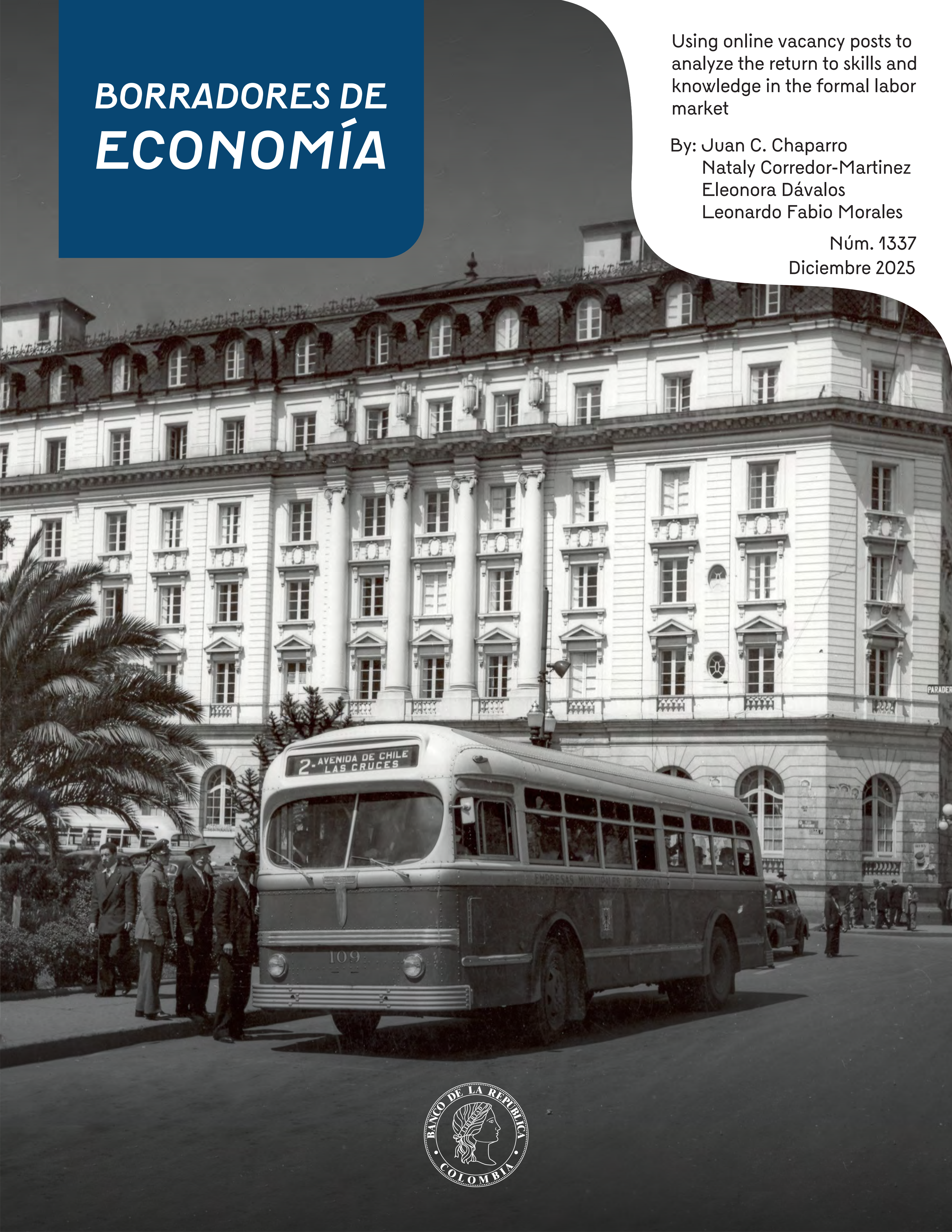
BORRADORES DE ECONOMÍA

Using online vacancy posts to analyze the return to skills and knowledge in the formal labor market

By: Juan C. Chaparro
Nataly Corredor-Martinez
Eleonora Dávalos
Leonardo Fabio Morales

Núm. 1337
Diciembre 2025



BANCO DE LA REPÚBLICA
COLOMBIA

Using online vacancy posts to analyze the return to skills and knowledge in the formal labor market [1]

Juan C. Chaparro, Nataly Corredor-Martinez, Eleonora Dávalos & Leonardo Fabio Morales

Abstract

The advent of digital technology and widespread internet access has transformed how firms advertise job vacancies and how job seekers conduct their searches. This paper presents a novel methodology for identifying job-related skills and knowledge from the textual descriptions of online job postings, using data from Colombia's largest public online job board and the O*NET database. By analyzing posted wages and applying a data-driven approach to identify skill requirements, the study estimates the wage returns associated with specific skills and knowledges in the formal Colombian labor market. The findings indicate that, within a given occupation, the prevalence of certain basic skill categories such as mathematics, writing, time management, and instructing is associated with higher wage returns. Notably, most of the skill categories with significant positive returns are not occupation-specific; rather, they represent transferable capabilities that enhance performance across a wide range of tasks and occupations.

Key words: Online job portals, Vacancies, Occupational skill requirements, Wage disparities, Labor markets in developing countries.

JEL Codes: J23, J24, J31, J63, O15

[1] The series Borradores de Economía is published by the Banco de la República; the documents are provisional, and their authors are fully responsible for the opinions expressed in them, as well as for remaining errors. This version: July 31st, 2025. Juan Camilo Chaparro: Universidad EAFIT jcchaparrc@eafit.edu.co. Nataly Corredor-Martinez: Universidad Nacional de Colombia lncorredorma@unal.edu.co. Eleonora Dávalos: Universidad EAFIT edavalosa@eafit.edu.co. Leonardo Fabio Morales: Banco de la República lmoralzu@banrep.gov.co.

Uso de vacantes publicadas en línea para analizar el retorno de habilidades y conocimientos en el mercado laboral formal[2]

Juan C. Chaparro, Nataly Corredor-Martinez, Eleonora Dávalos & Leonardo Fabio Morales

Resumen

El auge de la tecnología digital y el acceso generalizado a internet ha transformado la forma en que las empresas anuncian vacantes laborales y cómo los buscadores de empleo realizan sus búsquedas. Este artículo presenta una metodología para identificar habilidades y conocimientos relacionados con el trabajo a partir de las descripciones textuales de ofertas de empleo en línea, utilizando datos del mayor portal público de empleo en línea de Colombia y la base de datos O*NET. Al analizar los salarios publicados y aplicar un enfoque basado en datos para identificar los requisitos de habilidades, el estudio estima los retornos salariales asociados con habilidades y conocimientos específicos en el mercado laboral formal colombiano. Los hallazgos indican que, dentro de una ocupación determinada, la prevalencia de ciertas categorías de habilidades básicas como matemáticas, redacción, gestión del tiempo e instrucción está asociada con mayores retornos salariales. Cabe destacar que la mayoría de las categorías de habilidades con retornos positivos significativos no son específicas de una ocupación; más bien, representan capacidades transferibles que mejoran el desempeño en una amplia gama de tareas y ocupaciones.

Palabras clave: Portales de empleo en línea, Vacantes, Requisitos de habilidades ocupacionales, Disparidades salariales, Mercados laborales en países en desarrollo.

JEL Codes: J23, J24, J31, J63, O15

[2] La serie Borradores de Economía es publicada por el Banco de la República; este documento es provisional y las opiniones y errores expresados aquí son exclusivos de los autores. Esta versión: 31 de julio de 2025. Juan Camilo Chaparro: Universidad EAFIT jcchaparrc@eafit.edu.co. Nataly Corredor-Martinez: Universidad Nacional de Colombia lncorredorma@unal.edu.co. Eleonora Dávalos: Universidad EAFIT edavalosa@eafit.edu.co. Leonardo Fabio Morales: Banco de la República lmoralzu@banrep.gov.co.

1. Introduction

Technological advances and the widespread availability of the internet have brought profound changes in how firms advertise open labor positions and how workers search for jobs. The use of online job portals has significantly increased in recent decades. A study by Kuhn and Mansour (2014) using US data from 2010 reported that as many as 76 percent of the unemployed population searched for jobs exclusively through online platforms. The information provided in job postings holds great value as it encapsulates the main specific requirements for a job position. Consequently, job vacancy descriptions serve as an alternative source for identifying the necessary skills and knowledge associated with a given job. This information offers a distinct advantage over traditional skill measurements based on occupations (Autor et al., 2003; Acemoglu & Autor, 2011), as it allows for a more detailed analysis of skill variations across different job roles within occupations.

In recent years, online job board information has emerged as a valuable resource for labor market research. For instance, researchers have used these data to examine the effects of unemployment benefits, as demonstrated by Marinescu (2019). Other studies have focused on investigating the impact of public intermediation through online job boards, as conducted by Morales et al. (2024), and exploring the efficiency of online job search in relation to unemployment durations, as explored by Kuhn and Skuterud (2004) and Kuhn and Mansour (2014). Additionally, previous literature has examined the role of posted wages and vacancy titles in the recruitment process, as evidenced by Marinescu and Wolthoff (2020).

In this study, we present a methodology for identifying skill and knowledge requirements from the textual descriptions of online job vacancies. To accomplish this, we use data from the main public online job board in Colombia—*Servicio Público de Empleo* or SPE. Furthermore, we incorporate the skill and knowledge categorizations provided by the Occupational Information Network (O*NET), the primary source of occupational information in the United States (Peterson et al., 2001; Acemoglu & Autor, 2011). By identifying the demand for specific skills and knowledge, along with posted wages, we estimate the wage returns associated with skill and knowledge acquisition in the Colombian formal labor market.

The existing literature on returns to skills based on job board information is limited. However, recent studies have employed vacancy descriptions to identify the demand for and returns to skills. For instance, Bennett et al. (2022) used data from online job boards in Uruguay to examine the dynamics of skill demand. The study argues that one of the advantages of this approach is its reliance on readily available data in numerous countries worldwide, enabling detailed analysis of skill dynamics over time.

Deming and Kahn (2018) explored the variation in skill demands for professionals across firms, focusing on social and cognitive skills. They used a database from the largest job board in the United States. Although wage information for each vacancy was not available, the study relied on aggregated wages at the occupation level (6-digit) for their analysis. The findings of Deming and Kahn (2018) revealed that cognitive and social skills possess significant predictive power in explaining wage disparities across occupational labor markets in the United States. Hansen et al. (2021) identified skill requirements in top managerial occupations using job descriptions from online boards in the US and the O*NET database. One of the most important findings of that study is the greater emphasis on social skills requirements in these specific positions.

In a related study, Ziegler (2022) examined returns on a more comprehensive set of skills. In line with previous research, the paper argues that job vacancy data allows for greater granularity in characterizing skills demand within occupations compared to traditional occupation-based classifications. Importantly, this study incorporated wage information at the vacancy level, making the vacancy the unit of analysis. Ziegler (2022) provides robust evidence that the observed results are unlikely to be explained by measurement error. In another related study, Deming and Noray (2020) used online vacancy data for college graduates in STEM professions to investigate the impact of changing job skills on wage dynamics.

