fully competitive market (Weeden 2002). From these perspectives, occupations affect wages not just because they differ in skill content, but also because they vary in the strength of the institutions that regulate them.

With a direct measure of the skill profiles of jobs, we can make a first step in assessing the relevance of these alternative views of occupations. If occupations are indeed primarily bundles of skills, our job-level measure of skill content should be largely redundant: it should not improve the explanation of wage variance beyond what occupations already capture. This expectation aligns with Weeden and Grusky (2005), who argue that job-level work conditions add little explanatory power once occupations are taken into account (pp. 191–192). More generally, such a finding would reinforce the emphasis on between-occupation wage inequality in the literature (Mouw and Kalleberg 2010; Williams 2013; Liu and Grusky 2013), as well as the common practice of aggregating task or skill information to the occupational level (e.g., Cheng and Park 2020; Haslberger 2021; Lin and Hung 2022). Conversely, if the effect of occupations on wages is not reducible to skills, this would lend support to research traditions that conceptualize occupations more broadly.

With respect to skills, the extent to which heterogeneity in skill profiles within occupations drives income inequality beyond occupations remains largely unknown, due to the lack of prior research using detailed job-level information. Existing studies rely on coarse measures of job task content and find only a modest increase in the overall explanatory power of their models when job-level task characteristics are included (Autor and Handel 2013; Cassidy 2017; Rohrbach-Schmidt 2019).[3] Using a more comprehensive and fine-grained measure of skills required at work, we can better assess whether prior studies have underestimated the explanatory value of job-level characteristics due to measurement limitations.

Data

Lightcast job posting data

We use a unique source of online job ads collected in the United Kingdom by Lightcast, a job market analytics company formerly known as Burning Glass Technologies.[4] Lightcast spiders more than 51,000 job-related websites daily to identify online job postings. After deduplicating the identified job postings, it extracts, parses, and codes their data into a machine-readable format to create labor market analytics (see Carnevale, Jayasundera, and Repnikov 2014:3–5 for details). The Lightcast data offer important advantages for our research. First, because of its broad coverage, the dataset is claimed to encompass the near universe of jobs posted online, avoiding reliance on a single job platform (Cammeraat and Squicciarini 2021:11). Second, Lightcast data extracts or derives fine-grained information from job postings, including the job title, salary offer, and the occupational code according to the UK Standard Occupational Classification (SOC). Most important for our study, Lightcast extracts the skill requirements from the open text of each job posting, using a dynamic taxonomy of over 32,000 distinct skill keywords.[5] Skill requirements are standardized[6] but unique, offering granular information about the skill profile of jobs. This data is ideal for exploring the underlying skill structure in the labor market without depending on pre-existing skill classifications. Furthermore, unlike DOT or O*NET, Lightcast data allows us to analyze skills at the job level and thus compare the skill profiles of jobs *within* occupational groups.

Lightcast data has been increasingly used in labor market research, albeit mostly by economists (e.g., Deming and Kahn 2018; Hershbein and Kahn 2018; Modestino, Shoag, and Ballance 2020). The use of job postings in general has been limited in sociology (for exceptions, see Salvisberg and Sacchi 2014; Buchmann et al. 2020), with most studies opting for expert-coded or worker self-report data (e.g., Liu and Grusky 2013; Fernández-Macías and Hurley 2016; Martin-Caughey 2021). Job postings offer a different facet of the labor market, as part of the "official front stage of the recruitment process" (Hamann and Beljean 2021:50). Employers write them to attract the right candidates for the position, mobilizing explicit categories that *they* perceive as most relevant. Such categories are not only based on the specific requirements of the job but are also shaped by broader organizational factors, norms, and institutions (Sostero and Fernández-Macías 2021:9–12). This can lead to a number of biases, such as when employers use job postings to market their company brand (Backhaus 2004). One potential concern for our study is that the skills listed in job postings may not accurately reflect the skill composition of occupations. However, analyses comparing the skill requirements in online job posting data with the tasks documented by workers in surveys find that job postings do offer a relatively accurate picture of what people report doing on the job (Sostero and Fernández-Macías 2021; Daly, Groes, and Jensen 2025).

The advantages of Lightcast data come with a few limitations. Lightcast data only includes online job postings, which can provide an inaccurate representation of labor demand. According to Carnevale et al. (2014:11), between 60 and 70 percent of job openings were posted online in 2013. Although this share has likely increased over time, recent research shows that Lightcast data is still skewed toward certain occupations (Cammeraat and Squicciarini 2021), since in some industries (e.g., construction), the hiring process is primarily offline. Elementary and skilled trades occupations are under-represented in online job postings, to the advantage of professional occupations (Cammeraat and Squicciarini 2021). This lack of representativeness is acknowledged in previous literature (e.g., Hershbein and Kahn 2018; Modestino et al. 2020), and few studies address this issue in practice (for exceptions, see Blair and Deming 2020; Cammeraat and Squicciarini 2021). In our study, we tackle this issue by using a stratified random sample, as we describe in more detail below.

Another source of bias stems from the way Lightcast extracts and codes the information contained in job postings. Some characteristics may not be correctly identified or classified by Lightcast's parsers and custom machine-learning algorithms (see Carnevale et al. 2014:16). Of particular concern for our study is the occupational coding of job postings, which could be incorrectly imputed. Quality tests showed that the accuracy level for the most granular occupation classification (six-digit) was relatively low (73 percent) in the US Lightcast data in 2013 (Carnevale et al. 2014:16). However, as Lightcast's algorithms are continuously trained on new data, their accuracy has likely improved over the past decade. Furthermore, occupational misclassifications are not specific to Lightcast data: the mismatch rates between two independent coders can be as high as 50 percent for the four-digit occupational level in traditional surveys (Kim et al. 2020).

Sample description

We use the 2019 Lightcast data, which includes more than 6.9 million job postings. For a portion of these, Lightcast did not identify any skill requirements ($N = 599,541$, 8.6 percent). This can be because the job title or job description conveys sufficient information or because they include skill requirements that are not recognized by Lightcast's algorithms. While the latter case can be problematic, the large skill taxonomy used by Lightcast captures the most essential information about the skill content of jobs. This is supported by supplementary analysis of a sample of original job posting texts, which shows that those without identified skill requirements are typically for jobs that do not require specific qualifications.[7] Finally, we excluded job postings for which Lightcast could not infer the occupational code ($N = 5,220$, 0.1 percent), yielding a final sample of 6.3 million job postings.

Because the methods used in this article are computationally intensive, we draw a random sample of 600,000 job postings (or 10 percent) to perform our analysis. This sample size ensures optimal statistical properties and preserves the diversity of skill requirements in the sample: the random sample includes 9,065

Table 1: Variables used in the Lightcast UK dataset, January-December 2019.

Variable	Characteristics	Coverage (% of job ads)
Job title	Clean job title, $N = 1,9$ million distinct job titles	100
Occupation	UK SOC Classification (1–4 digits)	99
Skill requirements	Lightcast taxonomy of skills, $N = 11,583$ distinct skill requirements	91
Wage	Minimum and maximum hourly wage	62

of the 11,583 (78.2 percent) distinct skill categories present in the job postings in 2019.[8] We use stratified sampling to make the sample of job postings representative of the occupational structure of the UK labor market, using employment figures from the 2019 UK Annual Population Survey (APS).[9] While previous studies have performed comparisons at the one-digit level only (Sostero and Fernández-Macías 2021; Cammeraat and Squicciarini 2021), we stratify the Lightcast data at the two-digit level to correct for imbalances within some one-digit groups.[10] Figures S1 and S2 in the online supplement indeed show that the over- or under-representation of some major (one-digit) groups in job postings is not uniformly distributed at the sub-major (two-digit) level. An overview of the variables used in our study is shown in Table 1.

To analyze the skill content of occupations, we use the skill requirements extracted by Lightcast from each job posting. The list of identified skills is generally quite short, with job postings in the sample having an average of five skill requirements. However, the distribution is skewed to the right (Figure S3): while 75 percent have 7 or fewer skill requirements, a few job postings (1.6 percent) have more than 20. The distribution of the number of skill requirements is not uniform across occupations. Occupational groups such as plant and machine operators and elementary occupations tend to feature fewer skill requirements than the others ($Q2 = 2$), especially compared to managers, directors, and senior officials ($Q2 = 6$, see Figure S4). The limited information on skill content in the job posting data—and in some occupational groups in particular—can affect the learning of topics. In the next section, we explain how we address this problem by leveraging the co-occurrence patterns of skills at the aggregate level.

To analyze skill content within and between occupations, we use the different levels of the UK SOC: major groups (one-digit, $N = 9$), sub-major groups (two-digit, $N = 25$), minor groups (three-digit, $N = 90$), and unit groups (four-digit, $N = 369$). In the last part of the analysis, we restrict the sample to job postings that propose a wage (62.2 percent), which results in a restricted sample of about 373,500 job postings. Descriptive statistics do not reveal significant differences between the restricted and full samples in terms of occupational or industrial composition (see comparison in Tables S1 and S2). However, job postings that include a wage tend to offer more stable jobs than those that do not, with more permanent positions (+17,7pp) and more full-time contracts (+12.7pp). This means that the results of our wage predictions should not be generalized beyond the available sample—a point we will return to in our discussion.

Lightcast data include both the minimum and maximum hourly wages, but the distributions of these two variables are fairly similar (see Figures S5 and S6). About a third of job postings (37.4 percent) have the same value for minimum and maximum hourly wages, suggesting that recruiters listed a single wage rather than a range. We use the logarithm of the minimum hourly wage in our main analysis and include the results for the maximum wage in the online supplement. As expected, given the similarity of the two variables, the results are very comparable.

Analytical strategy

Identifying skill profiles

The notion of skill profiles builds on prior work mapping occupations in a multidimensional skill space (Yamaguchi 2012; Liu and Grusky 2013), identifying occupational skill portfolios (Poletaev and Robinson 2008; Geel and Backes-Gellner 2011), and detecting data-driven skill communities (Anderson 2017; Alabdulkareem et al. 2018; Djumalieva and Sleeman 2018; Stephany and Teutloff 2024). We extend this work in three ways. First, unlike studies that focus mainly on transversal skills (e.g., Poletaev and Robinson 2008; Yamaguchi 2012), we include the full set of job requirements, combining generic and job-specific skills, since jobs with similar general skills can differ substantially in specialized knowledge. Second, we depart from top-down approaches that predefine skill dimensions; assuming a small set of dimensions risks misrepresenting the skill structure and obscuring how skills co-occur and evolve amid rapid changes in demand. Finally, unlike prior network-based studies that treat skill differences as a partition problem (Anderson 2017; Alabdulkareem et al. 2018; Djumalieva and Sleeman 2018; Stephany and Teutloff 2024), we allow skills—particularly soft or basic ones used across diverse jobs (Burning Glass Technologies 2015)—to belong to multiple latent categories, as implemented with our Latent Dirichlet Allocation (LDA) approach outlined below.

The LDA approach

Text mining has been increasingly used in sociology in the last decade (Macanovic 2022), showing the great potential of computational text analysis (e.g., Kozlowski, Taddy, and Evans 2019). In particular, topic models can be used to identify hidden thematic structures in a set of documents—latent "topics"—hat are both interpretable and sociologically meaningful (DiMaggio, Nag, and Blei 2013; Nelson 2021). Our empirical strategy relies on a widely used Bayesian statistical model, LDA, which is based on two underlying principles (Blei, Ng, and Jordan 2003). First, LDA assumes that in a given text corpus, documents "share the same set of topics, but each document exhibits those topics in different proportion" (Blei 2012:79). Documents are more likely to contain certain topics than others, and they also vary in the extent to which they mix different topics. Second, each topic is defined by a distribution over the corpus vocabulary. Words with high probability on the same topic tend to co-occur frequently in the documents—that is, more often than expected by chance. Following this specification, LDA estimation yields

Table 2: Definition (a) and illustration (b) of the document-topic matrix.

(a) Document-topic matrix, definition

	t_1	t_2	.	.	.	t_K
d_1	p_{d_1,t_1}	.	.	.	.	p_{d_1,t_K}
d_2	.	.				.
.	.		.			.
.	.			.		.
.	.				.	.
d_D	p_{d_D,t_1}	.	.	.	.	p_{d_D,t_K}

(b) Document-topic matrix, illustration

	"Software Engineering"	"Management"	"Graphic Design"
Web developer	0.7	0.1	0.2
Software manager	0.5	0.4	0.1
Graphic designer	0.1	0.1	0.8

Note: We note $d_1, ...d_D$ the D documents, $t_1, ...t_K$ the K topics, and p_{d_i,t_i} the probability that document d_i contains topic t_i. The distribution of each document over the topics adds up to one, i.e., $\sum_{k=1}^{K} p_{d_m,t_k} = 1$ for all topic d_m in $d_1, ...d_D$. The example is based on a fictional case with job postings for three job titles (Web developer, Software manager, Graphic designer) and three topics ("Software engineering", "Management", "Graphic Design").

the following two key outputs: the document-topic matrix, which provides the distributions of documents over the set of topics, and the topic-term matrix, which provides the distributions of topics over the set of words (see Tables 2-3).

By treating job postings as documents and skill requirements as words, we use LDA to accomplish two key objectives: (a) identify a data-driven categorization of skills by detecting latent skill topics (henceforth referred to as skill categories) and (b) operationalize the notion of skill profile.

First, LDA reduces the complexity of job skills data inductively, based on the relations between skills within job postings. This data-driven dimension reduction avoids the need to *a priori* decide which skills are the most relevant. In particular, we are able to include both general *and* job-specific skills, rather than solely concentrating on the former (see, e.g., Poletaev and Robinson 2008; Yamaguchi 2012). Additionally, LDA identifies latent skill categories based on the relations between skills in job postings and their shared context. This approach goes beyond the mere categorization of skills that are similar in standard skill taxonomies, as it also clusters skills that are *complementary*, enabling the study of different combinations of skills (e.g., see the topic "Graphic Design" in Table 3b). Moreover, through the topic-term matrix, LDA provides a flexible classification where skills belong to all categories with different probabilities. This departs from the unrealistic assumption that each skill fits into a single category, which is made in virtually all existing skill classifications (e.g., Tippins and Hilton 2010; Le Vrang et al. 2014; Djumalieva and Sleeman 2018). Instead, it highlights whether and how the same skill can be used differently in different job contexts.