Overall, these studies demonstrate the usefulness of vacancy information in explaining wage variations, even after controlling for standard covariates such as education, experience, and occupation fixed effects. Furthermore, a subset of the literature exploiting postings from online job boards has focused on studying changes in skill demands prompted by recessions. Some papers within this line of research argue that crises may be perceived by firms as opportunities to revise their skill requirements (Hershbein & Kahn, 2018; Modestino et al., 2020).

Our paper uses online vacancy postings from Colombia's public online job board, managed by the Colombian Public Employment Services (SPE for its acronym in Spanish), to identify skills by extracting relevant keywords from the vacancy descriptions. More specifically, we use data from the publicly available job board managed by SPE, also known as SISE. Although our study is based on data from a single job board, we contribute to the existing literature in several ways. Unlike previous approaches, we have access to posted wage information within each vacancy. Therefore, our methodological design avoids the limitation of associating wages from third-party external sources with the content of vacancy descriptions. In our skill identification process, we adopt a data-driven approach by using the keywords and expressions associated with each skill in the O*NET database. After applying standard text processing techniques, we search for these keywords within the vacancy description text. By incorporating this methodology, we argue for a more objective and systematic skill identification process compared to using ad hoc keyword dictionaries.

Additionally, we estimate returns to knowledge categories in a similar manner, using the categorizations and keywords provided by O*NET. Consistent with existing literature, which suggests that skills contribute to wage differentials even after controlling for education levels, experience, and occupation fixed effects, our findings confirm the same conclusion, not only for skills but for different types of knowledge as well. Lastly, to the best of our knowledge, this study represents the first investigation into the returns to skills and knowledge in a developing country context.

Our findings show that, for a given occupation in a firm, the aggregate skills that contribute the most to explaining wage variation are systems skills. Additionally, other skills such as content and technical skills have significant and sizeable returns. These returns are driven by the demand for more specific skills in vacancy posts, including mathematics, writing, instructing, technology design, and operations monitoring. Regarding knowledge, for a given occupation in a firm, aggregated categories such as engineering, humanities, communications, and law have the highest returns. These returns are driven by the dominance of more specific types of knowledge, such as English, engineering and technology, economics, history, management, public safety, and certain science categories.[3]

The most outstanding finding in our study is that, for a given occupation, the dominance of some basic skill categories has important returns in the labor market. This is the case, for instance, with mathematics, writing, time management, and instructing skills. A similar situation occurs with knowledge in English, history, economics, and technology. Posted vacancies requiring these mentioned skills and knowledge offer a wage premium within the same occupation. Therefore, for any given job, it is preferable to have a candidate with a wide range of skills and knowledge beyond the basic requirements inherent to the occupation. Most of the categories with significant and positive wage returns are not specific to a given occupation; instead, they are tools that can be beneficial in the performance of a variety of tasks.

The remainder of the paper is organized as follows. Section 2 presents a simple theoretical framework of the demand for skills in the labor market. Section 3 provides an overview of the data sources utilized in this study. In Section 4, an algorithm is described that enables the identification of skills and knowledge from the text descriptions of online-posted vacancies. Section 5 outlines the empirical estimation approach employed to identify the returns to skills in the labor market. The main results of the returns to skills and knowledge are presented in Section 6. Section 7 examines the heterogeneity of these results across educational attainment levels and different levels of experience. Section 8 presents a series of robustness checks conducted to validate the robustness of our findings. Lastly, in

[3] Recent literature on human capital reassesses the conception that workers accumulate a homogeneous and rigid amount of human capital, which strictly determines the kind of job they can have. Alternative views consider that labor markets match workers' skills to potential tasks; therefore, the requirement from vacancies is a set of skills that would allow workers to develop specific tasks (Lise and Postel-Vinay, 2020). Furthermore, the set of skills that a worker has might be multidimensional and possibly independent across different occupational categories. Workers have different skills, and they use them proportionally depending on the tasks that the job requires (Sanders and Taber, 2012). Our findings support these alternative views of human capital accumulation.

Section 9, we provide concluding remarks and discuss policy implications derived from our research.

2. Theoretical Framework

Following Acemoglu & Autor (2011), we assume that a firm demands a total amount of labor L, and workers are distributed among those who possess different types of skills.[4] Therefore, $L = S_1 + S_2 + \cdots + S_N$, where S_i is the demand for workers with skill i. Skills are considered imperfect substitutes in the production process, and firms combine workers with different skills to achieve their production goals. The production function of the representative firm maps the combination of skill demands to the firm's outcomes and can be represented by a Cobb-Douglas function as follows:

$$Y = (A_1 S_1)^{\alpha_1} \cdot (A_2 S_2)^{\alpha_2} \cdot \ldots \cdot (A_k S_k)^{\alpha_k} \quad (2.1)$$

In the previous equation, technology is a labor-augmenting factor; therefore, changes in parameters A_j increase labor productivity of the j^{th} ability. Labor markets are competitive, and wages are determined exogenously in each skill segment of the market. If the firm solves a standard optimization problem of maximizing profits, the problem and the first-order conditions for skill S_i can be represented as:

$$\max p(A_1 S_1)^{\alpha_1} \cdot (A_2 S_2)^{\alpha_2} \cdot \ldots \cdot (A_k S_k)^{\alpha_k} - w_1 S_1 - \cdots - w_k S_k$$

$$w_i = p\alpha_i A_i^{\alpha_i} (S_i)^{\alpha_i - 1} \cdot \prod_{j \neq i} \left(A_j S_j\right)^{\alpha_j} \quad (2.2)$$

Equation (2.2) describes the standard optimization condition in which, for a profit-maximizing firm, the value of the marginal product of skill labor j should be equal to the wage for that specific type of skill. This equation implies that the wage for each skill should be a function of the parameters A_i, α_i, and the demand for all skills S_i. Building on this theoretical framework, in the empirical section, we estimate proxies of equation (2.2) where posted wages are a function of observed firm skill demands, and technological factors are incorporated by including firm and occupation fixed effects.

3. Data

3.1 Colombian employment services

This paper uses information on vacancies posted on the Colombian Public Employment Service (SPE) website. The SPE was established by Law 1636 of 2013 to create

[4] As explained in Acemoglu & Author (2011), the canonical model explains the changes in wage structure and skill differentials, based on the ideas proposed in Tinbergen (1974, 1975), Welch (1973), Katz and Murphy (1992), and Card and Lemieux (2001), among others.

a tool for active search labor market policies. The SPE's primary function is to make it easier for workers to find suitable jobs and for employers to hire adequate personnel. Under this initiative, Colombia established the obligation to report job vacancies to the Public Employment Service. Law 1636 requires companies to register and report all job vacancies through a network, including public and private service providers. The network comprises public and private institutions, such as online job portals, university job boards, and social services providers for workers (Morales et al., 2021). Job vacancies reported to the SPE cover various industries and job types, from entry-level positions to high-skilled professional roles, providing a comprehensive view of the Colombian labor market.

The SPE centralizes job vacancies throughout an automated online system, which provides an interface that facilitates interaction with various web pages from network members, providing links that direct job seekers to online platforms where job offers have been posted. Alternatively, the SPE has its own job board, known as the Information System of the Public Employment Service (SISE for its acronym in Spanish). The SISE facilitates the online connection between job seekers and employers in the labor market. Employers interested in filling specific job vacancies register in the SISE information system, where they can post all the information related to each vacancy. Likewise, workers searching for job opportunities must register in the system to apply for vacancies according to their profile. This job matching process ensures that job seekers are directed at the most suitable vacancies, and employers can find the most qualified candidates for their job openings.

Once a vacancy of interest has been selected, the searcher can access its detailed description, which, among other information, includes the following items: the vacancy registered name, a detailed description of the vacancy including the occupation, the employer's name, the posted salary in terms of minimum wages, and the minimum educational attainment and experience required. We provide an example of the information provided by the SPE for a vacancy in SISE, both in Spanish and English, in Appendix A. For this paper, we use the SISE information provided by the SPE on their public website: https://www.buscadordeempleo.gov.co.

Between 2013 and 2023, we collected a total of 184,000 job vacancies listed on SISE for 23 Colombian cities, including Medellín, Barranquilla, Bogotá, Cartagena, Tunja, Manizales, Florencia, Popayán, Valledupar, Montería, Quibdó, Neiva, Riohacha, Santa Marta, Villavicencio, Pasto, Cúcuta, Armenia, Pereira, Bucaramanga, Sincelejo, Ibagué, and Cali. Over this period, job postings surged significantly, escalating from 1,088 in 2014 to a staggering 64,138 in 2022. Bogotá, Colombia's capital, consistently leads annual job postings, highlighting its importance in national employment.

3.2 O*NET skills and knowledge classification

The O*NET program, conducted by the United States Department of Labor, produces an annual database that offers comprehensive information on the main characteristics of all occupations covered by the Standard Occupational Classification System (SOC). The O*NET Content Model is based on previous research from industrial and organizational psychology on job analysis (Peterson et al., 2001; Schmitt, Highhouse & Weiner, 2013).

Table 1. The O*Net Content Model

O*NET Data Collection Program Questionnaire	Number of Descriptors	Number of Scales per Descriptor	Total Number of Scales	Data Sources
Skills	35	2	70	Analysts
Knowledge	33	2	66	Job incumbents
Work Styles	16	1	16	Job incumbents
Education and Training	5	1	5	Job incumbents
Generalized Work Activities	41	2	82	Job incumbents
Work Context	57	1	57	Job incumbents
Abilities	52	2	104	Analysts
Tasks	Varies	2	Varies	Job incumbents
Total (not including tasks)	239	NA	400	NA

Notes: Based on Table 4-3, National Research Council (2010, p. 74).

The Content Model serves as the foundation for organizing the occupational information within the database. It presents a framework that encompasses two perspectives: job-oriented descriptors and worker-oriented descriptors. The job-oriented descriptors focus on occupational requirements, characteristics of the workforce, and occupation-specific information. Conversely, the worker-oriented descriptors concentrate on worker characteristics and experience requirements. These descriptors are further categorized into domains, facilitating the grouping of key attributes of workers and occupations into specific information areas. Table 1 presents a list of the main domains in the O*NET Content Model and the number of descriptors of each domain within the publicly available database.

The classification of skills and knowledge in our study aligns with the worker-oriented descriptors found in O*NET. There are in total 35 skills and 33 knowledge descriptors in the O*NET content model (first two rows in Table 1). Each descriptor has two scales of measurement: an importance score and a level rating.[5] The 35 specific skills in O*NET are classified into seven categories, and the 33 specific knowledges are grouped into ten main knowledge categories.[6] For a detailed breakdown of the distribution of skills within these groups, please refer to Appendix B (Table B1); likewise, a detailed breakdown of the distribution of knowledge groups is provided in Appendix B (Table B2).