Table 3: Definition (a) and illustration (b) of the topic-term matrix.

(a) Topic-term matrix, definition

	w_1	w_2	.	.	.	w_N
t_1	p_{t_1,w_1}	.	.	.	.	p_{t_1,w_W}
t_2	.	.				.
.	.		.			.
.	.			.		.
.	.				.	.
t_K	p_{t_K,w_1}	.	.	.	.	p_{t_K,w_W}

(b) Topic-term matrix, illustration

	Budgeting	Creativity	Java	Adobe InDesign
"Management"	0.9	0.1	0.0	0.0
"Graphic Design"	0.0	0.4	0.0	0.6
"Software Engineering"	0.0	0.0	0.8	0.2

Note: We note $t_1, ...t_K$ the K topics, $w_1, ...w_N$ the N words from the corpus vocabulary and p_{t_i,w_i} the probability that topic t_i contains word w_i. The distribution of each topic over the words adds up to one, i.e., $\sum_{n=1}^{N} p_{t_k,w_n} = 1$ for all topic t_k in $t_1, ...t_K$. The example is based on a fictional case with three topics ("Software engineering", "Management", "Graphic design") and a vocabulary of four skill words (Budgeting, Creativity, Java, AdobeInDesign).

Second, LDA enables us to represent the job postings in a skill space that is both low-dimensional and meaningful. The document-topic matrix describes each job posting by its probabilities of containing the latent skill categories. This reduces the dimension of the skill space from 9,065—the number of distinct skills in the data—to k, the number of topics (see illustration Figure 1). In addition to being parsimonious, this representation of job postings provides a rich description of the type of skills desired by employers. It not only indicates what categories of skill are most prevalent in a given job posting but also the extent to which it mixes multiple skill categories. For instance, certain job postings may mainly contain one skill category (e.g., the Graphic designer in Table 2b), while others may require a hybrid set of skills that span several skill categories (e.g., the Software manager). Such a representation of job postings is consistent with the conceptualization of job skills outlined above; it acknowledges that the execution of jobs depends on a specific combination of skills that cannot be unbundled (Lazear 2009; Autor and Handel 2013). We therefore measure the skill profiles of job postings by their probability distributions across the latent skill categories.

The strengths of the LDA model also come with a limitation: it infers topics from word co-occurrence patterns at the document level, which can be sparse in short documents like those in the Lightcast data. To circumvent this issue, we use a variant designed for short texts (Yan et al. 2013), the biterm topic model (BTM). For technical details and a comparison with standard LDA, see Section 2.1 in the online supplement.

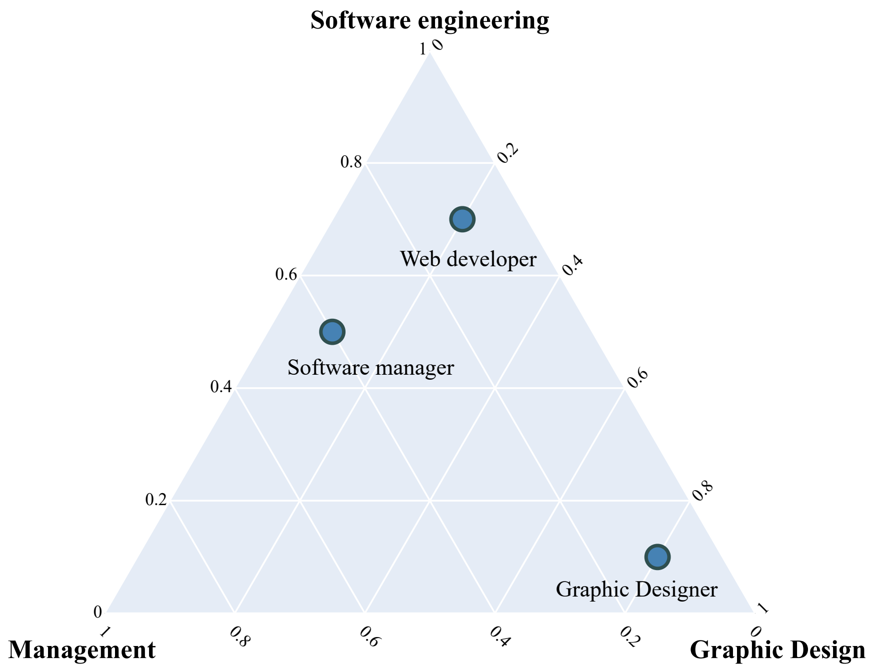


Figure 1: Latent skill space, illustration.
Note: We represent the latent skill space defined by the three topics "Software Engineering", "Management", "Graphic Design" defined in Table 3b. This takes the shape of a triangle, each corner of which corresponds to a topic. Job postings can be positioned in this space (blue circles) depending on their distribution over the topics, as defined in Table 2b. The scale on each side indicates the probability of containing each topic, from 0 to 1. For example, the job posting for a Graphic designer is in the bottom right corner because it contains almost exclusively the topic "Graphic Design", with a probability of 0.8.

Model parameters

Several factors influence the learning of topics, the most prominent being the number of topics K, which needs to be specified in advance. As DiMaggio et al. (2013:582) note, choosing K is a matter of finding the right lens for interpretation: it is neither inherently good nor bad but depends on the desired level of granularity. Quantitative measures like perplexity can guide this choice, but optimizing them does not guarantee semantically meaningful topics (Chang et al. 2009). Evaluating coherence using reference corpora such as Wikipedia (Röder, Both, and Hinneburg 2015) is not feasible here due to the lack of a suitable benchmark for assessing the interpretability of skill categories.[11] We therefore prioritized the model's ability to identify semantically *significant* and *distinct* topics, assessed via manual inspection of the topic-word matrix and statistical checks. Section 2.1.3 in the online supplement details the selection of our 19-topic solution and other model parameters.

Comparing occupations

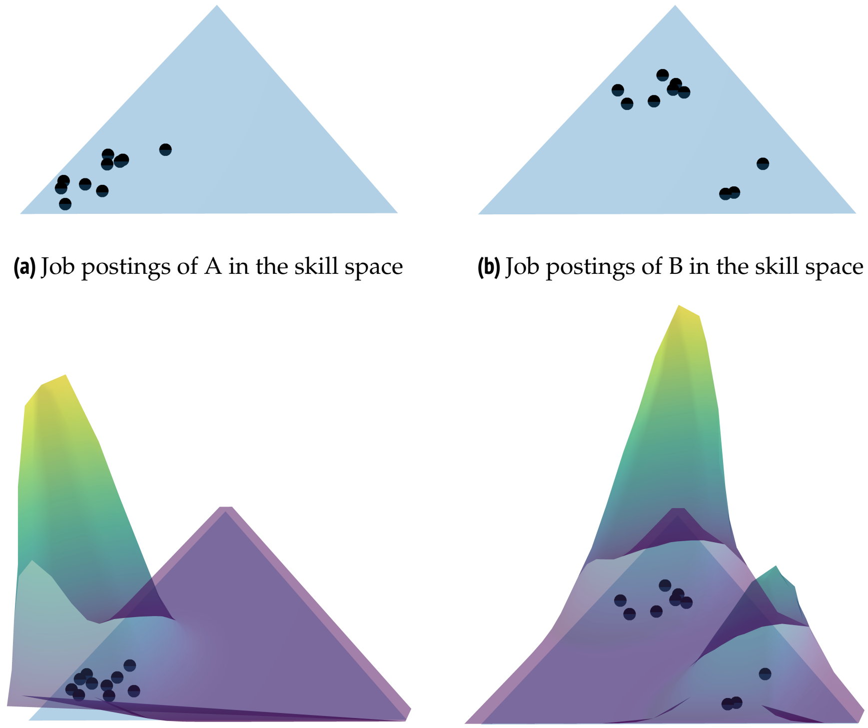
The BTM characterizes each job posting by its skill profile, that is, its probability distribution over 19 latent skill categories. To assess whether occupations are coherent skill bundles, we aggregate these job-level profiles at the occupation level.

Aggregation must account for the non-Euclidean nature of the 19-dimensional probability space, which makes standard measures like the mean inappropriate (Greenacre 2021). It must also preserve significant variation in skill profiles *within* occupations, since our goal is to study potential heterogeneity within occupational categories.

We compare the skill profiles of occupations in two steps. First, we construct empirical distributions over the job postings at the occupational level using kernel density estimation. For each occupation, we construct an empirical distribution that gives the probability of occurrence of each skill profile in the job postings of this occupation. These empirical distributions enable us to retain a maximum amount of information on the diversity of skill combinations within each occupational category, without making specific assumptions about the shape of those distributions. Second, we compare these occupational-specific empirical distributions using the maximum mean discrepancy (MMD) distance, which can be technically defined as a kernel-based metric on the space of probability distributions (Schrab et al. 2023:3).[12] The MMD distance is very common in machine learning because of its interesting features: it is flexible, robust, and implementable in high dimensions (Gretton et al. 2012). Intuitively, the MMD distance is inversely proportional to the amount of shared volume between two distributions.[13] The more two occupational categories contain job postings with similar skill profiles, the smaller the MMD distance between them.

Our approach is illustrated in Figure 2 with a latent skill space of three topics. We distinguish between two occupations, A and B, each defined by 10 job postings with contrasting skill profiles. While the job postings in occupation A are all focused on the left-hand topic (b), the distribution of job postings in occupation B is less concentrated: most job postings contain the upper topic, but a small fraction is more focused on the right-hand topic (b). The empirical distributions shown in the lower panels capture these different patterns, with a unimodal distribution centered in the left-hand topic for A (c) and a bimodal distribution spread over the other two topics for B (d). In this example, the MMD distance between the two occupations is expected to be very large because there is only marginal overlap in the empirical distributions of A and B. We construct such empirical distributions for each occupation on our 19-dimensional skill space and calculate the MMD distance between each pair of occupations to assess the extent to which they have distinct skill profiles.

This approach does not provide direct information on whether the skill profiles are homogeneous *within* occupational categories (cohesion) but rather indicates the extent to which the skill profiles are distinguishable between them (separation). However, we compute and compare the MMD distance between occupations at different levels of the occupational classification to approximate the cohesion of occupational categories. The intuitive idea is that if occupational categories are good clusters of skills, the skill profiles in a given sub-occupational category (e.g., sub-major group 21) should be closer to sub-occupational categories in the same occupational category (the major group 2, i.e., 22, 23, 24) than to other sub-occupational categories (e.g., 31, 32). This mimics the logic behind standard measures of cluster consistency, such as the silhouette width (Rousseeuw 1987), without making specific assumptions about the shapes of the empirical distributions of the occupational categories.[14]



(a) Job postings of A in the skill space

(b) Job postings of B in the skill space

(c) 3D empirical distribution of A

(d) 3D empirical distribution of B

Figure 2: Illustrations of the empirical distributions of two occupations A and B, defined on a latent skill space with three topics.
Note: Flat triangles in (a) and (b) represent the latent skill space, with each corner corresponding to a topic (see Figure 1). There are 10 job postings per occupation, whose positions in the skill space are represented by black dots. Using kernel density estimation, we construct 3-dimensional empirical distributions in (c) and (d) that give the probability of occurrence of each combination of topics (or skill profile) in the job postings of each occupation. The color gradient is only there to facilitate the interpretation of the 3D distributions.

Predicting the wage offered in job postings

Our topic modeling approach allows us to capture within-occupation variation at a granular level, but we must ensure that observed differences in skill profiles "capture substantive differences, rather than noise, in job content" (Autor and Handel 2013:81), since some variation may reflect idiosyncratic differences between jobs that are not economically or sociologically meaningful.

To validate our empirical approach, we assess the effect of the skill profiles of jobs on the wage offered in the job posting. In a standard OLS regression model, we predict the logged minimum hourly wage[15] offered in the job postings by their probabilities (or *loads*) on the $K - 1 = 18$ topics. This leads to the model:

$$log(w) = \alpha + \sum_{k=1}^{18} \beta_k p_k + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2) \tag{1}$$

where $log(w)$ is the logged hourly wage offered in the job posting and p_k is the probability of the job posting to contain the latent skill category k. We use the coefficient of determination R-squared (R^2) to assess the explanatory power of the skill profiles given by $p_1, ...p_{18}$. As a benchmark, we also estimate the model with the occupational categories, which are often identified as a major source of wage inequality across workers (e.g., Mouw and Kalleberg 2010; Williams 2013):

$$log(w) = \alpha + \sum_{s=1}^{n_{soc}-1} \gamma_s \mathbb{1}_{\{soc=s\}} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2) \tag{2}$$

where soc is the Standard Occupational Classification coding of the job postings. We estimate model (2) with each of the four levels of the occupational classification, from 1 ($n_{soc} = 9$) to 4 digits ($n_{soc} = 369$), resulting in four regressions. In a final step, we use the skill profiles and occupational coding of jobs as simultaneous predictors to analyze the extent to which they provide redundant or complementary information in predicting the offered wage:

$$log(w) = \alpha + \sum_{k=1}^{18} \beta_k p_k + \sum_{s=1}^{n_{soc}-1} \gamma_s \mathbb{1}_{\{soc=s\}} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2) \tag{3}$$

As in Model (2), we estimate a separate regression for each level of the occupational classification, resulting in four regressions. If, as is often assumed, the share of wage variation explained by occupational categories is mainly due to skill differences between occupational groups, the coefficients of determination of the regressions from model (3) should improve only marginally over those from model (2). Conversely, a significant increase in the R^2 coefficients would provide evidence that the explanatory value of occupations is mainly driven by characteristics *other than* skill content. We therefore take advantage of our detailed and direct measure of skill content to determine whether occupations predict wages either *because* or *independently* of the different skills that jobs require.

Results

Describing job ads with latent skill categories

A data-driven classification of skills

Table 4 displays the 19 latent skill categories (or topics) identified by the LDA biterm model, listing the 15 highest-ranked skill requirements (or terms) for each category.[16] To facilitate the interpretation, the ranking is based on the *relevance* of terms for each topic, which corrects the probability of terms for topics by their prevalence in the whole corpus (see Sievert and Shirley 2014:66). This limits the overrepresentation of the most common skill requirements in the job ads, such as "communication skills" (present in 24.0 percent of the sampled job postings) and "organizational skills" (11.2 percent). The latent skill categories have been labeled by ChatGPT using the top-100 highest term probabilities for each topic.[17]

Table 4: Description of the 19 latent skill categories obtained from the biterm topic model, using the 15 most relevant skills by decreasing order.

	Latent skill category	The 15 most relevant skills, by decreasing order ($\lambda = 0.7$)
1	Project Management	budgeting, project management, planning, communication skills, stakeholder management, people management, building effective relationships, key performance indicators KPIs, quality management, staff management, problem solving, teamwork/collaboration, performance management, change management, organizational skills
2	Office administration and management	Microsoft Excel, administrative support, organizational skills, Microsoft Office, communication skills, detail-orientated, Microsoft Word, secretarial skills, typing, Microsoft Powerpoint, customer service, spreadsheets, data entry, Microsoft Outlook, general office duties
3	Communication and Interpersonal Abilities	communication skills, detail-orientated, teamwork/collaboration, customer service, organizational skills, problem solving, writing, English, verbal/oral communication, time management, Microsoft Excel, listening, French, German, Spanish
4	Sales and Business Development	sales, business development, sales management, sales goals, account management, customer service, product sales, communication skills, business-to-business, prospective clients, building effective relationships, customer contact, client-base retention, telesales, teamwork/collaboration
5	Caregiving and Support Services	teaching, working with patient and/or condition: mental health, care planning, childcare, dementia knowledge, communication skills, nursing home, autism diagnosis treatment care, creativity, planning, English, home management, social services, learning disability, organizational skills
6	Customer Service and Retail Operations	cleaning, customer service, communication skills, cooking, retail industry knowledge, food safety, teamwork/collaboration, organizational skills, stock control, cash handling, food preparation, store management, English, housekeeping, detail-orientated
7	Digital Marketing and Content Strategy	social media, marketing, digital marketing, creativity, marketing management, Google Analytics, market strategy, content management, copy writing, editing, writing, email marketing, budgeting, social media tools, research
8	Financial Operations	accounting, finance, account reconciliation, balance sheet, Microsoft Excel, budgeting, VAT returns, financial reporting, financial accounting, bank reconciliation, accruals, statutory accounts, communication skills, detail-orientated, variance analysis
9	Web Development and Software Engineering	Javascript, Microsoft hash, software development, Java, NET, SQL, Git, software engineering, DevOps, Scrum, Python, active server pages ASP, ASP.NET, continuous integration CI, AngularJS
10	Logistics and Supply Chain Management	procurement, purchasing, supply chain knowledge, logistics, supply chain management, enterprise resource planning ERP, SAP, procurement contracts, planning, manufacturing resource planning MRP, contract management, Microsoft Excel, key performance indicators KPIs, material requirement planning MRP, communication skills
11	Facility Maintenance	plumbing, preventive maintenance, predictive preventative maintenance, electrical work, carpentry, painting, HVAC, boilers, wiring, communication skills, hand tools, heating systems, customer service, cleaning, emergency lighting
12	Healthcare and Patient Care	patient care, surgery, anaesthesiology, communication skills, working with patient and/or condition: trauma, dentistry, orthopaedics, xrays, critical care, paediatrics, primary care, rehabilitation, gynecology, blood pressure measurement, urology

Table 4: (*Continued*)

	Latent skill category	The 15 most relevant skills, by decreasing order ($\lambda = 0.7$)
13	Business strategy	risk management, communication skills, business development, research, due diligence, project management, building effective relationships, analytical skills, economics, accounting, Microsoft Excel, asset management industry knowledge, insurance underwriting, teamwork/collaboration, planning
14	Engineering and Technical Expertise	AutoCAD, mechanical engineering, engineering design and installation, commissioning, engineering design, calculation, civil engineering, project management, mechanical design, Revit, systems engineering, simulation, communication skills, electronics industry knowledge, planning
15	Manufacturing and Engineering	computer numerical control CNC, machining, engineering drawings, welding, lathes, manufacturing processes, quality assurance and control, ISO9001standards, MIG and TIG welding, lean manufacturing, milling cutters, machine tools, computerised numerical control lathes, quality management, problem solving
16	Data Management and Analysis	SQL, python, machine learning, data science, tableau, data warehousing, Microsoft Power BI, extraction transformation and loading ETL, data analysis, business intelligence, big data, Apache Hadoop, SQL server reporting services SSRS, data architecture, Microsoft SQL server integration services SSIS
17	Technical Support and Troubleshooting	Microsoft Active Directory, VMware, Windows Server, Cisco, troubleshooting, ITIL, domain name system DNS, Microsoft Windows, IT support, wide area network WAN, technical support, dynamic host configuration protocol DHCP, transmission control protocol/internet protocol TCP/IP, Linux, Microsoft Exchange
18	Graphic Design and Creative Media	Adobe Photoshop, Adobe InDesign, Adobe Acrobat, Adobe Creative Suite, Adobe Illustrator, creativity, graphic design, Adobe After Effects, digital design, typesetting, animation, editing, creative design, detail-orientated, teamwork/collaboration
19	Scientific Research and Laboratory Work	research, chemistry, biology, clinical trials, biotechnology, biochemistry, molecular biology, experiments, clinical research, cancer knowledge, oncology, drug development, bioinformatics, high performance liquid chromatography HPLC, drug discovery

Note: We use the *relevance* of the skills as defined by Sievert and Shirley (2014) to account for the fact that some skill requirements (e.g. communication skills) are listed much more frequently than others, which mechanically increases their probability of belonging to any latent skill category. The parameter λ determines the weight given to the probability of a term under a given topic relative to its marginal probability in the corpus (Sievert and Shirley 2014:66). We observed that $\lambda = 0.7$ provides optimal interpretability, but note that the latent topics are easily interpretable without this correction (see Table S3 in the online supplement).

We observe a few categories of mainly generic skills, such as the categories "Project Management" and "Communication and Interpersonal Abilities." They include both interpersonal (e.g., people management) and intrapersonal (e.g., problem-solving) skills, as well as basic ICT tools (e.g., Microsoft Word, Excel) and basic activities (e.g., planning, writing, typing). The other latent categories focus on job-specific skill requirements, including specialized knowledge (e.g., supply chain knowledge, category 10), specific techniques (e.g., MIG and TIG welding, 15), and know-how (e.g., blood pressure measurement, 12). Specific software and ICT tools often act as discriminating skill requirements, even in non-ICT-focused categories (e.g., AutoCAD in 14 or the Adobe Suite in 18).

Importantly, the latent categories that emphasize specialized and technical skills are also characterized by specific generic skills. Indeed, we can observe that some intrapersonal skills are prominent in some categories but not in others, such as "creativity" for the categories related to marketing and graphic design (categories 7 and 18). This is consistent with the idea that specialized and generic skills are closely intertwined at the job level (Fernández-Macías and Bisello 2020). Other generic skills are much less distinctive, such as "communication skills" and "collaboration/teamwork," which rank high in most latent skill categories. While employers may include such requirements in their job descriptions for strategic purposes (Backhaus 2004), the ubiquity of these terms may also reflect the growing importance of so-called "social skills" in the labor market (Deming 2017).

Table 4 suggests that the latent skill categories vary greatly in the type of skill requirements they contain, yet the table only includes the most discriminating terms for defining the topics. To get a more comprehensive overview, we analyze the composition of latent skill categories by looking at their full distribution over the skills. For the sake of illustration, we use simple dummies from Lightcast that identify two types of skills: software skills and baseline skills. The latter are defined as skills that employers seek across various occupations and that are not typically taught in training programs, such as "organizational skills" or "meeting deadlines" (Burning Glass Technologies 2015:7). We refer to skills that do not fit within either of these categories as "job-specific".

Figure 3 sums the probabilities of skill requirements for each latent skill category based on their categorization into the three skill types. Even with this crude measure, we can observe important variations in the composition of the latent skill categories. In particular, categories that appear similar at first glance, such as "Web development and Software engineering" and "Technical Support and Troubleshooting," differ substantially in their proportions of software skills, with 0.59 and 0.31, respectively. Another example is the skill category "Healthcare and Patient Care," which is significantly less likely to contain a baseline skill (0.16) than the other health-related category "Caregiving and Support Services" (0.30). This means that the same health-specific skill (e.g., "dementia knowledge") will be more likely to be featured with other expert knowledge in the category "Healthcare and Patient Care" than in the category "Caregiving and Support Services," which contains more non-health-specific general skills. This highlights the heterogeneity of latent skill categories, which are characterized not only by different most likely

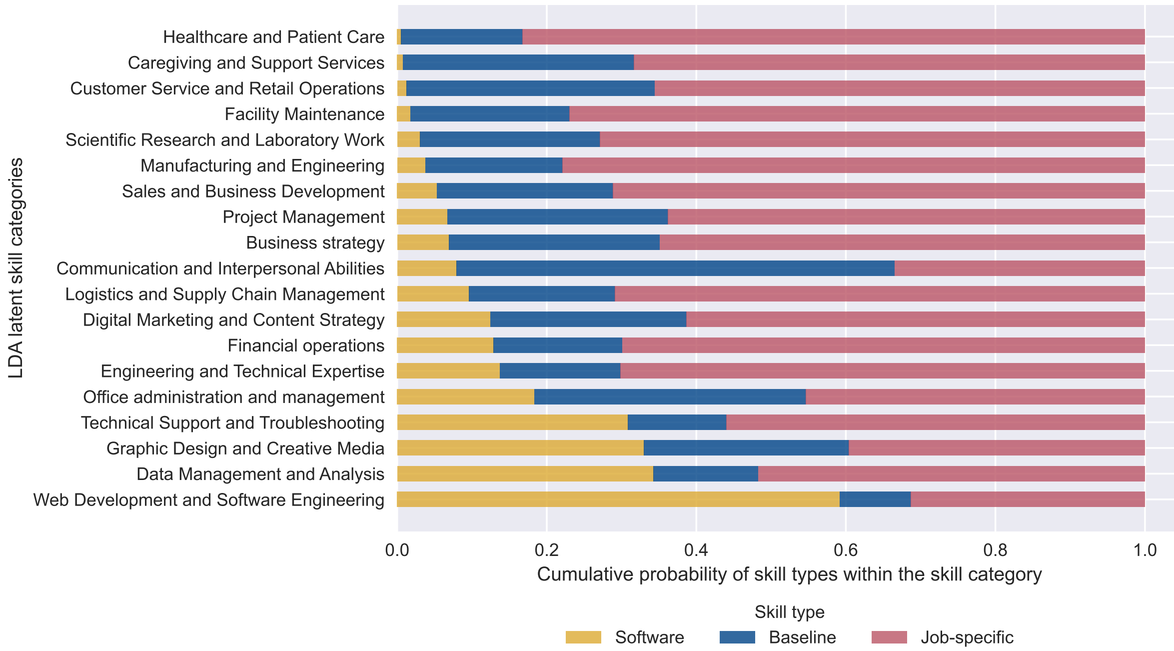


Figure 3: Cumulative probability of three skill types within each latent skill category, based on Lightcast coding of skill requirements.
Note: Lightcast data includes dummy variables for identifying software and baseline skill requirements. For a definition of the latter, see Burning Glass Technologies (2015:7).

skill requirements as illustrated in Table 4 but also by how and to what extent they articulate different types of skills.

Jobs as skill profiles

The LDA biterm model provides meaningful latent categories that capture both the similarity and complementarity of skills at the job level. We can then position job postings within this latent skill space. In other words, we can analyze whether and how job postings tend to draw on different latent skill categories. Figure S13 gives the distribution of the highest probability for topics within the sample. The median for the highest probability is 0.54, which means that half of the sampled job postings have a majority skill category that is more likely to be featured than all the others combined. The probability for the most likely category even exceeds 0.8 for about 20 percent of the job postings, suggesting that they are highly specialized. At the other end of the spectrum, a few job postings (about 5 percent) have a maximum probability below 0.2, which implies that their distribution across topics is roughly uniform. Overall, these findings point to substantial heterogeneity in the skill profiles of jobs. While half of the job postings focus predominantly on a single latent skill category, the other half mix several categories in varying proportions.

To further illustrate the differences between specialized and mixed skill profiles, we provide examples in Tables 5 and 6 from two job postings advertising a job within the unit group 8133 (“Routine inspectors and testers”). The job postings

Table 5: Example of a specialized skill profile for a Legionella Risk Assessor, with the skill requirements extracted from the original job posting (a), and the corresponding skill profile (b). p gives the probability that the job posting contains a given latent skill category; categories with a probability below 0.00 are grouped together in the last row.

(a) Extracted skill requirements	
Water Treatment	
Cooling Towers	
Working With Patient And/Or Condition: Legionella	
Equipment / Instrument Sterilization	
Organizational Skills	
Water Sampling	
(b) Corresponding skill profile	
Latent skill category	p
Facility Maintenance	0.89
Healthcare and Patient Care	0.03
Office administration and Management	0.02
Customer Service and Retail Operations	0.02
Sales and Business Development	0.02
Project Management	0.01
Engineering and Technical Expertise	0.01
Other skill categories	0.00

Table 6: Example of a generalist skill profile for a Legionella Risk Assessor, with the skill requirements extracted from the original job posting (a), and the corresponding skill profile (b). p gives the probability that the job posting contains a given latent skill category; categories with a probability below 0.00 are grouped together in the last row.