4. Identifying demand for skills and knowledge in the formal labor market

4.1 Text preprocessing

To obtain useful information that facilitates the classification process, we filter and process texts, the vacancy description, and the O*NET keywords and expressions by following 4 steps:

1) Translation and synonyms: Given that the online vacancy descriptions were written in Spanish, we translated the O*NET English expressions and keywords into Spanish to create the skill and knowledge keywords. Recognizing that each category can be expressed in multiple ways, we also searched for synonyms corresponding to the 35 specific skills and 33 disaggregated knowledge categories (Appendix B). By incorporating these synonyms, we expanded the range of keywords associated with each category. This process enhanced the likelihood of identifying the relevant keywords and expressions within the vacancy description text.[7]

2) Text normalization: This stage involved several transformations to standardize the text. First, we converted all texts from upper to lower case to ensure consistency. Second, we removed accents from characters (e.g., transforming "á" to "a"). Then, we employed a character encoding process to address recurring text errors. Finally, we eliminated the text's line breaks, bullets, and special characters.

3) Elimination of stop words: we removed all stops words to enhance the text quality and relevance. These stop words typically consist of prepositions and conjunctions in the Spanish language that may introduce noise into the text. We used the stop-word corpus from the Python NLTK library, specifically designed for the Spanish language. In addition, we

[5] O*Net data are collected using "behavioral anchored rating scales" (Peterson et al., 2001, p. 474). For a brief explanation of the difference between an importance score and a label rating in O*NET, see Chaparro (2016, pp. 13 – 14).

[6] The seven skills categories in O*Net are the following: (1) content skills, (2) process skills, (3) social skills, (4) complex-problem solving skills, (5) technical skills, (6) system skills and (7) resource management skills. The ten knowledge categories in O*Net are: (1) Arts and humanities, (2) Administration and management, (3) Communications, (4) Education and training, (5) Engineering and technology, (6) Health services, (7) Law and public safety, (8) Manufacturing and production, (9) Mathematics and science and (10) Transportation.

[7] The search for direct synonyms is done manually on the websites: www.wordreference.com, https://synonyms.reverso.net https://www.sinonimosgratis.com/.

excluded numbers and words that did not convey meaningful information and were not part of the stop-word corpus from the Python NLTK library. [8]

4) Creation of unigrams and bigrams for skills and knowledge keywords: The primary objective of our algorithm was to identify O*NET descriptor keywords within the vacancy description text. Our algorithm searched for single words (unigrams) and combinations of two words (bigrams). To accomplish this, we used the original keywords and their synonyms. While most descriptors and their synonyms consist of single words, some expressions are composed of multiple words, such as *reading comprehension*, *learning strategies*, and *financial resources management*. In such cases, we created bigrams by combining the words from the original O*NET categorization or their synonyms. Subsequently, we used the unigrams and all possible combinations of bigrams to search within a vacancy description. This process was performed after text normalization, removal of stop words, and elimination of regular expressions. For instance, for the skill *reading comprehension*, a set of possible bigrams would include reading interpretation, interpretation reading, reading understanding, and so on. For the skill *financial resources management*, the algorithm searched for combinations such as management of financial resources, administration resources financial, financial management, financial administration, resources administration, and so forth.

We also incorporated another technique in our robustness checks. Along with text normalization and the removal of standard expressions, we applied a lemmatization procedure to both O*NET descriptors and vacancy texts. To accomplish this, we used the spaCy Python package. Lemmatization is a Natural Language Processing technique that involves converting each word to its grammatical root or lemma, which represents the word's most basic form or dictionary form. By employing lemmatization, we aimed to increase the matching likelihood between vacancy descriptions and O*NET texts, particularly when the keywords in the text differ due to verb conjugations, tense variations, or subtle grammatical differences related to time or quantity.

Furthermore, as an additional robustness check, we also incorporated the consideration of trigrams during the processing of O*NET keywords. Therefore, we expanded our search to include combinations of three consecutive words. By including trigrams, we enhanced the comprehensiveness of the matching algorithm and further captured relevant keyword combinations within a vacancy description.

4.2 Using the Jaccard Function to classify skills and knowledge requirements

After applying text processing, we obtained a collection of keywords and synonyms associated with each of the 35 O*NET skill categories and 33 O*NET knowledge categories. The next step involved searching for these keywords within the processed description text of each vacancy. A vacancy is considered to possess a skill or knowledge requirement if at least

[8] Examples of prepositions are: a, ante, con, de, desde, durante, en, para, por. Examples of conjuntions are: y, e, ni, o, si, como, así.

one of the associated expressions, keywords, or synonyms is identified within the job description. We used the Jaccard similarity index to determine the strength of the relationship between a requirement (skill or knowledge) and the vacancy description. The Jaccard index calculates the ratio of the intersection's cardinality between two sets to the union's cardinality of both sets (Baharav et al., 2020; Leskovec et al., 2014). By employing the Jaccard index, we can assess the degree of similarity between a skill or knowledge requirement and a vacancy description. This index allows us to identify the most relevant requirements that align with the description and provide valuable insights into the skill and knowledge demands of the given vacancy. Formally, for two sets A and B, the Jaccard similarity index can be expressed as follows:

$$J(A,B) = \frac{\# \, (A \cap B)}{\# \, (A \cup B)} \quad (4.1)$$

In our specific case, sets A and B correspond to the skill or knowledge requirements set and the vacancy description text, respectively. The set of requirements encompasses both unigrams and bigrams, including all possible combinations. For each combination of vacancy *i* and skill/knowledge category *k*, we calculated the Jaccard index ($J_{i,k}$). This index ranges between 0 and 1, where a value of 0 indicates no intersection between A and B ($A \cap B = \emptyset$), and a value of 1 signifies that A and B are identical (A=B). By computing the similarity index for all vacancies and requirements, we obtained an N × K matrix, where N is equal to the number of vacancies, and K equals the number of skill and knowledge categories. This matrix captured the similarity between each vacancy and each requirement, allowing us to quantify the matching degree between vacancy descriptions and specific skill/knowledge categories. This matrix can be summarized as follows:

$$J_{N \times K} = \begin{bmatrix} j_{11} & \cdots & j_{1K} \\ \vdots & \ddots & \vdots \\ j_{N1} & \cdots & j_{NK} \end{bmatrix} \quad (4.2)$$

To determine the skill and knowledge requirements for each vacancy, we examined the row-wise maximum values of the Jaccard similarity matrix $J_{N \times K}$. Thus, we looked for the category with the highest Jaccard index for each vacancy among all skill and knowledge categories. Such skill is defined as the most important skill requirement for vacany *i*:

$$S_{1i} = \max_{j \in 1,2,..,K} (j_{i,1}, \dots, j_{i,K}) \quad (4.3)$$

For our baseline estimations, we identified the top three most important skills and knowledge categories (S_{1i}, S_{2i}, S_{3i}) by identifying the categories with the three highest Jaccard index for each vacancy. Finally, we followed the same approach for the lemmatized texts, which have undergone normalization and lemmatization procedures. The robustness checks section provides further details and insights.

5. Returns to skills and knowledge in the formal labor market

To examine the impact of skills on wages announced in a vacancy description, we estimate the following general wage equation (5.1). The wage posted for vacancy *i*, in city *c*, at time *t*, in occupation *o*, and by firm *f*, ($Wage_{i,c,t,o,f}$), is modeled as a function of skill requirements and control covariates. We allowed up to three skill or knowledge categories that a vacancy might require, as indicated by index $k = 1, 2, 3$. Wages are measured as multiples of the national monthly minimum wage (MW), because all vacancy descriptions include such information.

$$Wage_{i,c,t,o,f} = \alpha + \sum_{k=1}^{3} Skill_{k,i,c,o,f} \cdot \beta_{k,i} + X'_{i,c,t,o,f}\gamma + \delta_c + \tau_t + \pi_o + \mu_f + u_{i,c,t,o,f} \quad (5.1)$$

where, $Skill_{i,c,o,f}$ is a vector of required skills, aggregated following the O*NET classification for skill groups as follows: (1) content, (2) process, (3) complex problem-solving skills, (4) resource management skills, (5) social skills, (6) systems skills, and (7) technical skills. In Appendix B (Table B1), we present the complete description of skills from O*NET. Vector $X'_{i,c,t,o,f}$ contains a set of educational attainment dummy variables and experience required in the vacancy description. We include a series of fixed effects $\delta_c, \tau_t, \pi_o, \mu_f$ by city, time, occupation, and firm, respectively. Finally, $u_{i,c,t,o,f}$ stands for the regression error. Since we are interested in the return of a given skill, regardless of the order in the identification process described in section 4, the estimate of interest can be represented as: $\beta_i = \sum_{k=1}^{3} \beta_{k,i}$.

Similarly, we estimate an analogous equation to determine the returns for knowledge, which is represented by equation (5.2). This equation follows the same structure as before, but replaces the $Skill\_Group_{k,i,c,o,f}$ vector with a vector of required group knowledge dummy variables denoted as $Knowledge_{k,i,c,o,f}$. As before, we use description of the knowledge categories in O*NET which are presented in Appendix B (Table B2).

$$Wage_{i,c,t,o,f} = \alpha + \sum_{k=1}^{3} Knowledge_{k,i,c,o,f} \cdot \beta_{k,i} + X'_{i,c,t,o,f}\gamma + \delta_c + \tau_t + \pi_o + \mu_f + u_{i,c,t,o,f} \quad (5.2)$$

Based on the theoretical framework presented in section 2, we argue that wages are influenced by skill and knowledge requirements, as well as labor productivity which may vary across firms. Potential additional heterogeneity at the firm level, such as wage bargaining power or market power, is controlled through firm fixed effects. We also account for unobserved heterogeneity at the occupation level, which could be correlated with skill or knowledge requirements, by including occupation fixed effects. By leveraging on the granularity of our fixed effects, we are able to identify wage changes associated with skill and knowledge requirements within cities, firms, and occupations. As detailed in the subsequent section, skills and knowledge significantly explain wage variations within cities, firms, and occupations, even after controlling for educational and experience requirements.

6. Results

6.1 Returns to skills

Table 2 presents the estimation results of Equation 5.1. In column (1), we report the results without control variables or fixed effects. In column (2), we include control variables, and in columns (3) to (5), we estimate nested models that progressively incorporate city, time, occupation, and firm fixed effects. The most comprehensive specification is presented in column (6), which includes control variables and city, time, occupation, and firm fixed effects. The dependent variables in all specifications are posted wages measured as multiples of the national monthly minimum wage (MW).

In general, across all specifications, the estimated coefficients for the skill requirements are positive and mostly statistically significant. Our preferred specification, which includes all fixed effects, provides the most robust results. Social skills are used as the reference category among the skill requirements. In Figure 1, we present the estimates of skill returns ($\beta_i = \sum_{k=1}^{3} \beta_{k,i}$ in Equation 5.1). In Figure 2, we present the p-values of the test of statistical difference in the returns. All returns of aggregated skill categories are statistically different from social skills` returns; nevertheless, most skill categories exhibit a subtle difference compared to this reference category. For instance, process skills, linked to learning capacity, contribute an additional 0.15 MW to the predicted average salary compared to social skills. Content skills, which encompass general skills applicable across various domains, contribute an additional 0.17 MW; resource management skills add an extra 0.18 MW. Finally, complex problem-solving skills account for a 0.23 MW increase in average wages compared to social skills (see Figure 1). Figure 2 summarizes the statistical tests of differences in wage returns across all skill categories. Apart from system and social skills, the wage return of all skill categories are statistically similar.