(a) Extracted skill requirements	
HSG (Health & Safety Guidance)	
Microsoft Word	
Site Surveys	
Microsoft Excel	
Quality Assurance and Control	
Working With Patient And/Or Condition Legionella	
(b) Corresponding skill profile	
Latent skill category	p
Facility Maintenance	0.39
Office administration and Management	0.25
Project Management	0.19
Engineering and Technical Expertise	0.09
Sales and Business Development	0.02
Communication and Interpersonal Abilities	0.02
Logistics and Supply Chain Management	0.01
Manufacturing and Engineering	0.01
Caregiving and Support Services	0.01
Other skill categories	0.00

advertise the same job ("Legionella Risk Assessor") and have the same number of skill requirements (n=6), yet their content differs significantly: while the first job posting focuses primarily on technical knowledge and know-how (e.g,. water sampling, equipment sterilization, Table 5), the second includes office tools and emphasizes regulatory aspects, such as the Health & Safety Guidance (Table 6). This suggests that the first position requires on-site work, while the second involves more office work and coordination. These differences are reflected in the latent skill profiles of the job postings. The first job posting gets a very high load on the topic "Facility Maintenance" ($p = 0.9$), while the probability of the second job posting is distributed over three main topics: "Facility Maintenance" ($p = 0.4$), "Office Administration and Management" ($p = 0.2$), and "Project Management" ($p = 0.2$). This illustration shows that our measure of skill profile captures granular differences in the skill content of job postings, reflecting whether and how they articulate different categories of skills.

Figure S15 shows the marginal probability distributions per topic, which allows us to analyze in more detail how skill categories are distributed within the job postings. As expected, the distributions of the generic skill categories (1, 2, 3 in Table 4) have most of their mass between 0 and 0.4, which indicates that they mainly *complement* other categories and rarely characterize a job posting on their own. This is especially true for the category "Communication and Interpersonal Abilities," which has no mass above 0.8. Other distributions (e.g., for 9, 11, 12) follow a U-curve: these skill categories tend to play either a negligible or a predominant role in the definition of job postings. As one might expect, these are the ones that include the most job-specific skills, such as the category "Healthcare and Patient Care" (see Figure 3). In contrast, some topic distributions are fairly flat, which implies that skill categories can either define a job by themselves or complement other skills. This is, for example, the case for "Financial Operations," which can be either the primary or secondary activity of a job, such as for a corporate accountant or business manager, respectively. Finally, some topic distributions have most of their mass around 0, suggesting that the skill category may only play a role in a small fraction of ads (e.g., "Scientific Research and Laboratory Work").

These findings underscore the relevance of our conceptualization of jobs as skill profiles. First, they highlight the heterogeneous nature of skill requirements: job postings typically bundle different types of skills (e.g., knowledge, know-how, professional attitudes) and vary in their combinations of different skill domains (e.g., office administration, logistics, manufacturing). This heterogeneity is largely hidden when mostly generic skills are considered, as is often done in the literature, at the expense of job-specific skills (e.g., Liu and Grusky 2013; Cheng and Park 2020; Lin and Hung 2022). Second, the topic distributions indicate that the same skill category can play different roles depending on the type of job: while some skills are the primary feature of some job postings, they appear only secondarily in others. This supports the idea that jobs differ in the extent to which they require different and/or complementary skills (Anderson 2017; Stephany and Teutloff 2024). With the BTM, we can map the skill content of jobs without obscuring these heterogeneous patterns, which may appear not only between but also within occupations.

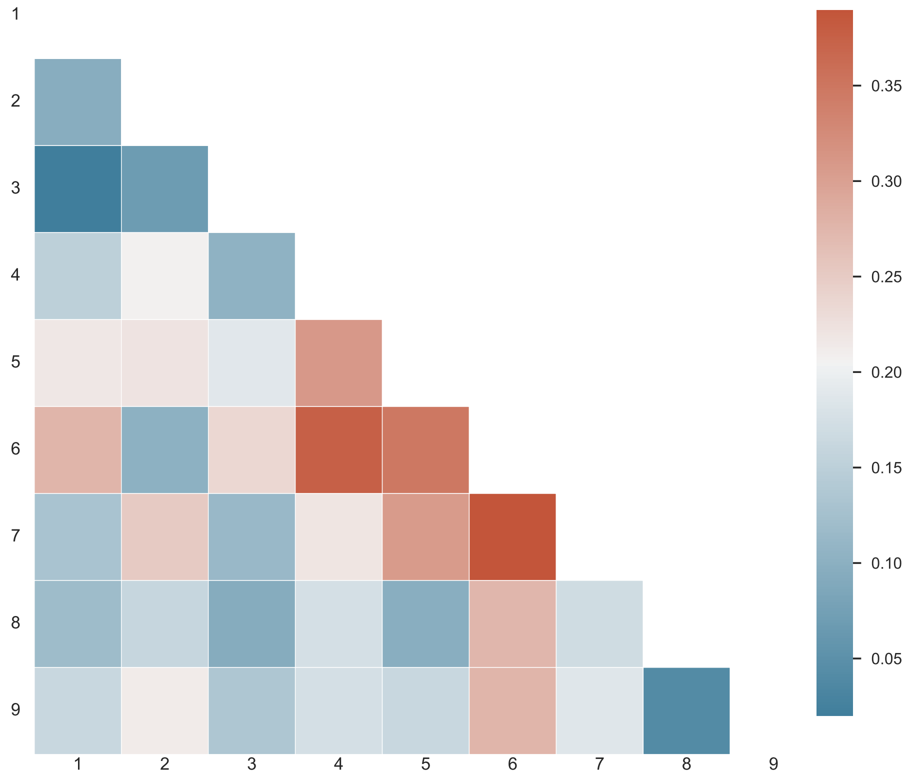
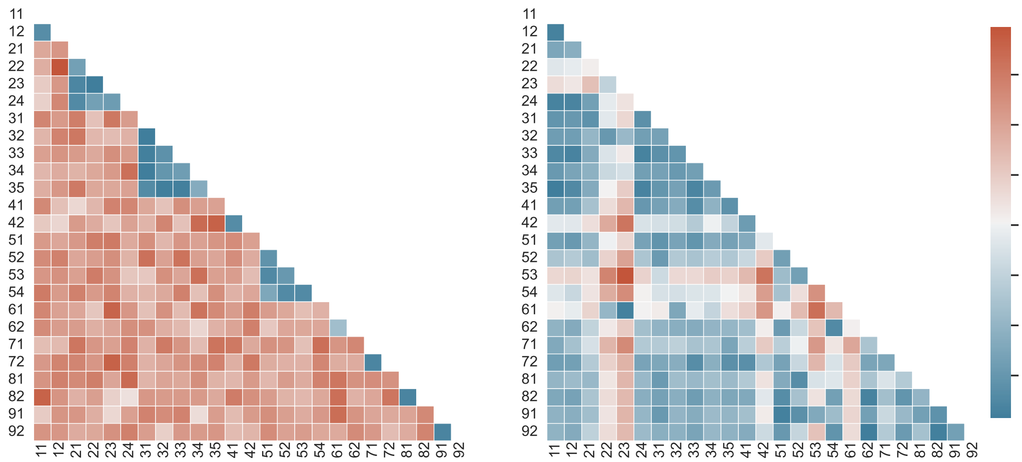


Figure 4: Empirical heat map for the MMD distance between major groups.
Note: each cell represents the Maximum mean discrepancy pairwise distance between major groups, with a kernel bandwidth of $\lambda = 1.49$.

Skill profiles within and between occupations

Next, we analyze whether occupations are distinct and homogeneous skill bundles based on the representation of their respective job postings in our 19-dimensional latent skill space. We calculate the MMD pairwise distance for all pairs of occupational categories, starting with the most aggregated level of the SOC classification, the nine major groups.

Figure 4 shows the distance between each major group with a heat map, where the bluer the smaller and the redder the greater the MMD distance. We can see from the predominant light blue pattern that the skill distributions of the major groups are overall at a similar distance from each other. Major group 6, "Caring, Leisure and Other Service," shows a distinctive pattern and tends to be further away from all other major groups, notably 4 ("Administrative and Secretarial") and 7 ("Sales and Customer Service"). The patterns as found in Figure 4 do not align with the implicit hierarchy of occupational categories, which would imply that the categories furthest apart in the SOC classification, such as the managers and directors (1) and elementary occupations (9), are also the furthest apart in terms of skill profiles. However, we lack a tangible benchmark to assess the significance



(a) Simulated heat map **(b)** Empirical heat map

Figure 5: Simulated and empirical heat maps for the MMD distance between sub-major groups.
Note: The x- and y-axes are labeled from the coding in the SOC occupational classification. The simulated heat map (a) represents the theoretical expectation that sub-major occupations in the same major group should be closer to each other than to sub-major occupations in other major groups. It is based on a simulation with normal Gaussian distributions and is for illustrative purposes only. In the empirical heat map (b), each cell indicates the maximum mean discrepancy pairwise distance between sub-major groups based on their representation in the latent skill space. We use a kernel bandwidth of $\lambda = 1.49$.

of these (dis)similarities. Furthermore, Figure 4 does not give information about the variation of skill profiles *within* major groups.

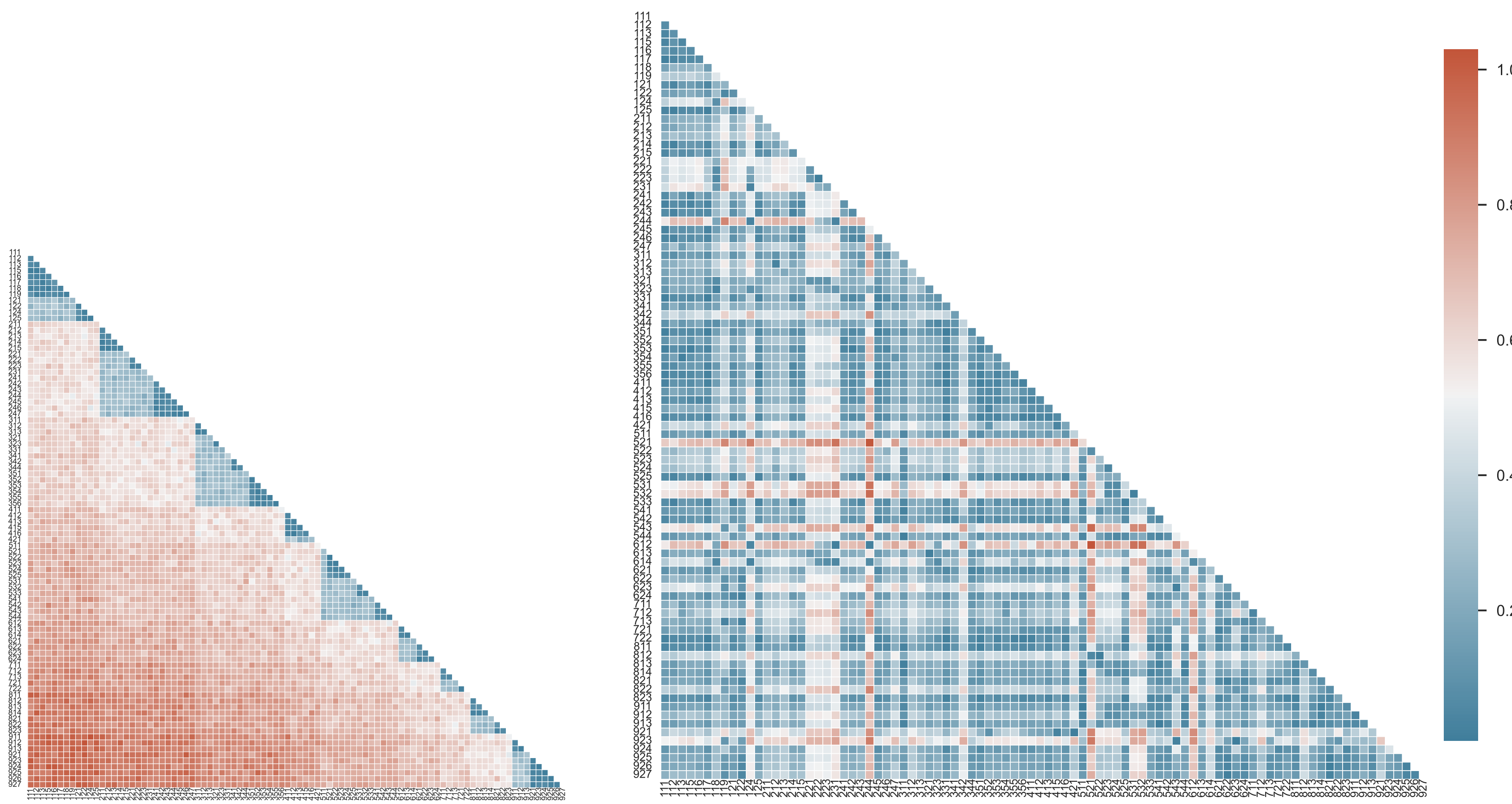
To explore the internal consistency of major groups, we delve deeper into the classification, examining the MMD distance between the 25 sub-major groups. Figure 5(a) provides a simulated heat map to ease the interpretation. It represents the theoretical expectation of major groups as coherent clusters of skills, with sub-major occupations in the same major group being closer to each other than to occupations in other major groups. The simulated heat map and the empirical heat map (Figure 5(b) show very distinct patterns. Most sub-major occupations are at fairly similar distances from each other, regardless of whether they are part of the same major group. Similar to what we observed in Figure 4, there are a few sub-major groups that stand out and are far away from most of the other sub-major groups. This is particularly the case for occupations 22 ("Health Professionals"), 23 ("Teaching and Educational Professionals"), and 53 ("Skilled Construction and Building Trades"). Conversely, we observe that some sub-major groups are remarkably close to each other even though they are far apart in the SOC classification, such as the Health and Teaching professionals (22, 23) with 61 "Caring Personal service occupations." This is coherent from a skill perspective, as workers in sub-major group 61 are notably described as assisting health professionals and teachers in the care of patients and children.

These results challenge the conventional view of occupational major groups as bundles of skills. Figure 4 shows no evidence that the skill profiles of one-digit occupations are well separated, while Figure 5 highlights a low degree of cohesion within these categories when looking at the skill profiles of the two-digit occupations they contain. However, it could be argued that due to the high level of aggregation of the major groups, their skill profiles are necessarily highly heterogeneous. In fact, most researchers rely on the two- or three-digit level when analyzing the effect of occupations on labor market outcomes (e.g., Mouw and Kalleberg 2010; Avent-Holt et al. 2020; Haslberger 2022). We therefore replicate the analysis at a more detailed level and examine the relevance of the sub-major and minor groups for inferring the skill content of jobs.