The skills that explain the variation in wages the most are systems skills, which are associated with the ability to comprehend, monitor, and enhance technical and social systems. Systems skills explain an average wage increase of 0.74 MW. Overall, all aggregated skills positively contribute to wage variation within firms and occupations, with effects ranging from 0.15 MW to 0.74 MW. Further examination is needed to understand the underlying reasons for these findings. Therefore, we estimate regressions with a more detailed set of skills, considering all 33 basic skills proposed by the O*NET categorization.

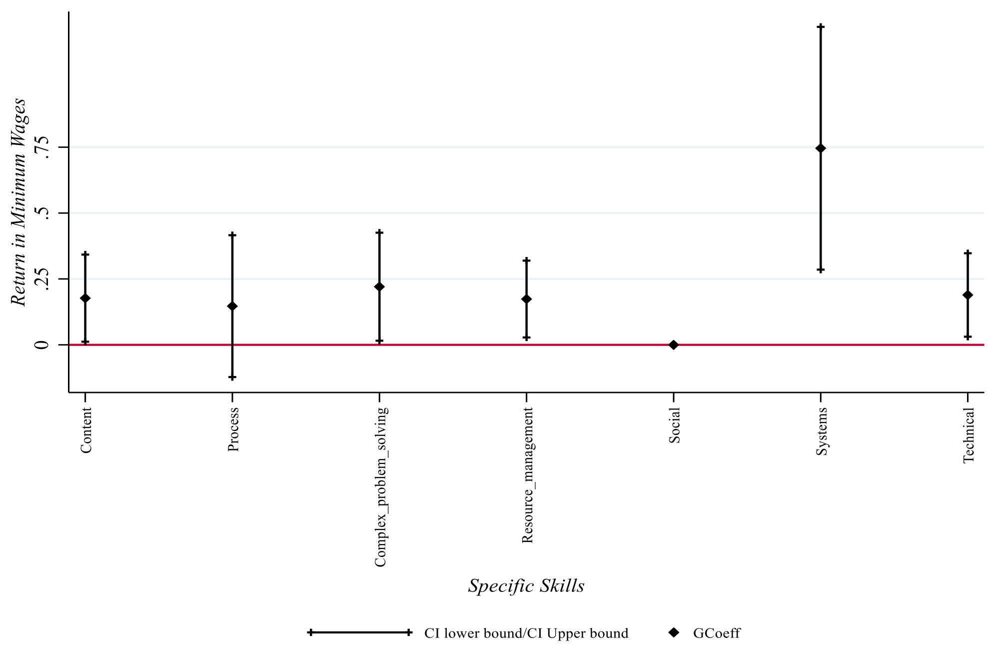
Table 2. *Skills' Returns Regression*

Variables	(1)	(2)	(3)	(4)	(5)	(6)
			Wage			
Skills						
1st Skill						
Content	0.27***	0.18***	0.16***	0.16***	0.09***	0.05**
	(0.02)	(0.01)	(0.05)	(0.05)	(0.03)	(0.02)
Process	0.67***	0.18***	0.15	0.15	0.12**	0.06
	(0.03)	(0.03)	(0.11)	(0.10)	(0.05)	(0.04)
Complex problem solving skills	0.77***	0.41***	0.47**	0.46**	0.30*	0.22
	(0.08)	(0.07)	(0.23)	(0.22)	(0.16)	(0.14)
Resource management skills	0.63***	0.21***	0.18***	0.17***	0.10***	0.06**
	(0.01)	(0.01)	(0.05)	(0.05)	(0.03)	(0.02)
Systems skills	1.16***	0.60***	0.62***	0.60***	0.39***	0.36***
	(0.08)	(0.07)	(0.16)	(0.16)	(0.11)	(0.11)
Technical skills	0.65***	0.33***	0.31***	0.30***	0.16***	0.12***
	(0.01)	(0.01)	(0.07)	(0.06)	(0.03)	(0.02)
2nd Skill						
Content	0.33***	0.25***	0.23**	0.22***	0.13**	0.10***
	(0.02)	(0.01)	(0.09)	(0.09)	(0.05)	(0.04)
Process	0.43***	0.15***	0.14***	0.13***	0.12***	0.08***
	(0.03)	(0.02)	(0.04)	(0.04)	(0.04)	(0.03)
Complex problem solving skills	0.37***	0.09**	0.17**	0.17**	0.07	0.01
	(0.06)	(0.05)	(0.09)	(0.08)	(0.06)	(0.07)
Resource management skills	0.44***	0.17***	0.14***	0.14***	0.09***	0.08***
	(0.01)	(0.01)	(0.04)	(0.04)	(0.03)	(0.02)
Systems skills	0.84***	0.50***	0.48***	0.46***	0.29***	0.24***
	(0.06)	(0.06)	(0.11)	(0.11)	(0.08)	(0.08)
Technical skills	0.34***	0.16***	0.14***	0.14***	0.08***	0.06***
	(0.01)	(0.01)	(0.03)	(0.03)	(0.02)	(0.02)
3rd Skill						
Content	0.23***	0.12***	0.10*	0.09**	0.04*	0.02
	(0.02)	(0.02)	(0.05)	(0.05)	(0.02)	(0.02)
Process	0.36***	0.08***	0.05	0.05	0.04	0.01
	(0.03)	(0.02)	(0.04)	(0.04)	(0.04)	(0.03)
Complex problem solving skills	0.30***	0.15***	0.22**	0.20**	0.07	-0.00
	(0.05)	(0.04)	(0.10)	(0.08)	(0.06)	(0.04)
Resource management skills	0.37***	0.12***	0.10***	0.09***	0.05***	0.04**
	(0.02)	(0.01)	(0.03)	(0.02)	(0.02)	(0.02)
Systems skills	0.63***	0.31***	0.29***	0.28***	0.17***	0.14***
	(0.04)	(0.04)	(0.07)	(0.06)	(0.05)	(0.04)
Technical skills	0.27***	0.08***	0.06**	0.06**	0.03	0.02
	(0.01)	(0.01)	(0.03)	(0.03)	(0.02)	(0.02)
Education						
None		0.34***	0.19	0.24	0.32	0.23
		(0.09)	(0.25)	(0.24)	(0.20)	(0.21)
High School		-0.04	-0.25	-0.20	-0.08	-0.07

		(0.09)	(0.23)	(0.22)	(0.17)	(0.18)
Technical		-0.14	-0.29	-0.24	-0.18	-0.16
		(0.09)	(0.23)	(0.23)	(0.18)	(0.18)
College		1.17***	0.99***	1.05***	0.65***	0.61***
		(0.09)	(0.21)	(0.21)	(0.20)	(0.19)
Years work experience		0.47***	0.47***	0.48***	0.44***	0.40***
		(0.01)	(0.06)	(0.06)	(0.06)	(0.05)
Constant	1.30***	0.94***	1.15***	1.11***	1.36***	1.48***
	(0.02)	(0.09)	(0.22)	(0.21)	(0.19)	(0.21)
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	98,277	98,277	98,277	98,276	98,215	97,636
R-squared	0.05	0.40	0.42	0.43	0.49	0.56
City-FE	No	No	Yes	Yes	Yes	Yes
Time-FE	No	No	No	Yes	Yes	Yes
Occupation-FE	No	No	No	No	Yes	Yes
Firm-FE	No	No	No	No	No	Yes

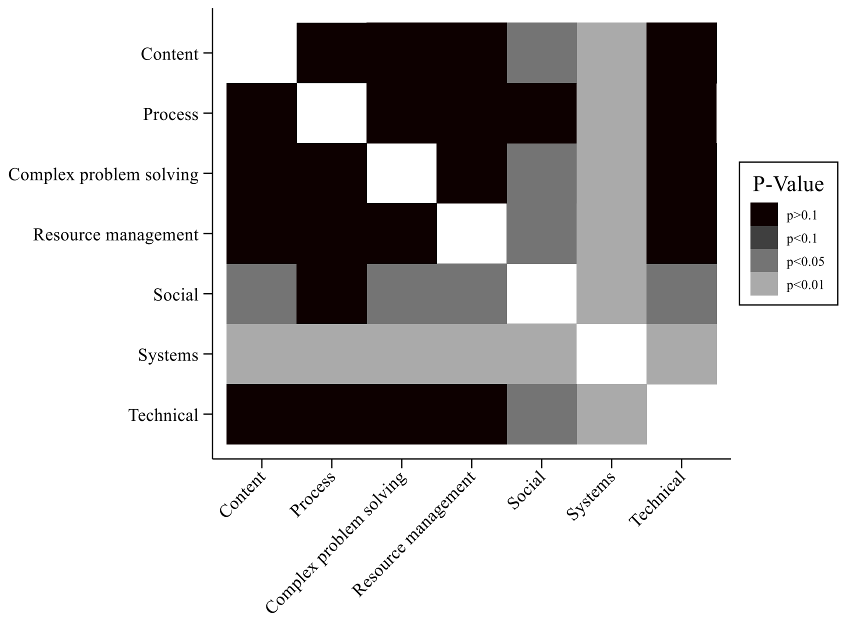
Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses. This table shows the estimation of Equation 5.1. The dependent variable is the wage expressed in terms of minimum wages. Regression in column 1 does not consider control variables or fixed effects, while column 2 includes control variables. From columns 3 to 6, we progressively incorporate fixed effects, culminating in regression 6, the most comprehensive specification, including city, time, occupation, and firm. Standard errors are clustered at the occupation level.

Figure 1. Returns of skills by group



Notes: This figure shows the estimates of the returns for aggregated skill categories ($\beta_i = \sum_{k=1}^{3} \beta_{k,i}$ in Equation 5.1). The dependent variable is the wage expressed in terms of minimum wages, in which we use the most comprehensive specification, including city, time, occupation, and firm. The lines represent 95% confidence interval. Standard errors are clustered at the occupation level.

Figure 2. Differences across skills by group return estimates

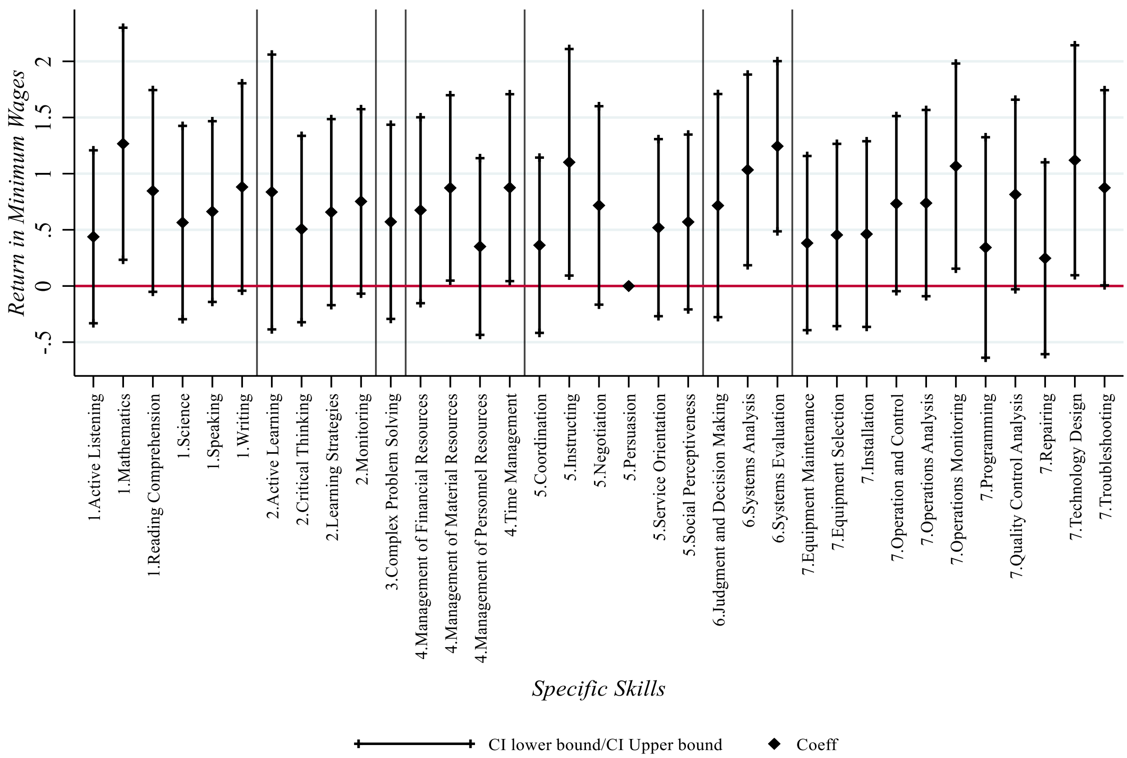


Notes: This figure shows the p-values of the statistical test for differences in the returns across aggregated skill categories. A p-value below 0.05 indicates a statistically significant difference between the returns of the respective skill categories.

6.2 Returns to granular skills

In Figure 3, we provide a detailed breakdown of the returns for each skill group into their respective essential skill components. To achieve this, we estimate an equation similar to Equation 5.1. Instead of using the vector of aggregated skills, we employ the vector of specific skills, comprising 35 individual categories. Figure 3 illustrates the wage return for each specific skill category. As before, returns can be interpreted compared to a reference category, persuasion, which belongs to the aggregate category of social skills. The coefficients are labeled with their corresponding group numbers, such as 1 for content, 2 for process, 3 for problem-solving, 4 for resource management, 5 for social skills, 6 for systems skills, and 7 for technical skills. The returns for specific skills exhibit sizeable variability, ranging from 0.25 MW to 1.27 MW. As indicated by the heatmap in Figure 4, there is a lot of heterogeneity in the returns for specific skills; when it comes to the analysis of granular skill categories, a sizeable share of these returns is statistically different across the categories. Let us consider, for instance, mathematical skills, which have the highest estimated return (second row in the matrix); this estimate is statistically different from 20 of the 33 estimates for the return of other skill categories.

Figure 3. Returns of specific skills by group



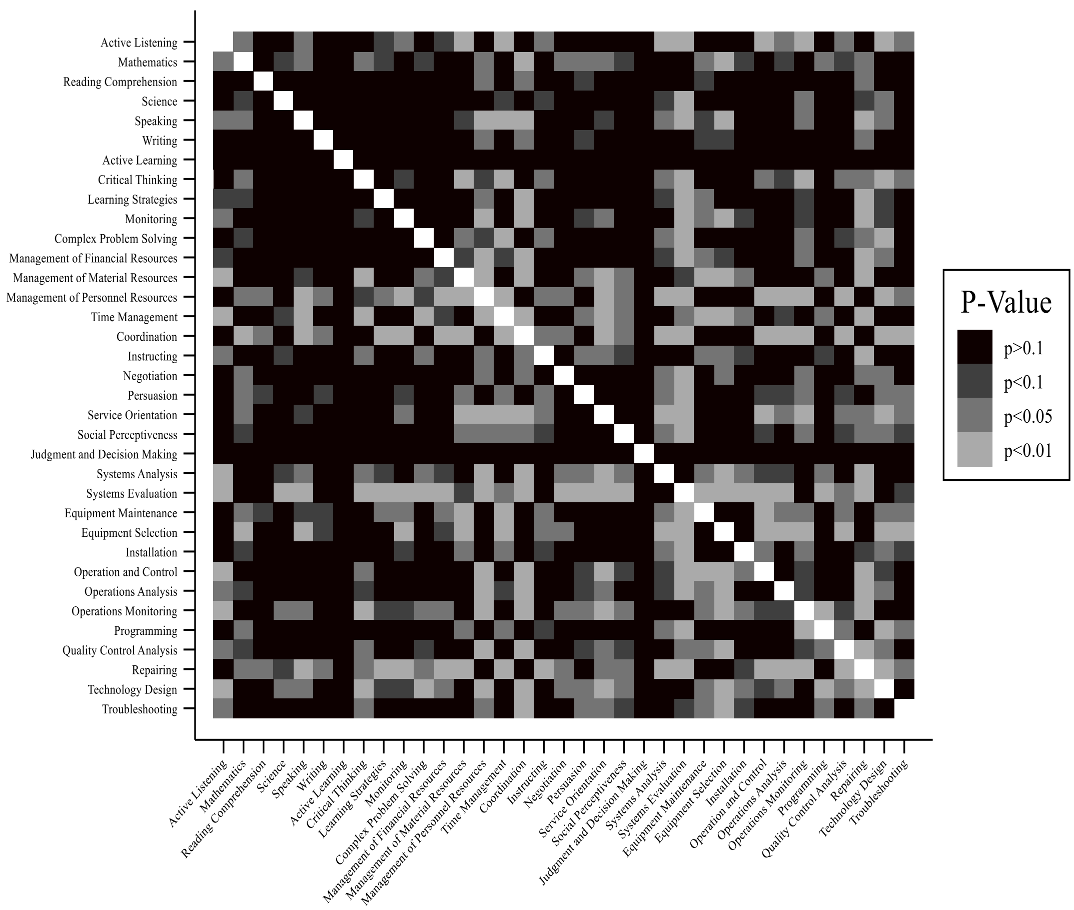
Notes: This figure shows the estimates of the returns for aggregated and disaggregated skill categories, using the vector of specific skills, comprising 35 individual categories ($\beta_i = \sum_{k=1}^{3} \beta_{k,i}$ in Equation 5.1). The dependent variable is the wage expressed in terms of minimum wages, in which we use the most comprehensive specification, including city, time, occupation, and firm. The lines represent 95% confidence interval. Standard errors are clustered at the occupation level.

Within the content category (1), which encompasses foundational skills essential for working with and acquiring more skills in other domains, the highest point estimate returns are associated with mathematics (1.27 MW), which exhibits one of the highest returns among all specific skills. As indicated in Figure 3, the return in this category is statistically different from most other categories, including the reference category (persuasion). Other categories such as writing and reading comprehension have important returns (around 0.85 MW) and are statistically significant at the 90% confidence level. These results describe how, within an occupation and conditional on educational attainment and work experience required by the vacancy, having general skills in math and writing provides one of the highest wage returns we document in this study. Content skills as the later ones can be associated with the development of various tasks and are fundamental in the acquisition of additional skills.

In the process category (2), which comprises skills that contribute to acquiring knowledge and skills in other domains, the highest point estimate return is observed for active learning (0.87 MW), but it is not statistically different from the reference category. Regarding complex problem-solving skills, which consist of a single unique skill related to solving ill-defined problems in complex settings, the return is 0.57 MW; nevertheless, this return is not statistically different from the reference category. Regarding resource management skills, the wage returns for managing material resources and time management display the highest returns, around 0.88 MW, in both cases statistically significant.

As previously discussed, the social skills group exhibits the lowest aggregated return. We observe higher returns and statistically significant returns for instructing (1.10 MW). The systems skill group, defined by the capability to utilize, comprehend, monitor, and enhance socio-technical systems, displays the highest returns estimated from Equation 5.1. This outcome can be attributed to the system-evaluation and system-analysis skills, which exhibits one of the highest returns among the specific skills (1.24 MW and 1 MW, respectively). Lastly, regarding technical skills, the highest returns are associated with operation monitoring and technology design.

Figure 4. Differences across specific skills return estimates



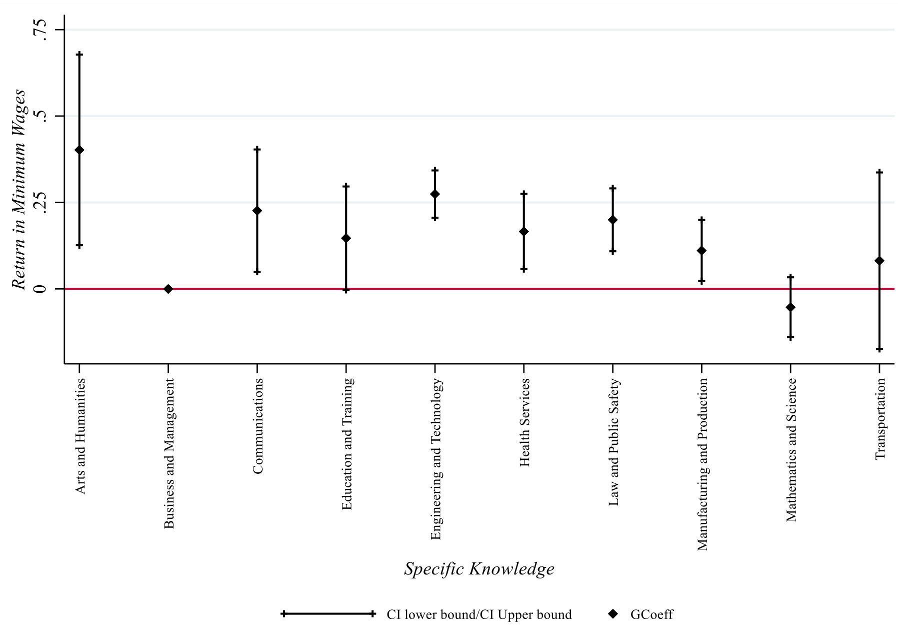
Notes: This figure shows the p-values of the statistical test for differences in the returns across the 35 disaggregated skill categories. A p-value below 0.05 indicates a statistically significant difference between the returns of the respective skill categories.

6.3 Return to knowledge

Table 3 provides the estimation results of Equation 5.2 using the O*NET knowledge categories. Similar to the previous analysis, we examine different specifications starting from column (1), which does not include control variables or fixed effects, and progressively incorporate them in subsequent columns. As before, the most comprehensive specification, presented in column 6, includes control variables and fixed effects for city, time, occupation, and firm.