To assess the heterogeneity in skill profiles within sub-major groups, we examine the MMD distance between the 90 minor groups in Figure 6. As previously, we simulate a heat map (a) to give a visual representation of good cluster quality for major and sub-major groups: minor groups should be very close to minor groups in the same sub-major group, relatively close to other minor groups in the same major group, and further away from minor groups in a different major group. The empirical heat map (b) shows starkly different patterns. While there is evidence of some clusters of minor groups, they do not intersect with the sub-major groups. The most salient example is provided by the minor groups 221 ("Health Professionals") to 231 ("Teaching and Education professionals"), which are not only close to each other but also have similar pairwise distances with all other minor groups. This suggests that these minor groups do share a distinctive skill profile, despite being placed in different sub-major groups (22 and 23). Another example is sub-major group 53, where two out of the three minor groups show clustering patterns. These are occupations 531 ("Construction and Building Trades") and 532 ("Building Finishing Trades"), which both include different types of construction workers, such as roofers, carpenters, and plasterers. In contrast, occupation 533 groups workers with a supervisory and control role ("Construction and Building Trades Supervisors"), leading to a distinct skill profile.

Other minor groups tend to be far away from most other groups, such as 244 ("Welfare Professionals"), 521 ("Metal Forming, Welding and Related Trades"), and 612 ("Childcare and Related Personal Services"). However, beyond these few examples, Figure 6(b) shows that most minor groups are at relatively similar distances from each other. This does not mean that their skill content is comparable, but rather that the extent to which the skill content varies between minor groups remains relatively stable across the occupational structure. In other words, most three-digit occupations are as (dis)similar to each other when they are in the same one- or two-digit occupational category as when they are in different occupational categories. This indicates that the skill structure of the labor market is largely misaligned with the occupational structure as measured by the major or sub-major SOC groups.

Importantly, the same observation holds when we examine the MMD distance between the 369 unit groups to assess the cluster quality of minor groups (see Figure S18): we observe similar clusters to those identified in Figure 5, but they do not coincide with the boundaries of the minor group classification. This indicates that even at the three-digit level, occupational categories are not a good proxy for the skill content of jobs. While some previous research has argued that there are homogeneity-inducing mechanisms at the detailed occupational level (Weeden and



(a) Simulated heat map

(b) Empirical heat map

Figure 6: Simulated and empirical heat maps for the MMD distance between minor groups.
Note: The x- and y-axes are labeled from the coding in the SOC occupational classification. The simulated heat map (a) represents the theoretical expectation that minor occupations in the same sub-major group should be closer to each other than to minor occupations in other minor groups. It is based on a simulation with normal Gaussian distributions and is for illustrative purposes only. In the empirical heat map (b), each cell indicates the maximum mean discrepancy pairwise distance between minor groups based on their representation in the latent skill space. We use a kernel bandwidth of $\lambda = 1.49$.

Grusky 2005:149), our results suggest that these mechanisms are not related to the type and combination of skills that workers need to perform on the job. On the contrary, the homogeneity in skill content across jobs appears to be related to other dynamics that have yet to be explored.

All these results point to the same key finding: occupations are neither distinct nor coherent bundles of skills. This is true for both the major groups, which represent a big-class map of the labor market (Erikson and Goldthorpe 1992), and the detailed minor groups, which come close to the microclass approach championed by Weeden and Grusky (2005). Jobs in the same occupational category tend to require dissimilar skill profiles, which may reflect jobs of varying levels of complexity and authority. This is particularly true when occupational categories include jobs with different levels of seniority, some of which consist of managing or supervising the work of others (see, e.g., sub-major groups 53, 62, 71). Conversely, jobs that are far apart in the occupational hierarchy may still consist of similar bundles of tasks, such as the elementary cleaning occupations (minor group 923) and the housekeeping and related services (minor group 623). Our bottom-up approach further shows that occupations are not reliable categories for conveying information about the skill content of jobs. Our latent skill space—which is agnostic to pre-existing occupational classifications—leads to a different picture of the skill structure of the labor market than is usually assumed. While it is beyond the scope of this article to analyze the actual skill structure of the labor market (see Anderson 2017; Alabdulkareem et al. 2018), our results support previous findings that workers navigate a dense network of skills across jobs, with substantial overlap in their skill profiles (Poletaev and Robinson 2008; DeMaria, Fee, and Wardrip 2020).

Predicting wage with occupations and skill profiles

Our measure of skill profiles captures granular differences in the skill requirements of job postings. This allows us to measure virtually all observable variation in skill content within occupational categories. But is this variation relevant or important? While jobs may require highly specific bundles of skills, to what extent does this translate into unequal socioeconomic outcomes for workers? As a final step, we test the relevance of our novel measure and assess the explanatory power of skill profiles on the wage offered in the job postings. If our measure mainly captures idiosyncratic variations between job postings, skill profiles should explain a negligible fraction of the variation in wages across jobs. Table 7 compares the results of the OLS regressions of the logged minimum hourly wage using the equations (1)-(3).

Model (1) includes the skill profile of job postings, that is, their probabilities on the K-1=18 latent skill categories. This can be contrasted with Model (2), which is four regressions including the occupational coding of the job posting for each level of the SOC classification, from the major (one-digit) to the unit (four-digit) groups. We find that our measure of skill content accounts for slightly less than a third of the variance in wages across job postings ($R^2 = 0.30$). This level of explanatory power is similar to that of major ($R^2 = 0.28$) and sub-major groups ($R^2 = 0.32$) but is less precise than the most detailed levels of the SOC classification. Specifically, the 369 unit groups explain about 40 percent of the variance in the minimum hourly

Table 7: Adjusted coefficient of determination R-squared (R^2) and degrees of freedom (Df) for predicting the logged minimum hourly wage in job postings, comparing models (1)-(3) defined in the Analytical Strategy section.

	Model (1) *Skill profiles*		Models (2) *SOC coding*		Models (3) *(1) and (2) combined*	
	R^2	Df	R^2	Df	R^2	Df
	0.301	18				
SOC level						
Major groups			0.289	8	0.403	26
Sub-major groups			0.320	24	0.414	42
Minor groups			0.356	89	0.432	107
Unit groups			0.396	368	0.460	386

Notes: Degrees of freedom represent the number of effective parameters in the models.

wage. This indicates that improving the explanatory quality of Model (1) comes at the cost of a sharp increase in the number of parameters, expending large degrees of freedom. Overall, these findings show that skill profiles capture relevant differences between job postings and compete well with traditional measures. At the same time, they provide much more direct information about the skills people need to perform on the job.

Our estimates in Model (2) also confirm the strong explanatory power of occupations. They align with studies in other contexts, such as the work of Mouw and Kalleberg (2010), who found that the 469 US Census occupational codes accounted for 43 percent of the wage gap between workers in 2008. However, the extensive literature on the explanatory power of occupations has tended to leave aside the question of *why* occupations are important in the first place (Williams 2013:852). Our detailed measure of skill content gives us a unique opportunity to better understand what occupations actually proxy. In particular, it allows us to assess whether part or most of the effect of occupations on wage inequality is attributable to differences in skill content between occupational groups, as is often assumed (e.g., Goos, Manning, and Salomons 2009; Liu and Grusky 2013). If occupations mainly proxy the type of skills and tasks that are required in jobs, skill profiles should mostly add redundant information to occupations in predicting wages.

Model (3) includes the skill profiles of the job postings as well as their occupational coding for each level of the SOC classification. The skill profiles increase the R^2 from 0.29 to 0.40 when the major groups are used as predictors, representing a 40-percent increase in the explanatory power of the model. At a higher level of granularity in the SOC classification, the skill profiles still increase the R^2 of the minor and unit groups by 25 and 16 percent, respectively. Likelihood ratio tests confirm that the addition of skill profiles significantly improves the goodness of fit for all levels (Table S6). Regarding the regression coefficients, they decrease moderately when compared to Models (1) and (2) but remain in the same order of magnitude (see the full models in Tables S4 and S5).

These findings are robust to the inclusion of a number of job characteristics (type of contract, job hours, industrial code, and geographical county) that may jointly shape skill requirements and the offered wage (see Table S15 and Section 4.2.1), and we obtain nearly identical estimations with an out-of-sample R^2 through cross-validation (see Table S16 and Section 4.2.2). We also assessed the sensitivity of our findings to alternative numbers of topics k, ranging from $k = 5$ to $k = 100$. Overall, the patterns observed for $k = 19$ remain consistent for $k \geq 10$ (see Figures S19 and S20 and Section 4.2.3). The results further show that, with less emphasis on interpretability and scalability, a more granular topic solution (e.g., $k = 50$) would further improve the predictive power of the model across all occupational levels.

Altogether, our wage analysis suggests that *both* occupations and skill profiles are important for understanding wage inequality between job postings. Contrary to what we would expect if occupations were good proxies for the skills and tasks performed at work, our results show that a direct measure of the skill content of jobs significantly improves the prediction of wages compared to including the occupational code alone. Strikingly, combining the skill profiles with the 9 major groups has as much explanatory power as the four-digit unit groups alone, with a much smaller number of parameters ($Df = 26$ vs. 468). This means that our latent skill space captures relevant differences in the skill content of job postings that translate into different socioeconomic outcomes for workers. It is also consistent with the theoretical expectation that wage differentials between workers at least partially arise from the different skills required by their jobs (Haupt and Ebner 2020), although the moderate overall predictive quality of our models ($R^2 = 0.46$ at most) suggests that other important predictors of wages remain unobserved.

Furthermore, occupations and skill profiles appear to be complementary predictors of wages, each providing a substantial share of distinct information. This confirms earlier studies showing, with less precise measures, that job tasks at the individual level are informative about wages even within detailed occupational groups (Autor and Handel 2013; Cassidy 2017; Rohrbach-Schmidt 2019). Our results show that while occupations were designed to group sets of jobs with highly similar tasks and duties (International Labour Office 2012:11), these categories do not correlate well with the very tangible skills that jobs require. As we will elaborate in the discussion, this raises the important question of what occupations capture, if not differences in skills and tasks across jobs.

Limitations

The Lightcast job posting data used in this article provide valuable insights into the skill content of jobs. However, it is important to note that these data were not initially developed for research purposes and have several limitations. First, the data indicate whether a particular skill is required for a job, but do not provide information on the level of skill intensity or ability required. This implies that we consider skills in a binary way, ignoring the fact that jobs with similar skill requirements may necessitate varying levels of proficiency. We measure skills at such a granular level that we do not expect this to introduce substantial bias, but we cannot rule out the possibility that we are underestimating both the heterogeneity in skill content within occupational categories and the explanatory power of skill profiles.

Second, our data capture only the demand side of the labor market. We measure the skill content of jobs as it is described by employers, which may not correspond to what jobs may require in practice and may be even further removed from what workers actually do on the job. More generally, we have largely taken skills for granted, as if they were objectively and unambiguously defined in relation to the task content of jobs. Yet, this neglects the role of institutions in shaping how employers perceive and use skills to construct a given job, and how workers respond to them in return. Our exploratory approach has shed light on the great heterogeneity of skills within occupational categories but falls short in contextualizing how the skill profiles of jobs are created and socially validated. We see this as a very fruitful avenue for further research, as we might expect the skill content of jobs to vary greatly depending on the institutional context in which they are embedded. For example, the presence of unions, occupational licensing, or gender norms may all affect the definition of skills and their strategic use in the labor market (Bol and Weeden 2015; Martin-Caughey 2021).

Third, the wage offered in job postings may differ from the wage paid to incumbents. Our wage equations should therefore be interpreted with caution, as the mechanisms generating wage inequality between jobs may differ from those generating inequality between workers. For a given wage offered, workers have different bargaining power depending on the scarcity and suitability of their skill profiles, among other organizational factors (see, e.g., Sakamoto and Wang 2017). Replicating this study with employee surveys would therefore be worthwhile, provided that they include detailed task data. Moreover, our wage equations are only estimated for a subset of job postings that propose a wage. As discussed in the Data section, these job postings tend to advertise more permanent, full-time contracts, which could affect the relationship between skills, occupations, and wages. While we have no strong intuition about how this might bias the result, more comprehensive data on wages could put our findings on a firmer footing.

Beyond these limitations inherent in job posting data, our findings rely on the assumption that the skills listed in job postings accurately reflect the actual skill profiles of jobs. On the one hand, employers use job postings as a screening device and are therefore expected to include the most relevant skill requirements for the job (Jackson 2007). On the other hand, they may omit skill requirements that they feel are obvious or implied by the job title or qualification requirements. We expect this potential omission bias to be less prevalent in the United Kingdom, where the general education system is weakly tied to the labor market (Bol and Weeden 2015). For example, less than 20 percent of employers indicate a preferred qualification requirement in the Lightcast data. However, for the so-called regulated professions, employers may omit the skill qualifications that are legally required to undertake such professional activities. While it is challenging to measure the extent of this potential bias in a systematic way, descriptive analysis of a sample of regulated professions suggests that it is unlikely to substantially alter our findings (see Section 4 in the online supplement). Yet, further investigation of omission bias in job postings is important for understanding the caveats with which such data should be used for research purposes.

Finally, it is important to note that our skill-profile measure was not intended for direct reuse by other researchers. The purpose of this study was to highlight the potential of a data-driven and granular measure of job skills compared to conventional proxies. However, some of its strengths limit its broader applicability. Namely, while our 19-dimensional skill space yields a conceptually sound and interpretable mapping of job skills, it is inherently shaped by the particular pool of job postings analyzed. As such, it is by no means a definitive or universal partition of the skills landscape. In addition, for the purposes of our study, the skill profiles are defined at the job level. This complicates the export of our measures to other databases and their comparison with more traditional measures of skill or task content. Practically, the high dimensionality of our measure may also require more degrees of freedom than many studies can accommodate. Such limitations are common to other bottom-up approaches to skills (e.g., Anderson 2017; Stephany and Teutloff 2024). Thus, the next step in extending our exploratory approach is to develop more accessible and scalable skill-profile measures for social science research.

Discussion and conclusion

Key substantive findings

Our central finding is that occupations are not bundles of skills: even detailed occupational categories do not consist of homogeneous or clearly distinct skill profiles. Crude conceptualizations and measures of skills miss a key aspect of their structure: most jobs require combinations of specialized knowledge and general skills, and only half of the jobs in our sample rely primarily on a single latent skill category. This heterogeneity is systematically obscured when skill content is reduced to the presence or absence of a small number of predetermined skill types (Liu and Grusky 2013).