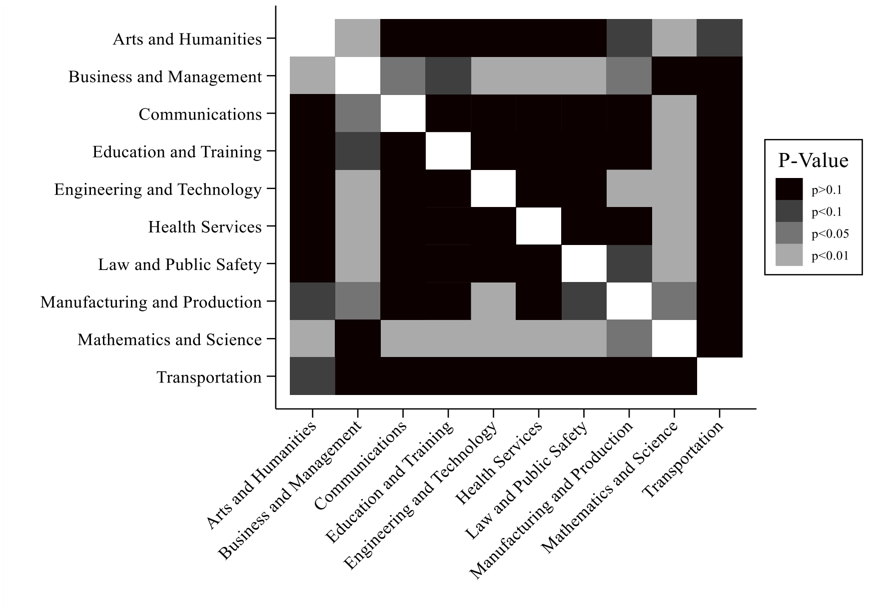
In Figure 5, we present the estimates of aggregated knowledge returns ($\beta_i = \sum_{k=1}^{3} \beta_{k,i}$ in equation 5.2), and in Figure 6, we present the p-values for the test of statistical difference among these estimates. Across the different specifications, the effects of knowledge requirements generally remain positive and significant for most knowledge categories. In addition, the returns to these aggregated skills are statistically different across several groups. Among the various knowledge categories, business and management knowledge is employed as the reference category in the analysis. After incorporating, city, occupation and firm fixed effects, aggregated categories, such as Arts and Humanities or Engineering and Technology, show substantial returns. Knowledge in Arts and Humanities consistently contributes to wage differences, with a cumulative return of 0.41 MW, while expertise in Engineering and Technology leads to a rise of 0.27 MW. Likewise, fields like Communications and Law exhibit positive returns, with values around 0.20 MW.

Figure 5. Returns of knowledge by group



Notes: This figure shows the estimates of the returns for aggregated knowledge categories ($\beta_i = \sum_{k=1}^{3} \beta_{k,i}$ in equation 5.2). The dependent variable is the wage expressed in terms of minimum wages, in which we use the most comprehensive specification, including city, time, occupation, and firm. The lines represent 95% confidence interval. Standard errors are clustered at the occupation level.

Figure 6. Differences across knowledge return estimates



Notes: This figure shows the p-values of the statistical test for differences in the returns across aggregated knowledge categories. A p-value below 0.05 indicates a statistically significant difference between the returns of the respective knowledge categories.

Most of the ten aggregated knowledge categories contribute positively and significantly to wage variation within firms and occupations, with effects up to 0.41 MW relative to the reference category. As before, to gain a deeper understanding of the underlying factors driving these findings, we conduct regressions using a more detailed breakdown of knowledge groups. Specifically, we consider all 33 basic knowledge categories proposed by the O*NET content model.

Table 3. Knowledge' returns regression

Variables	(1)	(2)	(3)	(4)	(5)	(6)
			Wage			
Knowledge						
1st Knowledge						
Arts and Humanities	0.24***	0.16***	0.13	0.13	0.17**	0.14*
	(0.05)	(0.04)	(0.13)	(0.13)	(0.09)	(0.08)
Communications	0.31***	0.11***	0.12	0.12*	0.12***	0.12**
	(0.03)	(0.02)	(0.07)	(0.07)	(0.04)	(0.06)
Education and Training	0.05*	-0.14***	-0.18*	-0.18*	0.10***	0.04
	(0.03)	(0.03)	(0.09)	(0.09)	(0.04)	(0.04)
Engineering and Technology	0.86***	0.49***	0.44***	0.43***	0.21***	0.19***
	(0.02)	(0.02)	(0.11)	(0.11)	(0.04)	(0.03)
Health Services	0.44***	0.18***	0.18**	0.17**	0.11***	0.09***
	(0.02)	(0.02)	(0.09)	(0.09)	(0.03)	(0.03)

Law and Public Safety	0.17***	0.25***	0.24***	0.24***	0.21***	0.13***
	(0.02)	(0.02)	(0.08)	(0.08)	(0.06)	(0.03)
Manufacturing and Production	-0.00	0.07***	0.05*	0.04*	0.04**	0.03*
	(0.02)	(0.01)	(0.03)	(0.03)	(0.02)	(0.02)
Mathematics and Science	-0.02	0.07***	0.03	0.03	0.03	0.02
	(0.02)	(0.01)	(0.04)	(0.04)	(0.04)	(0.02)
Transportation	-0.29***	-0.01	-0.04	-0.04	0.04	0.06
	(0.02)	(0.03)	(0.07)	(0.07)	(0.07)	(0.06)
2nd Knowledge						
Arts and Humanities	0.23***	0.17***	0.14	0.15	0.18*	0.10
	(0.05)	(0.04)	(0.12)	(0.12)	(0.09)	(0.06)
Communications	0.14***	0.06***	0.08**	0.09**	0.07***	0.07***
	(0.03)	(0.02)	(0.04)	(0.04)	(0.02)	(0.03)
Education and Training	0.01	-0.01	-0.02	-0.03	0.05	0.06
	(0.02)	(0.02)	(0.05)	(0.05)	(0.04)	(0.04)
Engineering and Technology	0.32***	0.17***	0.15***	0.15***	0.03	0.03
	(0.02)	(0.01)	(0.03)	(0.03)	(0.03)	(0.02)
Health Services	0.59***	0.13***	0.12*	0.12*	0.06**	0.06
	(0.02)	(0.02)	(0.07)	(0.07)	(0.03)	(0.03)
Law and Public Safety	0.13***	0.11***	0.10***	0.09***	0.07***	0.05***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)
Manufacturing and Production	0.14***	0.07***	0.04*	0.05*	0.04*	0.04
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)
Mathematics and Science	-0.15***	-0.05***	-0.08**	-0.07**	-0.06**	-0.04
	(0.01)	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)
Transportation	-0.27***	0.04*	-0.03	-0.03	0.06	0.09
	(0.03)	(0.02)	(0.11)	(0.11)	(0.09)	(0.08)
3rd Knowledge						
Arts and Humanities	0.31***	0.22***	0.22***	0.22***	0.25***	0.17***
	(0.05)	(0.04)	(0.06)	(0.06)	(0.06)	(0.06)
Communications	-0.01	-0.00	0.02	0.02	0.02	0.04
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Education and Training	-0.02	-0.03	-0.01	-0.01	0.06	0.04
	(0.03)	(0.02)	(0.05)	(0.05)	(0.04)	(0.03)
Engineering and Technology	0.20***	0.12***	0.09***	0.09***	0.03	0.05***
	(0.02)	(0.01)	(0.03)	(0.02)	(0.02)	(0.02)
Health Services	0.45***	0.06***	0.07	0.06	0.03	0.02
	(0.03)	(0.02)	(0.05)	(0.05)	(0.04)	(0.04)
Law and Public Safety	0.11***	0.06***	0.05*	0.05*	0.04*	0.02
	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)
Manufacturing and Production	0.10***	0.10***	0.08**	0.08**	0.05**	0.04
	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)
Mathematics and Science	-0.20***	-0.07***	-0.09***	-0.09***	-0.06***	-0.04**
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Transportation	-0.47***	-0.17***	-0.18***	-0.18***	-0.10**	-0.06**
	(0.02)	(0.02)	(0.05)	(0.04)	(0.05)	(0.03)
Education						
None		0.18**	0.08	0.12	0.22	0.15
		(0.08)	(0.22)	(0.22)	(0.18)	(0.19)

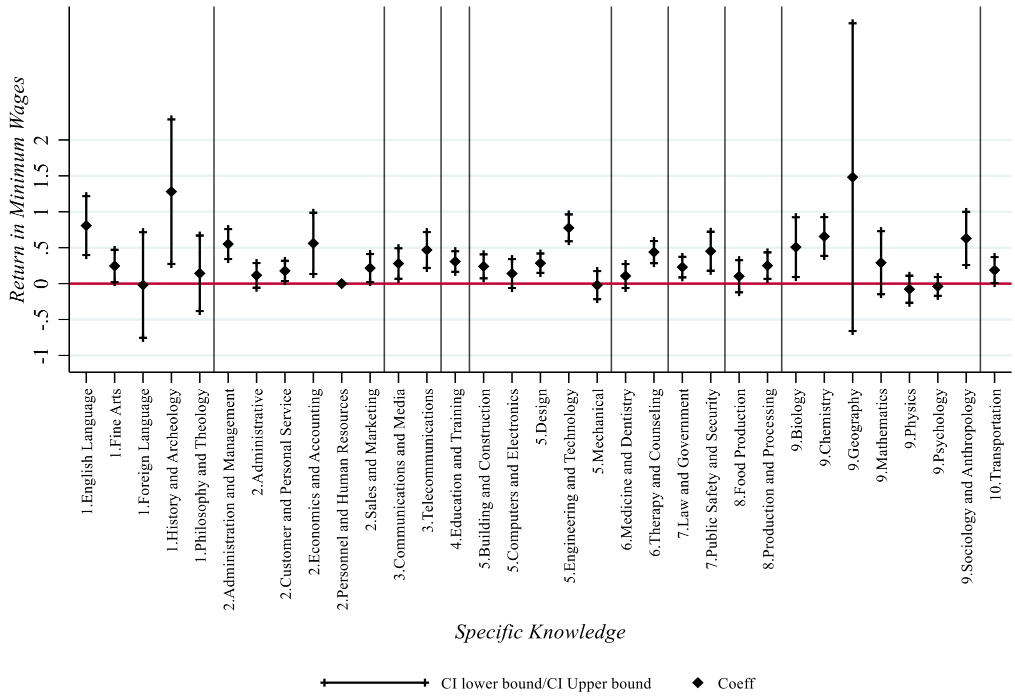
	(1)	(2)	(3)	(4)	(5)	(6)
High School		-0.18**	-0.32*	-0.29	-0.16	-0.14
		(0.08)	(0.19)	(0.19)	(0.15)	(0.17)
Technical		-0.27***	-0.37*	-0.33*	-0.27*	-0.23
		(0.08)	(0.20)	(0.19)	(0.16)	(0.16)
College		1.04***	0.91***	0.95***	0.58***	0.55***
		(0.08)	(0.18)	(0.18)	(0.18)	(0.17)
Years work experience		0.47***	0.48***	0.48***	0.45***	0.41***
		(0.01)	(0.05)	(0.05)	(0.05)	(0.05)
Constant	2.00***	1.29***	1.44***	1.40***	1.54***	1.60***
	(0.01)	(0.08)	(0.18)	(0.17)	(0.16)	(0.18)
Observations	106,777	106,777	106,777	106,772	106,707	106,120
R-squared	0.07	0.42	0.43	0.44	0.49	0.57
City-FE	No	No	Yes	Yes	Yes	Yes
Time-FE	No	No	No	Yes	Yes	Yes
Occupation-FE	No	No	No	No	Yes	Yes
Firm-FE	No	No	No	No	No	Yes

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses. This table shows the estimation of Equation 5.2. Regression in column 1 does not consider control variables or fixed effects, while column 2 includes control variables. From columns 3 to 6, we progressively incorporate fixed effects, culminating in regression 6, the most comprehensive specification, including city, time, occupation, and firm. Standard errors are clustered at the occupation level.