While previous studies have emphasized the importance of skill combinations (Poletaev and Robinson 2008; Geel and Backes-Gellner 2011), a key contribution of this article is an empirical measure fully aligned with its theoretical formulation, thereby strengthening the link between theory and empirical evidence (Lundberg, Johnson, and Stewart 2021). Topic modeling is central to this approach, as it avoids predefined categories and captures the contextual dependence and complementarities of skills at the job level. Although we present it here as a proof of concept using job postings, the approach lays the groundwork for constructing lower-dimensional and less computationally intensive measures. In the meantime, researchers should exploit all available skill information rather than relying solely on occupational classifications. For example, surveys such as the British Employer Skills Survey or the German BIBB/BAuA Employment Survey include detailed skill and task dimensions that could be used to construct occupation-independent skill measures.

Our detailed measure of skill profiles also carries explanatory value: it accounts for substantial wage variation across jobs, complementing rather than substituting for the effect of occupational categories. Controlling for skill profiles and one-digit occupational codes predicts wages at least as well as four-digit codes alone. Yet, occupations remain significant predictors of wages even when job-level skill content is comprehensively accounted for. These findings complicate prevailing interpretations of why occupations matter in the labor market and point to a research agenda that clearly distinguishes and theorizes the distinct roles of skills and occupations.

Implications and agenda for future research

Measuring skill content beyond occupations

Our findings question how occupational classifications are measured and how researchers use them (Sakamoto and Wang 2020; Martin-Caughey 2021). While occupations are explicitly defined to classify the type of work that people do, in practice, this classification relies on a highly truncated representation of the skill content of jobs. For example, we noted the coexistence of jobs with different levels of seniority within the same occupational category, leading to significant differences in skill content. This is consistent with Cassidy (2017:406), who observed that the "hierarchical level has a major impact on [the] reported task usages [of workers]." Occupational classifications tend to ignore the organizational setting in which skills are used (Jencks, Perman, and Rainwater 1988:1328; Tomaskovic-Devey and Avent-Holt 2019), neglecting that skills—and their returns—are highly context-dependent (Wilmers 2020; Stephany and Teutloff 2024).

Furthermore, occupations are defined by both the type of work performed by workers and their skill level, but the latter cannot be proxied by skill alone. As Hodge (1981:397) puts it, "Knowing the type of work [people] do and nothing more does not suffice to rank them." In practice, statistical organizations typically proxy skill levels with the amount of formal and informal education required to perform the jobs (Office for National Statistics 2010; International Labour Office 2012). This has important implications for scholars who use occupations as a proxy for skills: occupational classifications often separate jobs with similar skill content, reflecting not only horizontal differences in tasks but also an underlying educational hierarchy.

Thus, occupations are far from neutral analytical categories, despite their widespread use. They remain prominent in the study of income inequalities and inter- and intragenerational mobility (Kalleberg and Mouw 2018; Leicht 2020; Barone, Hertel, and Smallenbroek 2022) and are also widely used in economics, particularly in studies of technological change (Freeman et al. 2020). Our results show that occupations cannot simply serve as proxies for skills. Even when controlling for four-digit occupational codes, adding job-level skill profiles increases the explanatory power of our wage model by 16 percent. This added value is consistent with recent research moving beyond occupations to demonstrate that workers' productivity and comparative advantage depend on the complementarity of their skills (Anderson 2017; Stephany and Teutloff 2024) and the exclusivity of their skill profile (Wilmers 2020).

Beyond wages, our findings are also relevant for research on digitalization and robotization. This literature would benefit from a more holistic understanding of skills, as crude task-based measures may exaggerate the adverse effects of automation on occupations. As Dengler and Matthes (2018) show, even when some tasks are readily automated, most jobs combine tasks that cannot be performed by machines. Future research could build on our conceptual and methodological approach to skill profiles to better capture the hybrid nature of jobs' task compositions.

This is particularly important in light of the recent surge in AI, which occurred after the data for this study were collected. As generative AI increasingly performs tasks once considered non-automatable, a more nuanced understanding of which skill combinations are vulnerable becomes essential—for example, by distinguishing more precisely between different types of interpersonal and intrapersonal skills. Finally, although a temporal analysis was beyond the scope of this study, examining how skill profiles evolve over time would shed light on how recent labor market changes are reshaping the skill structure of jobs, beyond simply identifying which individual skills are rising or declining.

New perspectives for the study of labor market mobility

Our study opens new perspectives for the literature on labor market mobility. Our bottom-up approach allows us to visualize what a purely skill-based labor market would look like—without boundaries imposed by educational requirements and other institutional barriers. An important mechanism in understanding (a lack of) mobility between jobs is found in the (in)transferability of workers' skills (e.g., Poletaev and Robinson 2008; Gathmann and Schönberg 2010; Hanushek et al. 2017). Our mapping of skills on occupations suggests that there is, in fact, substantial overlap in the skill content of occupations. We found that many occupations require similar skills, even though they are far apart in occupational hierarchies. This finding needs to be corroborated with other types of data, such as detailed job task surveys, to rule out possible omission bias in job postings. Nevertheless, it is consistent with the earlier observation of large intra- and intergenerational mobility flows between microclasses, which highlights the permeability of some occupational boundaries even at the detailed occupational level (Lin and Hung 2022:1585, York, Song, and Xie 2025:1013).

In particular, our findings support the idea that upward mobility for workers in low-wage occupations would require only limited reskilling (See also DeMaria et al. (2020)). While this might sound striking, this finding is consistent with the permeability between occupations that workers themselves perceive (Hénaut et al. 2023). Especially at the bottom of the occupational structure, workers are found to "stretch status," that is, to identify with occupations of different status that involve similar tasks (Hénaut et al. 2023).

This suggests that there is much more potential for career mobility than we usually think. However, these skill-based transitions may be hindered in practice (DeMaria et al. 2020). Workers who have the right skill profile for a job may not have the certified credentials to access it (Lancee and Bol 2017). This draws attention to the potential barriers and institutional mechanisms that limit the occupational mobility of workers regardless of, or beyond, the matching of their skill profiles (Weeden and Grusky 2005; Bol et al. 2019).

More research is needed to assess how formal qualifications affect the transferability of skills and occupational mobility (Lancee and Bol 2017; Nedelkoska and Neffke 2019). Existing measures of skill mismatches often focus on general skills, neglecting job-specific knowledge (e.g., Guvenen et al. 2020; Lise and Postel-Vinay 2020). Our skills-based approach allows finer-grained identification of potential labor market transitions and horizontal skill transferability. Future work could, for example, compare occupational mapping based on skill proximity with actual mobility flows of workers between occupations (Cheng and Park 2020; Lin and Hung 2022).

Toward a sociological approach to skills and occupations

Our finding that skills and occupations play distinct roles invites scholars to reconsider their respective contributions to labor market inequality. We confirm the relevance of the skills used at work, also known as "task-based human capital" (Gathmann and Schönberg 2010; Schulz et al. 2023). Previous research has often linked differences in workers' career paths to the different skills they possess (see Hanushek et al. 2017), and primarily to the ones that they acquire through formal training (DiPrete et al. 2017). Our research suggests that another part of the story takes place in the workplace, where workers' skills take shape in a given hierarchical labor process (Avent-Holt et al. 2020). While a few studies have sought to conceptualize the allocation of work tasks into jobs and their labor market returns (Wilmers 2020), a sociological model that disentangles the effects of school-based and task-based human capital remains to be developed (Haupt and Ebner 2020; Schulz et al. 2023).

Our study demonstrates that online job postings are a valuable and underappreciated source for mapping skill structures. Unlike most employee surveys, they capture a broad range of skills beyond general categories, and they represent a wider segment of the labor market than typical online labor market datasets (Anderson 2017; Stephany and Teutloff 2024). Combining our comprehensive measure of job skill content with employee data at the job-title level could further illuminate the relationship between the skills workers possess and the skills they actually use on the job.

While our findings highlight the role of skills in labor market inequality, they also show that occupations matter even when job skill composition is comprehensively accounted for. Occupations influence inequality not merely through the tasks and skills they encompass, which then raises the question posed by Martin-Caughey (2021): "What's in an occupation?"

Our findings support theories that view occupations not as bundles of skills, but as entities embedded in institutions and societies' cultures (Haupt and Ebner 2020:24–47). As Weeden and Grusky (2005:153) argue, homogeneity may arise not per se because workers share the same working conditions, but because "employers (and, to some extent, workers) fashion jobs that correspond with ideal-typical occupational templates." From this perspective, occupational boundaries are produced by institutional mechanisms and cultural perceptions that assign greater value to some jobs over others. While we do not directly test these mechanisms, our study reinforces the idea that skills alone cannot explain wage differences and points to job postings as a rich source for studying occupational barriers, given the additional information they contain beyond skills, such as education and licensing requirements.

In sum, rather than opposing skill- and occupation-based perspectives, future research should explore their interaction, particularly how real or perceived differences in skills can be used to legitimize socio-economic inequalities between occupational groups (Braverman 1998).

Investigating the role of institutional factors

The skill formation system in the United Kingdom is characterized by a relatively open market economy, combined with an educational system that offers very little vocational and occupation-specific training (Hall and Soskice 2001a; Bol and Van De Werfhorst 2013). The generalist educational system of the United Kingdom is not strongly tied to the labor market, and in contrast to many continental European countries (Germany, the Netherlands), many jobs do not have specific educational requirements (Bol and Weeden 2015).

Compared to other countries, the weak linkage between specific educational credentials and specific occupations may exacerbate the within-occupation variation in the skill requirements. While it is likely that our results will be different in countries with stronger linkage between education and the labor market, we argue that our findings should be relatively similar in the United States, which—similar to the United Kingdom—lacks a strong vocational educational system and is characterized by diffuse pathways to employment (DiPrete et al. 2017). An interesting avenue for future research is to explicitly study these contextual factors: to what extent does the content of occupations depend on the institutional setup of a country's skill formation system?

In conclusion, this study opens up promising avenues for further research. It has shed a different light on occupational classifications, exploring new and "more complex ways in which two occupations can be 'close to' or 'distant from' one another" (Cheng and Park 2020:620). We have shown that when it comes to skills, occupational categories tend to harbor a great deal of heterogeneity. Should we then consider occupational analysis as an old VW Beetle from the 1960s and trade it in for a new model (Leicht 2020)? Before we make a hasty decision, let's first open the hood.

Notes

1 In this article, the term "skills" refers to the skills that are required to perform a given job, rather than the skills that workers possess (but may not use). This differs from a common distinction in the literature, which defines skills as characteristics of workers and tasks as characteristics of jobs (see Haupt and Ebner 2020). Our definition of skills should thus be understood as "skills used on the job," which then largely overlaps with the notion of job tasks. It is relevant in the context of our analytical approach, which seeks to analyze skills and their returns from the demand side: we focus on the sets of skills that

employers require for a given job and how they would reward them, regardless of the skills of the workers who will occupy that job.

2 Accordingly, in our data, the vast majority of employers (>80 percent) do not specify a preferred qualification requirement. As a robustness check, we analyzed the extent to which skills implied by job titles or qualifications are omitted in Section 4.1 of the online supplement.

3 At best, including job-level task characteristics alongside the same characteristics measured at the occupational level increases the R^2 from 0.27 to 0.34 in Autor and Handel (2013:85).

4 Lightcast is a merger of Emsi and Burning Glass Technologies. For more information, visit https://lightcast.io/about/data [accessed on 02/03/2026].

5 This taxonomy is updated every 2 weeks and is freely available via an Application Programming Interface (API): https://lightcast.io/open-skills/access [accessed on 02/03/2026].

6 Lightcast uses unique but standardized skill requirements in order to group synonyms. For example, "communication skills" and "ability to communicate" are coded in the same skill requirement.

7 The most represented job titles among the job postings for which Lightcast did not identify any skill requirements are, in descending order of importance: Laborer, Care Assistant, Support Worker, Warehouse Operative, and Cleaner. A visual inspection of the raw text of these job postings suggests that job descriptions are typically very short or unspecific. When they do include some detail, job postings either emphasize what workers are expected to do (e.g., "sort through bins full of metal cables"), or personality traits they should have (e.g., "you'll need to be caring, compassionate and empathetic"), leaving aside identifiable skill requirements.

8 The skill requirements excluded from the random sample tend to be so specific or technical that they only appear in a minority of job postings. Seventy-five percent of them appeared in fewer than eight job postings, and none appeared in more than a hundred job postings in 2019.

9 Using employment figures to correct the non-representativeness of job posting data has its limitations, as the two data sources measure different quantities: while job posting data measure the *flows* of vacancies, employment data inform about their *stock* at a given time. This is particularly problematic when job posting data are used as a proxy for the demand side. However, the purpose of our study is not to identify aggregate skill dynamics. Moreover, as Cammeraat and Squicciarini (2021:17) show, using employment data is the most appropriate option in the absence of commensurable data between job postings and openings.

10 We used ONS employment figures from the APS, which are available via Nomis on https://www.nomisweb.co.uk/datasets/aps168 [accessed on 02/03/2026]. Data were selected for the United Kingdom for the period January 2019 to December 2019.

11 We did not use external databases such as the O*NET because our goal is precisely to avoid a top-down approach to skill classification.

12 We presented the construction of the empirical distributions and the calculation of the distance between them as two separate steps but in practice, the MMD distance performs both steps simultaneously.

13 For a technical discussion of the MMD distance and parameter, see Section 2.2 in the online supplement.

14 We also analyzed the silhouette width of occupational categories to assess whether they are consistent clusters in terms of their skill composition. Yet, it is important to note that the silhouette width is biased toward convex or spherical clusters (Rousseeuw 1987:55), making it a conservative measure for assessing the cluster quality of occupations. In other words, occupations may have a poor silhouette width not because they are poor skill clusters, but because they are non-spherical skill clusters.

15 The analysis using the logged maximum hourly wage yields very similar results and is available in the online supplement, Tables S7 and S8.

16 An interactive visualization of the LDA BTM is available on https://surfdrive.surf.nl/files/index.php/s/T0uNAWrbuJkkIyl [accessed on 02/03/2026]. The skill requirements listed in Table 4 can be retrieved with $\lambda = 0.7$. See Sievert and Shirley (2014) for an introduction to the interface.

17 We gave ChatGPT the following instructions: "I would like you to find good names for the following latent topics, based on the top-100 word probabilities. The names should be precise but not too long (no more than 5 words)."

References

Acemoglu, Daron and David Autor, 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings". Pp. 1043–171 in *Handbook of Labor Economics*, Vol. 4. Elsevier. https://doi.org/10.1016/S0169-7218(11)02410-5.

Alabdulkareem, Ahmad, Morgan R. Frank, Lijun Sun, Bedoor AlShebli, César Hidalgo, and Iyad Rahwan. 2018. "Unpacking the Polarization of Workplace Skills." *Science Advances* 4(7):eaao6030. https://doi.org/10.1126/sciadv.aao6030

Anderson, Katherine A. 2017. "Skill Networks and Measures of Complex Human Capital." *Proceedings of the National Academy of Sciences* 114(48):12720–24. https://doi.org/10.1073/pnas.1706597114

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *The Quarterly Journal of Economics* 118(4):1279–333. https://doi.org/10.1162/003355303322552801

Autor, David and Neil Thompson. 2025. "Expertise." *Journal of the European Economic Association* 23(4):1203–71. https://doi.org/10.1093/jeea/jvaf023

Autor, David H. and Michael J. Handel. 2013. "Putting Tasks to the Test: Human Capital, Job Tasks, and Wages." *Journal of Labor Economics* 31(S1):S59–96. https://doi.org/10.1086/669332

Avent-Holt, Dustin, Lasse F. Henriksen, Anna E. Hägglund, Jiwook Jung, Naomi Kodama, Silvia M. Melzer, Eunmi Mun, Anthony Rainey, and Donald Tomaskovic-Devey. 2020. "Occupations, Workplaces or Jobs? An Exploration of Stratification Contexts using Administrative Data." *Research in Social Stratification and Mobility* 70:100456. https://doi.org/10.1016/j.rssm.2019.100456

Backhaus, Kristin B. 2004. "An Exploration of Corporate Recruitment Descriptions on Monster.com." *The Journal of Business Communication* 41(2):115–36. https://doi.org/10.1177/0021943603259585

Barone, Carlo, Florian R. Hertel, and Oscar Smallenbroek. 2022. "The Rise of Income and the Demise of Class and Social Status? A Systematic Review of Measures of Socio-economic Position in Stratification Research." *Research in Social Stratification and Mobility* 78:100678.https://doi.org/10.1016/j.rssm.2022.100678

Bills, David B., Valentian Di Stasio, and Klarita Gërxhani. 2017. "The Demand Side of Hiring: Employers in the Labor Market." *Annual Review of Sociology* 43(1):291–310. https://doi.org/10.1146/annurev-soc-081715-074255

Blair, Peter Q. and David J. Deming. 2020. "Structural Increases in Demand for Skill after the Great Recession." *AEA Papers and Proceedings* 110:362–65. https://doi.org/10.1257/pandp.20201064

Blau, P. and O. D. Duncan. 1967. *The American Occupational Structure*. New-York: John Wiley & Sons.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3:993–1022.

Blei, David M. 2012. "Probabilistic Topic Models." *Communications of the ACM* 55(4):77–84. https://doi.org/10.1145/2133806.2133826

Böhm, Miachel J., Hans-Martin Von Gaudecker, and Felix Schran. 2024. "Occupation Growth, Skill Prices, and Wage Inequality." *Journal of Labor Economics* 42(1):201–243. https://doi.org/10.1086/722084

Bol, Thijs, Christina Ciocca Eller, Herman G. Van De Werfhorst, and Thomas A. DiPrete. 2019. "School-to-Work Linkages, Educational Mismatches, and Labor Market Outcomes." *American Sociological Review* 84(2):275–307.https://doi.org/10.1177/0003122419836081

Bol, Thijs and Herman G. Van De Werfhorst. 2013. "Educational Systems and the Trade-Off between Labor Market Allocation and Equality of Educational Opportunity." *Comparative Education Review* 57(2):285–308. https://doi.org/10.1086/669122

Bol, Thijs and Kim A. Weeden. 2015. "Occupational Closure and Wage Inequality in Germany and the United Kingdom." *European Sociological Review* 31(3):354–69. https://doi.org/10.1093/esr/jcu095

Braverman, Harry. 1998. *Labor and Monopoly Capital: the Degradation of Work in the Twentieth Century*. 25th anniversary ed. New York: Monthly Review Press.

Buchmann, Marlis, Helen Buchs, and Ann-Sophie Gnehm. 2020. "Occupational Inequality in Wage Returns to Employer Demand for Types of Information and Communications Technology (ICT) Skills: 1991-2017." *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 72(S1):455–82. https://doi.org/10.1007/s11577-020-00672-5

Burning Glass Technologies. 2015. *The Human Factor. The Hard Time Employers Have Finding Soft Skills*. Tech. rep. Boston: Burning Glass Technologies.

Caines, Colin, Florian Hoffmann, and Gueorgui Kambourov. 2017. "Complex-task Biased Technological Change and the Labor Market." *Review of Economic Dynamics* 25, 298-319. https://doi.org/10.1016/j.red.2017.01.008

Cammeraat, Emile and Mariagrazia Squicciarini. 2021. *Burning Glass Technologies' Data Use in Policy-relevant Analysis: An Occupation-level Assessment*. OECD Science, Technology and Industry Working Papers 2021/05. Series: OECD Science, Technology and Industry Working Papers Volume: 2021/05. Organisation for Economic Co-operation and Development. https://doi.org/10.1787/cd75c3e7-en

Caplow, Theodore. 1954. *The Sociology of Work*. Minneapolis: University of Minnesota Press.

Carnevale, Anthony P., Tamara Jayasundera, and Dmitri Repnikov. 2014. *Understanding Online Job Ads Data. A Technical Report.* Tech. rep. Center on Education and the Workforce, Georgetown University.

Cassidy, Hugh. 2017. "Task Variation Within Occupations." *Industrial Relations: A Journal of Economy and Society* 56(3):393–410. https://doi.org/10.1111/irel.12179

Chang, Jonathan, Jordan Boyd-Graber, Sean Gerrish, Chong Wang, and David M. Blei. 2009. "Reading Tea Leaves: How Humans Interpret Topic Models." *Proceedings of the 22nd International Conference on Neural Information Processing Systems*, 288-296.

Cheng, Siwei and Barum Park. 2020. "Flows and Boundaries: A Network Approach to Studying Occupational Mobility in the Labor Market." *American Journal of Sociology* 126(3):577–631. https://doi.org/10.1086/712406

Daly, Moira, Fane Groes, and Mathias F. Jensen. 2025. "Skill Demand Versus Skill Use: Comparing Job Posts with Individual Skill Use on the Job." *Labour Economics* 92:102661. https://doi.org/10.1016/j.labeco.2024.102661

Damelang, Andreas, Andreas Haupt, and Martin Abraham. 2018. "Economic Consequences of Occupational Deregulation: Natural Experiment in the German Crafts." *Acta Sociologica* 61(1):34–49. https://doi.org/10.1177/0001699316688513

DeMaria, Kyle, Kyle Fee, and Keith Wardrip. 2020. *Exploring A Skills-Based Approach To Occupational Mobility*. Community Affairs Discussion Paper 89004. Federal Reserve Bank of Philadelphia.

Deming, David and Lisa B. Kahn. 2018. "Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals." *Journal of Labor Economics* 36(S1):S337–69. https://doi.org/10.1086/694106

Deming, David J. 2017. "The Growing Importance of Social Skills in the Labor Market*." *The Quarterly Journal of Economics* 132(4):1593–640. https://doi.org/10.1093/qje/qjx022

Dengler, Katharina and Britta Matthes. 2018. "The Impacts of Digital Transformation on the Labour Market: Substitution Potentials of Occupations in Germany." *Technological Forecasting and Social Change* 137:304–316. https://doi.org/10.1016/j.techfore.2018.09.024

DiMaggio, Paul, Manish Nag, and David Blei. 2013. "Exploiting Affinities between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Arts Funding." *Poetics* 41(6):570–606. https://doi.org/10.1016/j.poetic.2013.08.004

DiPrete, Thomas A., Christina C. Eller, Thijs Bol, and Herman G. Van De Werfhorst. 2017. "School-to-Work Linkages in the United States, Germany, and France." *American Journal of Sociology* 122(6):1869–938. https://doi.org/10.1086/691327

Djumalieva, Jyldyz and Cath Sleeman. 2018. "An Open and Data-driven Taxonomy of Skills Extracted from Online Job Adverts". Pp. 425–454 in *Developing Skills in a Changing World of Work*, edited by Christa Larsen, Sigrid Rand, Alfons Schmid, and Andrew Dean. Rainer Hampp Verlag. DOI: https://doi.org/10.5771/9783957103154-425

Durkheim, Emile. 1893. *The Division of Labor in Society*. 1933rd ed. New-York: Free Press.

Erikson, Robert and John H. Goldthorpe. 1992. *The Constant Flux: a Study of Class Mobility in Industrial Societies*. Oxford: Oxford University Press.

Fernández-Macías, Enrique and Martina Bisello. 2020. *A Taxonomy of Tasks for Assessing the Impact of New Technologies on Work*. Tech. rep. 2020/04. Seville: European Commission, Joint Research Centre (JRC).

Fernández-Macías, Enrique and John Hurley. 2016. "Routine-biased Technical Change and Job Polarization in Europe." *Socio-Economic Review* 15(3):563–85. https://doi.org/10.1093/ser/mww016

Freeman, Richard B., Ina Ganguli, and Michael J. Handel. 2020. "Within-Occupation Changes Dominate Changes in What Workers Do: A Shift-Share Decomposition, 2005-2015." *AEA Papers and Proceedings* 110:394-399. https://doi.org/10.1257/pandp.20201005

Ganzeboom, Harry B., Paul M. De Graaf, and Donald J. Treiman. 1992. "A Standard International Socio-economic Index of Occupational Status." *Social Science Research* 21(1):1–56. https://doi.org/10.1016/0049-089X(92)90017-B

Gathmann, Christina and Uta Schönberg. 2010. "How General Is Human Capital? A Task-Based Approach." *Journal of Labor Economics* 28(1):1–49. https://doi.org/10.1086/649786

Geel, Regula and Uschi Backes-Gellner. 2011. "Occupational Mobility within and between Skill Clusters: an Empirical Analysis based on the Skill-weights Approach." *Empirical Research in Vocational Education and Training* 3(1):21–38. https://doi.org/10.1007/BF03546496

Goldthorpe, John H. 2002. "Occupational Sociology, Yes: Class Analysis, No: Comment on Grusky and Weeden's Research Agenda." *Acta Sociologica* 45(3):211–217. https://doi.org/10.1177/000169930204500303

Goos, Maarten, Alan Manning, and Anna Salomons. 2009. "Job Polarization in Europe." *American Economic Review* 99(2):58–63. https://doi.org/10.1257/aer.99.2.58

Goos, Maarten, Alan Manning, and Anna Salomons. (2014). "Explaining Job Polarization: Routine-Biased Technological Change and Offshoring." *American Economic Review* 104(8):2509–26. https://doi.org/10.1257/aer.104.8.2509

Greenacre, Michael. 2021. "Compositional Data Analysis." *Annual Review of Statistics and Its Application* 8(1):271–99. https://doi.org/10.1146/annurev-statistics-042720-124436

Gretton, Arthur, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. 2012. "A Kernel Two-sample Test." *The Journal of Machine Learning Research* 13(1):723–73.

Guvenen, Fatih, Burhan Kuruscu, Satoshi Tanaka, and David Wiczer. 2020. "Multidimensional Skill Mismatch." *American Economic Journal: Macroeconomics* 12(1):210–44. https://doi.org/10.1257/mac.20160241

Hall, Peter A. and David W. Soskice. 2001a. "An Introduction to Varieties of Capitalism". Pp. 1–68 in *Varieties of Capitalism*, edited by Hall, P. A. and D. Soskice. 1st ed. Oxford: Oxford University Press. https://doi.org/10.1093/0199247757.003.0001

Hall, Peter A. and David W. Soskice. 2001b. *Varieties of Capitalism: the Institutional Foundations of Comparative Advantage*. Oxford: Oxford University Press.

Hamann, Julian and Stefan Beljean. 2021. "Career Gatekeeping in Cultural Fields." *American Journal of Cultural Sociology* 9(1):43–69. https://doi.org/10.1057/s41290-019-00078-7

Hanushek, Eric A., Guido Schwerdt, Ludger Woessmann, and Lei Zhang. 2017. "General Education, Vocational Education, and Labor-Market Outcomes over the Lifecycle." *Journal of Human Resources* 52(1):48–87.https://doi.org/10.3368/jhr.52.1.0415-7074R

Haslberger, Matthias. 2021. "Routine-biased Technological Change Does not always Lead to Polarisation: Evidence from 10 OECD Countries, 1995-2013." *Research in Social Stratification and Mobility* 74, 100623. https://doi.org/10.1016/j.rssm.2021.100623

Haslberger, Matthias. 2022. "Rethinking the Measurement of Occupational Task Content." *The Economic and Labour Relations Review* 33(1):178–199. https://doi.org/10.1177/10353046211037095

Haupt, Andreas. 2023. "Who Profits from Occupational Licensing?." *American Sociological Review* 88(6):1104–130. https://doi.org/10.1177/00031224231207395 10.1177/00031224231207395.