6.4 Granular knowledge returns

In Figure 7, we present a detailed analysis of the returns for each knowledge group, disaggregated into their components. To accomplish this, we estimate equation 5.2 using the vector of specific knowledge variables, which consists of 33 categories. This figure illustrates the decomposition of wage returns for each knowledge group. The coefficients are labeled with their respective group numbers, such as 1 for Arts and Humanities, 2 for Business and Management, 3 for Communications, 4 for Education, 5 for Engineering and Technology, 6 for Health Services, 7 for Law and Public Safety, 8 for Manufacturing and Production, 9 for Mathematics and Science, and finally 10 for Transportation. The heatmap in Figure 8 shows important heterogeneity in the returns for specific knowledge categories; most of the returns are statistically different across categories.

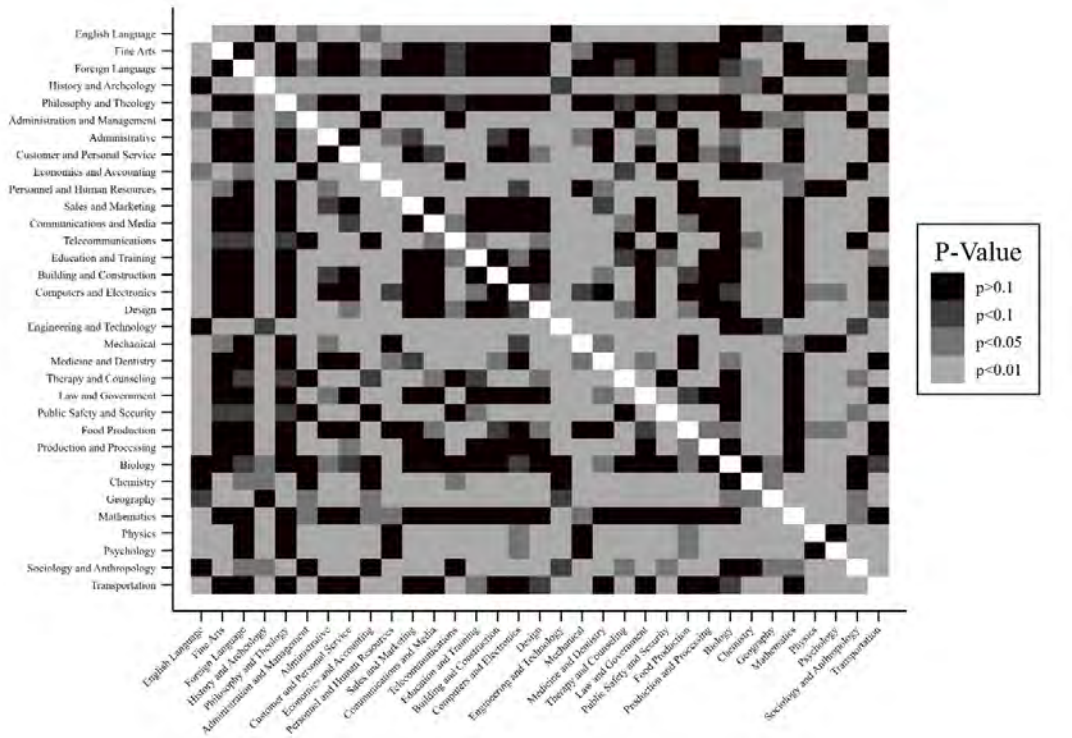
Figure 7. Returns to Specific Knowledge by Groups



Notes: The figure shows the graphical representation of the β coefficients for each of the 33 knowledge categories, obtained from estimating an equation like Equation 5.2 (instead of using the vector of aggregated knowledge, we employ the vector of specific categories). In this regression, we control by education level, years of experience, and fixed effects of city, period, occupation, and firm. Confidence intervals are constructed at 95%. Standard errors are clustered at the occupation level.

Within the Arts and Humanities category, the highest returns are associated with History and Archeology (1.28 MW) and English (0.81 MW); these categories statistically differ from 23 out of the 33 specific knowledge categories. Economics and Accounting demonstrate the highest return (0.56 MW) in the Business and Management knowledge group, while Administration exhibits the lowest return. In the Communications category, the higher returns are driven by Telecommunications, with a point estimate of 0.47 MW. Within the Engineering knowledge category, Engineering and Technology have the highest point estimate return (0.78 MW). In the Manufacturing and Production category, knowledge of production and processing stands out with the highest point estimate return of 0.25 MW. Finally, within the Science knowledge category, chemistry exhibits the highest statistically significant estimated return (1.48 MW), while Physics, Geography and Psychology do not show statistically significant returns compared to the reference category.

Figure 8. Differences Across Specific Knowledge Return Estimates



Notes: This figure shows the p-values of the statistical test for differences in the returns across the 33 disaggregated knowledge categories. A p-value below 0.05 indicates a statistically significant difference between the returns of the respective skill categories.

It is important to note that the return to knowledge categories identified in this paper explains the variation of wages within occupations. Therefore, a significant portion of the return associated with some of these knowledge categories is already accounted for by educational qualifications and the inclusion of occupation and firm fixed effects. For example, let us consider knowledge in Health Services; they show positive returns, although they are not substantially different from the average knowledge return. This finding aligns with the aggregated regression results presented in Table 3. Without fixed effects or other controls, the health service category displays one of the highest returns (1.48 MW). However, this return decreases to 0.17 MW once the regression incorporates controls and all fixed effects.

7. Heterogeneity

We conduct a heterogeneity analysis of the returns of skills and knowledge aggregated categories by different levels of education and experience, which are required information in the description of posted vacancies. We present the results of the estimation of equation 5.1 for each one of the analyzed categories in Appendix C (Table C1). Regarding skills, in section 6.1 we report that the aggregated skills that contribute the most to explaining the variation in wages are complex problem-solving skills and systems skills. Heterogeneity results show that these high returns are mainly driven by the sample of vacancies requiring post-secondary education (technical or college degrees). The returns of these latter skills are very low or not statistically significant for vacancies requiring lower educational attainment.

For vacancies requiring lower educational levels, the most important skills are technical, which turns out to be important for all educational levels. In the case of vacancies requiring a high school degree, process skills are the ones with the highest returns. In terms of experience required by the vacancies, we find that the highest returns in system and technical skills are explained by the sample of vacancies requiring some experience especially for those vacancies requiring more than 2 years of previous experience. In the case of vacancies which require no previous experience, the highest returns are for technical, resource management and content skills. In the latter case this is an expected finding given that content skill refers to foundational skills essential for working with and acquiring more skills in other domains.

Regarding knowledge categories, we present the results of the estimation of equation 5.1 for each one of the categories in Appendix C (Table C2). In our baseline estimations, within the dimensions of occupation and firm, the knowledge categories that exhibit the most notable returns are engineering, communications and law. Heterogeneity results indicate that the group of vacancies requiring college education are responsible for the higher returns. Regarding previous work experience, the effect of knowledge is concentrated for vacancies requiring low experience levels (one and two years), especially for engineering, humanities and law.

8. Robustness checks

Lemmatization

In Appendix D (Table D1), we present regression for disaggregated skill returns using a lemmatization procedure to both the O*NET descriptors and the vacancy text. Lemmatization is a Natural Language Processing technique that involves converting each word in a text into its grammatical root or lemma, representing the word’s basic form or dictionary form. The results are similar to our baseline estimation. Regarding knowledge, we apply the same robustness check, and the results are presented in Appendix D (Table D2). Overall, the results are also similar to the baseline estimation results.

Considering Trigrams

In an additional robustness check, we create trigrams by combining the words from the original O*NET categorization or their synonyms. Subsequently, we use the unigrams and all possible combinations of bigrams and trigrams to search within the vacancy description. As explained in section 4, this process is performed after text normalization, removing stop words, and eliminating regular expressions. Results of this process are presented in Appendix D, Figure D1 in the case of skills and Figure D2 in the case of knowledge. In both cases, results are similar to the baseline estimation.

Alternative skills and knowledge classifications

In this paper, we use the O*NET database of occupations because it has been widely used for similar purposes in the literature, and its category descriptions are very detailed. This latter feature is particularly important for our research, as more comprehensive category descriptions facilitate better identification of skills and knowledge. However, there is also a local database we can use: the Unique Classification of Occupations for Colombia (CUOC). This alternative database is a hierarchical classification system encompassing 670 occupations. Similar to O*NET, the CUOC describes skills and knowledge categories to identify the competencies required in various jobs. In this robustness check, we estimate the returns to skills and knowledge using the CUOC for the identification algorithm. As presented in Appendix E, the skills and knowledge categories and their descriptions are similar, with two main differences: the CUOC has additional categories in the most granular disaggregation, and the descriptions of each granular category are shorter.

Regarding specific skills, the estimated returns are similar to our baseline results. For instance, math and reading show the highest returns in the content group, while management of financial and time resources have the highest returns in the management skill group. Categories such as system evaluation and technology design have the highest significant returns in the system and technical skill groups. There are additional categories from COUC, which are not included in O*NET, that are statistically significant with considerable returns, such as “proactivity,” “work under pressure,” and “interpretation”.

Regarding granular knowledge categories, the results are similar as well. For instance, English has the highest return in the arts and humanities group; however, the history and archeology category is no longer significant, even though it has a positive return. In the business and management group, economics and accounting still have one of the highest significant returns. However, in the models using the CUOC, the highest return in this category belongs to administrative knowledge instead of administration/management as in the baseline estimation, which we attribute to the similarity of the terms. Engineering and technology have the highest significant return, both in this robustness check and in the baseline estimation. Finally, in the science knowledge group, as in the baseline estimation, most categories have positive returns. Nevertheless, using the CUOC database, some of the estimates are statistically significant, such as knowledge of geography.