Haupt, Andreas and Christian Ebner. 2020. "Occupations and Inequality: Theoretical Perspectives and Mechanisms." *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 72(S1):19–40. https://doi.org/10.1007/s11577-020-00685-0

Hauser, Robert M. and John R. Warren. 1997. "Socioeconomic Indexes for Occupations: A Review, Update, and Critique." *Sociological Methodology* 27(1):177–298. https://doi.org/10.1111/1467-9531.271028

Hénaut, Léonie, Jennifer C. Lena, and Fabien Accominotti. 2023. "Polyoccupationalism: Expertise Stretch and Status Stretch in the Postindustrial Era." *American Sociological Review* 88(5):872–900. https://doi.org/10.1177/00031224231190942

Hershbein, Brad and Lisa B. Kahn. 2018. "Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings." *American Economic Review* 108(7):1737–72. https://doi.org/10.1257/aer.20161570

Hodge, Robert W. 1981. "The Measurement of Occupational Status." *Social Science Research* 10(4):396–415. https://doi.org/10.1016/0049-089X(81)90012-0

Hurrell, Scott A., Dora Scholarios, and Paul Thompson. 2013. "More than a 'Humpty Dumpty' Term: Strengthening the Conceptualization of Soft Skills." *Economic and Industrial Democracy* 34(1):161–82. https://doi.org/10.1177/0143831X12444934

International Labour Office, ed. 2012. *International Standard Classification of Occupations 2008 (ISCO-08): Structure, group definitions and correspondence tables*. International Labour Office.

Jackson, Michelle. 2007. "How Far Merit Selection? Social Stratification and the Labour Market." *The British Journal of Sociology* 58(3):367–90. https://doi.org/10.1111/j.1468-4446.2007.00156.x

Janietz, Christoph and Thijs Bol. 2020. "Occupations, Organizations, and the Structure of Wage Inequality in the Netherlands." *Research in Social Stratification and Mobility* 70:100468. https://doi.org/10.1016/j.rssm.2019.100468

Jencks, Christopher, Lauri Perman, and Lee Rainwater. 1988. "What Is a Good Job? A New Measure of Labor-Market Success." *American Journal of Sociology* 93(6).

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. 1993. "Wage Inequality and the Rise in Returns to Skill." *Journal of Political Economy* 101(3):410–42. https://doi.org/10.1086/261881

Kalleberg, Arne L. 2009. "Precarious Work, Insecure Workers: Employment Relations in Transition." *American Sociological Review* 74(1):1–22. https://doi.org/10.1177/000312240907400101

Kalleberg, Arne L. and Ted Mouw. 2018. "Occupations, Organizations, and Intragenerational Career Mobility." *Annual Review of Sociology* 44(1):283–303. https://doi.org/10.1146/annurev-soc-073117-041249

Kalleberg, Arne L., Ted Mouw, and Michael A. Schultz. 2025. "Career Types and Labor Market Structure: Intragenerational Mobility in the United States." *Social Science Research* 128:103153. https://doi.org/10.1016/j.ssresearch.2025.103153

Kim, ChangHwan, Jibum Kim, and Mihee Ban. 2020. "Do You Know what You Do for a Living? Occupational Coding Mismatches between Coders in the Korean General Social Survey." *Research in Social Stratification and Mobility* 70:100467. https://doi.org/10.1016/j.rssm.2019.100467

Kozlowski, Austin C., Matt Taddy, and James A. Evans. 2019. "The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings." *American Sociological Review* 84(5):905–49. https://doi.org/10.1177/0003122419877135

Lancee, Bram and Thijs Bol. 2017. "The Transferability of Skills and Degrees: Why the Place of Education Affects Immigrant Earnings." *Social Forces* 96(2):691–716. https://doi.org/10.1093/sf/sox058

Lazear, Edward P. 2009. "Firm-Specific Human Capital: A Skill-Weights Approach." *Journal of Political Economy* 117(5):914–40. https://doi.org/10.1086/648671

Le Vrang, Martin, Agis Papantoniou, Erika Pauwels, Pieter Fannes, Dominique Vandensteen, and Johan De Smedt. 2014. "ESCO: Boosting Job Matching in Europe with Semantic Interoperability." *Computer* 47(10):57–64. https://doi.org/10.1109/MC.2014.283

Leicht, Kevin T. 2020. "Occupations and Inequalities in the 21st Century: What's in your Wallet?." *Research in Social Stratification and Mobility* 70:100550. https://doi.org/10.1016/j.rssm.2020.100550

Lin, Ken-Hou and Koit Hung. 2022. "The Network Structure of Occupations: Fragmentation, Differentiation, and Contagion." *American Journal of Sociology* 127(5):1551–601. https://doi.org/10.1086/719407

Lise, Jeremy and Fabien Postel-Vinay. 2020. "Multidimensional Skills, Sorting, and Human Capital Accumulation." *American Economic Review* 110(8):2328–76. https://doi.org/10.1257/aer.20162002

Liu, Yujia and David B. Grusky. 2013. "The Payoff to Skill in the Third Industrial Revolution." *American Journal of Sociology* 118(5):1330–74. https://doi.org/10.1086/669498

Liu, Zhaolu, Jonathan M. Clarke, Bertha Rohenkohl, and Mauricio Barahona. 2025. "Patterns of Co-Occurrent Skills in UK Job Adverts." *PLOS Complex Systems* 2(2). Petersen, A. Publisher: Public Library of Science (PLoS), e0000028. https://doi.org/10.1371/journal.pcsy.0000028

Lundberg, Ian, Rebecca Johnson, and Brandon M. Stewart. 2021. "What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory." *American Sociological Review* 86(3):532–65. https://doi.org/10.1177/00031224211004187

Macanovic, Ana 2022. "Text Mining for Social Science - The State and the Future of Computational Text Analysis in Sociology." *Social Science Research* 108:102784. https://doi.org/10.1016/j.ssresearch.2022.102784

Marinescu, Ioana and Ronald Wolthoff. 2020. "Opening the Black Box of the Matching Function: The Power of Words." *Journal of Labor Economics* 38(2):535–68. https://doi.org/10.1086/705903

Martin-Caughey, Ananda. 2021. "What's in an Occupation? Investigating Within-Occupation Variation and Gender Segregation Using Job Titles and Task Descriptions." *American Sociological Review* 86(5):960–99. https://doi.org/10.1177/00031224211042053

Modestino, A. S., D. Shoag, and J. Ballance. 2020. "Upskilling: Do Employers Demand Greater Skill When Workers Are Plentiful?." *The Review of Economics and Statistics* 102(4):793–805. https://doi.org/10.1162/rest_a_00835

Mouw, Ted and Arne L. Kalleberg. 2010. "Occupations and the Structure of Wage Inequality in the United States, 1980s to 2000s." *American Sociological Review* 75(3):402–431. https://doi.org/10.1177/0003122410363564

Mouw, Ted, Aren L. Kalleberg, and Michael A. Schultz. 2024. ""Stepping-Stone" versus "Dead-End" Jobs: Occupational Structure, Work Experience, and Mobility Out of Low-Wage Jobs." *American Sociological Review* 89(2):298–345. https://doi.org/10.1177/00031224241232957

Nedelkoska, Ljubica and Frank Neffke. 2019. "Skill Mismatch and Skill Transferability: Review of Concepts and Measurements." *Papers in Evolutionary Economic Geography (PEEG)* 1921.

Nelson, Laura K. 2021. "Cycles of Conflict, a Century of Continuity: The Impact of Persistent Place-Based Political Logics on Social Movement Strategy." *American Journal of Sociology* 127(1):1–59. https://doi.org/10.1086/714915

OECD. 2019. *OECD Skills Strategy 2019: Skills to Shape a Better Future*. OECD Publishing. https://doi.org/10.1787/9789264313835-en

OECD. 2024. *Insights from Skills Strategies in the European Union: Lessons Learnt for Developing and Implementing Effective Skills Policies*. OECD Skills Studies. OECD Publishing. https://doi.org/10.1787/0bf9e78e-en

Office for National Statistics. 2010. *Standard Occupational Classification 2010*. OCLC: 691102853. Basingstoke, Hampshire: Palgrave Macmillan.

Parkin, Frank. 1971. *Class Inequality and Political Order: Social Stratification in Capitalist and Communist Societies*. London: MacGibbon & Kee.

Poletaev, Maxim and Chris Robinson. 2008. "Human Capital Specificity: Evidence from the Dictionary of Occupational Titles and Displaced Worker Surveys, 1984-2000." *Journal of Labor Economics* 26(3):387–420. https://doi.org/10.1086/588180

Röder, Michael, Andreas Both, and Alexander Hinneburg. 2015. "Exploring the Space of Topic Coherence Measures." *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, 399-408. https://doi.org/10.1145/2684822.2685324

Rohrbach-Schmidt, Daniela. 2019. "Putting Tasks to the Test: The Case of Germany." *Social Inclusion* 7(3):122–35. https://doi.org/10.17645/si.v7i3.2025

Rousseeuw, Peter J. 1987. "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis." *Journal of Computational and Applied Mathematics* 20:53-65. https://doi.org/10.1016/0377-0427(87)90125-7

Sacchi, Stefan, Irene Kriesi, and Marlis Buchmann. 2016. "Occupational Mobility Chains and the Role of Job Opportunities for Upward, Lateral and Downward Mobility in Switzerland." *Research in Social Stratification and Mobility* 44:10-21. https://doi.org/10.1016/j.rssm.2015.12.001

Sakamoto, Arthur and Sharron X. Wang. 2017. "Occupational and Organizational Effects on Wages among College-educated Workers in 2003 and 2010." *Social Currents* 4(2):175–95. https://doi.org/10.1177/2329496516651643

Sakamoto, Arthur and Sharron X. Wang. 2020. "The Declining Significance of Occupation in Research on Intergenerational Mobility." *Research in Social Stratification and Mobility* 70, 100521. https://doi.org/10.1016/j.rssm.2020.100521

Salvisberg, Alexander and Stefan Sacchi. 2014. "Labour Market Prospects of Swiss Career Entrants after Completion of Vocational Education and Training." *European Societies* 16(2):255–74. https://doi.org/10.1080/14616696.2013.821623

Schrab, Antonin, Ilmun Kim, Mélisande Albert, Béatrice. Laurent, Benjamin Guedj, and Arthur Gretton. 2023. "MMD Aggregated Two-Sample Test." *Journal of Machine Learning Research* 24. Publisher: arXiv Version Number: 4:1-81. https://doi.org/10.48550/ARXIV.2110.15073

Schulz, Wiebke, Heike Solga, and Reinhard Pollak. 2023. "Vocational Education, Tertiary Education, and Skill Use across Career Stages." *European Sociological Review*, jcac074. https://doi.org/10.1093/esr/jcac074

Sievert, Carson and Kenneth Shirley. 2014. "LDAvis: A Method for Visualizing and Interpreting Topics." *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, 63-70. https://doi.org/10.3115/v1/W14-3110

Sostero, Matteo and Enrique Fernández-Macías. 2021. *The Professional Lens: What Online Job Advertisements Can Say about Occupational Task Profiles*. Tech. rep. JRC125917. Seville: European Commission.

Stephany, Fabian and Ole Teutloff. 2024. "What is the Price of a Skill? The Value of Complementarity." *Research Policy* 53(1):104898. https://doi.org/10.1016/j.respol.2023.104898

Tippins, Nancy T. and Margaret L. Hilton. 2010. *A Database for a Changing Economy: Review of the Occupational Information Network (O*NET)*. OCLC: 650842713. Washington, D.C.: National Academies Press.

Tomaskovic-Devey, Donald and Dustin Avent-Holt. 2019. *Relational Inequalities: An Organizational Approach*. 1st ed. Oxford University Press. DOI: https://doi.org/10.1093/oso/9780190624422.001.0001

Treiman, Donald J. 1977. *Occupational Prestige in Comparative Perspective*. Elsevier. https://doi.org/10.1016/C2013-0-11617-0

Weber, Max. 1979. *Economy and Society: an Outline of Interpretive Sociology*. Berkeley London: University of California Press.

Weeden, Kim A. 2002. "Why Do Some Occupations Pay More than Others? Social Closure and Earnings Inequality in the United States." *American Journal of Sociology* 108(1):55–101. https://doi.org/10.1086/344121

Weeden, Kim A. and David B. Grusky. 2005. "The Case for a New Class Map." *American Journal of Sociology* 111(1):141–212. https://doi.org/10.1086/428815

Weeden, Kim A., Young-Mi Kim, Matthew Di Carlo, and David B. Grusky. 2007. "Social Class and Earnings Inequality." *American Behavioral Scientist* 50(5):702–36. https://doi.org/10.1177/0002764206295015

Williams, Mark. 2013. "Occupations and British Wage Inequality, 1970s-2000s." *European Sociological Review* 29(4):841–57. https://doi.org/10.1093/esr/jcs063

Wilmers, Nathan. 2020. "Job Turf or Variety: Task Structure as a Source of Organizational Inequality." *Administrative Science Quarterly* 65(4):1018–57. https://doi.org/10.1177/0001839220909101

Yamaguchi, Shintaro. 2012. "Tasks and Heterogeneous Human Capital." *Journal of Labor Economics* 30(1):1-53. https://doi.org/10.1086/662066

Yan, Xiaohui, Jiafeng Guo, Yanyan Lan, and Xueqi Cheng. 2013. "A Biterm Topic Model for Short Texts." *Proceedings of the 22nd international conference on World Wide Web*, 1445-1456. https://doi.org/10.1145/2488388.2488514

York, Hunter, Xi Song, and Yy Xie. 2025. "Gradationalism Revisited: Intergenerational Occupational Mobility Along Axes of Occupational Characteristics." *American Journal of Sociology* 130(4):976–1027. https://doi.org/10.1086/733122

Acknowledgments: We are grateful to Luisa Burchartz, Viktor Decker, Thomas A. DiPrete, Fenella Fleischmann, Andreas Haupt, and Wouter Schakel for their helpful feedback on earlier drafts of this manuscript. This research was presented at the 2024 ISA RC28 Spring Meeting, the 2025 TASKS VII Conference, and workshops of the Institutions, Inequalities and Life Courses (IIL) research group at the University of Amsterdam, the Sciences Po Center for Research on Social Inequalities (CRIS), the Center for Research in Economics and Statistics (CREST), and the Federal Institute for Vocational Education and Training (BIBB); we thank participants for their constructive discussions. Marie Labussière gratefully acknowledges Pierre Alquier and Matteo Amestoy for their technical advice.

Funding: This work was supported by the ERC starting grant from School to Career: Towards a Career Perspective on the Labor Market Returns to Education (CAREER) (ID: 950189).

Marie Labussière: Sciences Po, Centre for Research on Social Inequalities (CRIS), E-mail: marie.labussiere@sciencespo.fr.
Thijs Bol: University of Amsterdam, Department of Sociology, E-mail: t.bol@uva.nl.