9. Conclusion and policy implications

This paper shows the importance of online vacancy posts as a valuable resource for analyzing labor markets in developing countries. The widespread use of the internet and online job portals has revolutionized how firms advertise job vacancies and how job seekers search for opportunities. The study emphasizes that the information provided in job postings is valuable as it encapsulates specific skills and knowledges required by firms for different job positions. Online vacancy descriptions offer a detailed analysis of skill variations across job roles within firms and occupations, providing an alternative and complementary source of information for labor market research. This research investigates the returns to skills and knowledge in the formal labor market. By analyzing a public online job board in Colombia and processing skill and knowledge categorizations from the O*NET occupational database, the study estimates the wage return of different skills and knowledge within firms and occupations. This provides valuable insights into the economic incentives and labor demand for the accumulation of skill and knowledges in formal labor markets.

The paper identifies the skills and knowledge categories with the highest impact on wages. To achieve this, the paper introduces a methodology for skill and knowledge identification from textual descriptions of online job vacancies. Unlike previous approaches, this study directly incorporates posted wage information within each vacancy, avoiding the limitation of associating wages from external sources with the content of vacancy descriptions. The data-driven approach using keyword extraction from the O*NET database enables a more objective and systematic skill identification process. We show that most of the main results are robust to variations in the methodology using alternative databases for the definition of skill or knowledge categories.

Since our analysis explains the variation of wages at a granular level, within occupations and firms, one of the most outstanding findings is that, for a given occupation, the mastery of some general skills and knowledges significantly contributes to explaining wage differentials within occupations. Skills and knowledge categories, such as mathematics, writing, English, economics, history, science, instructing, and management, offer wage premiums across various occupations, highlighting the importance of a diversified skill set for job candidates. This conclusion challenges the traditional notion of workers accumulating a homogeneous and rigid amount of human capital, instead supporting alternative views that emphasize the multidimensional nature of human capital accumulation and the flexibility of workers in utilizing different skills to develop several tasks. Moreover, our findings show that, in addition to the abilities needed for a given occupation, the formal labor market is willing to reward certain skill and knowledge categories that might boost the capabilities for performing occupation-related tasks. Most of the categories with significant and positive returns are not specific to a given occupation; instead, they are tools that can be beneficial in the performance of a variety of tasks.

In a heterogeneity analysis of these returns, we find that the highest return for aggregated skills is driven mainly by vacancies that require post-secondary education and previous work experience, especially more than two years. Similarly, heterogeneity results indicate that the group of vacancies requiring college education drives the higher returns for aggregated knowledge categories. Regarding experience, the effect of knowledge is concentrated on vacancies requiring low experience levels (one and two years), especially for engineering, humanities, and law.

The findings of this study have important implications for policymakers and stakeholders in the labor market. By identifying the skills and knowledge that lead to higher wages, policymakers can design targeted training and educational programs to enhance the employability and earning potential of the labor force. Furthermore, understanding the demand for specific skills and knowledge in the labor market can help guide workforce development policies and improve the matching of job seekers with job opportunities.

References

Acemoglu, D., & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of Labor Economics* (Vol. 4, Issue PART B). https://doi.org/10.1016/S0169-7218(11)02410-5

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics*, *118*(4). https://doi.org/10.1162/003355303322552801

Baharav, T. Z., Kamath, G. M., Tse, D. N., & Shomorony, I. (2020). Spectral Jaccard Similarity: A New Approach to Estimating Pairwise Sequence Alignments. *Patterns*, *1*(6). https://doi.org/10.1016/j.patter.2020.100081

Bennett, F., Escudero, V., Liepmann, H., & Podjanin, A. (2022). Using Online Vacancy and Job Applicants' Data to Study Skills Dynamics. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4198744

Card, D., & Lemieux, T. (2001). Can falling supply explain the rising return to college for younger men? A cohort-based analysis. *Quarterly Journal of Economics*, *116*(2). https://doi.org/10.1162/00335530151144140

Chaparro, J. C. (2016). Skills over the Life Cycle: Evidence from the United States and the Philippines (Doctoral dissertation, University of Minnesota). http://hdl.handle.net/11299/182724

Deming, D. J., & Noray, K. (2020). Earnings dynamics, changing job skills, and stem careers. In *Quarterly Journal of Economics* (Vol. 135, Issue 4). https://doi.org/10.1093/qje/qjaa021

Deming, D., & Kahn, L. B. (2018). Skill requirements across firms and labor markets: Evidence from job postings for professionals. *Journal of Labor Economics*, *36*(S1). https://doi.org/10.1086/694106

Hansen, S., Ramdas, T., Sadun, R., & Fuller, J. (2021). The demand for executive skills (No. w28959). *National Bureau of Economic Research*. https://doi.org/10.3386/w28959

Hershbein, B., & Kahn, L. B. (2018). Do recessions accelerate routine-biased technological change? evidence from vacancy postings. *American Economic Review*, *108*(7). https://doi.org/10.1257/aer.20161570

Katz, L. F., & Murphy, K. M. (1992). Changes in relative wages, 1963-1987: Supply and demand factors. *Quarterly Journal of Economics*, *107*(1). https://doi.org/10.2307/2118323

Kuhn, P., & Mansour, H. (2014). Is Internet Job Search Still Ineffective? *Economic Journal*, *124*(581). https://doi.org/10.1111/ecoj.12119

Kuhn, P., & Skuterud, M. (2004). Internet job search and unemployment durations. *American Economic Review*, *94*(1). https://doi.org/10.1257/000282804322970779

Leskovec, J., Rajaraman, A., & Ullman, J. D. (2014). Mining of Massive Datasets. In *Mining of Massive Datasets*. https://doi.org/10.1017/cbo9781139924801

Lise, J., & Postel-Vinay, F. (2020). Multidimensional skills, sorting, and human capital accumulation. *American Economic Review*, 110(8), 2328–2376. https://doi.org/10.1257/aer.20162002

Marinescu, I. (2019). Reprint of: The general equilibrium impacts of unemployment insurance: Evidence from a large online job board. *Journal of Public Economics*, *171*. https://doi.org/10.1016/j.jpubeco.2019.03.004

Marinescu, I., & Wolthoff, R. (2020). Opening the black box of the matching function: The power of words. *Journal of Labor Economics*, *38*(2). https://doi.org/10.1086/705903

Modestino, A. S., Shoag, D., & Balance, J. (2020). Upskilling: Do employers demand greater skill when workers are plentiful? *Review of Economics and Statistics*, *102*(4). https://doi.org/10.1162/rest_a_00835

Morales, L. F., Ospino, C., & Nicole, A. (2021). *Online Vacancies and its Role in Labor Market Performance* (No. 1174; Borradores de Economía). https://doi.org/10.32468/be.1174

Morales, L. F., Ospino, C., & Amaral, N. (2024). Online vacancies and their role in labor market performance. In B. Elsner & S. W. Polachek (Eds.), *Big data applications in labor economics* (Part A, Research in Labor Economics, Vol. 52A, pp. 119-159). Emerald Publishing Limited. https://doi.org/10.1108/S0147-91212024000052A023

National Research Council. (2010). A Database for a Changing Economy: Review of the Occupational Information Network (O*NET), N.T. Tippins & M.L. Hilton, eds. The National Academy Press, Panel Review of the Occupational Information Network. https://nap.nationalacademies.org/catalog/12814/a-database-for-a-changing-economy-review-of-the-occupational

Peterson, N. G., Mumford, M. D., Borman, W. C., Jeanneret, P. R., Fleishman, E. A., Levin, K. Y., & Dye, D. M. (2001). Understanding work using the Occupational Information Network (O*NET): Implications for practice and research. *Personnel Psychology*, 54(2), 451-492. https://doi.org/10.1111/j.1744-6570.2001.tb00100.x

Sanders, S., & Taber, C. (2012). Life-cycle wage growth and heterogeneous human capital. *Journal of Labor Economics*, 30(4), 697-722. https://doi.org/10.1146/annurev-economics-080511-111011

Schmitt, N. W., Highhouse, S. E., & Weiner, I. B. (2013). Handbook of psychology: Industrial and organizational psychology, Vol. 12. John Wiley & Sons, Inc. https://psycnet.apa.org/record/2012-28468-000

Tinbergen, J. (1974). Substitution Of Graduate by Other Labour. *Kyklos*, *27*(2). https://doi.org/10.1111/j.1467-6435.1974.tb01903.x

Tinbergen, J. (1975). *Income Differences: recent research*. North-Holland Publishing Company, Amsterdam.

Welch, F. (1973). Black-white differences in returns to schooling. *American Economic Review*, *63*(5).

Ziegler, L. (2022). Skill Demand and Wages. Evidence from Linked Vacancy Data. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4183423

Appendix A.
Example of information in SISE in Spanish with English translation.

Figure A1.

Illustrative Example of Information in SISE (Spanish)

Técnico en soporte y mantenimiento

Código: 1596587558-448

Empresa: Confidencial

Información adicional

Cargo Requerido:	Auxiliar de soporte técnico
Empresa:	Confidencial
Salario:	1 SMMLV
Tipo de Contrato:	Obra
Minimo nivel de estudio:	Técnica Profesional
Minima experiencia requerida (meses):	12
Distribución:	Departamento(s): ANTIOQUIA; Municipio(s): MEDELLÍN,
Fecha limite de envió de candidatos:	27 de Enero de 2017
Prestadores Asociados:	CAJA DE COMPENSACIÓN FAMILIAR DE ANTIOQUIA -COMFAMA - COMFAMA CARRERA 44-CLAUSTRO SAN IGNACIO
Empleo susceptible a teletrabajo:	No

Descripción de la vacante

Se solicita Técnico en soporte y mantenimiento para mantener en un estado óptimo los equipos, reparación de equipos, instalación de programas y aplicaciones, mantenimiento técnico de los equipos y los programas, soporte de informes.

Figure A2.

Illustrative Example of Information in SISE (English)

Support & Maintenance Technician

Code: 1596587558-448

Company: Confidential

Additional information

Required Charge:	Technical Support Assistant
Company:	Confidential
Wage:	1 SMMLV
Type of contract:	Construction Site
Minimum level of study:	Professional Technique
Minimum experience required (months):	12
Distribution:	Deparment: ANTIOQUIA; Municipality: MEDELLÍN
Deadline for submitting candidates:	January 27, 2017
Associate Providers:	CAJA DE COMPENSACIÓN FAMILIAR DE ANTIOQUIA –COMFAMA- COMFAMA CARRERA 44 CLAUSTRO SAN IGNACIO
Employment susceptible to teleworking :	No

Job Description:

Support and maintenance technician is request to maintain equipment in optimal condition, equipment repair, installation of programs and applications, technical maintenance of equipment and programs, report support.

Appendix B.
Skills and Knowledge Groups O*NET

Table B1

Skills Groups, Subgroups, and Specific Skills

Skills groups	Skill subgroups	Specific skills
Basic Skills	**Content** Background structures needed to work with and acquire more specific skills in a variety of different domains.	Active Listening Mathematics Reading Comprehension Science Speaking Writing
	Process Procedures that contribute to the more rapid acquisition of knowledge and skill across a variety of domains.	Active Learning Critical Thinking Learning Strategies Monitoring
Cross-Functional Skills	**Complex Problem-Solving Skills** Developed capacities used to solve novel, ill-defined problems in complex, real-world settings.	Complex Problem Solving
	Resource Management Skills Developed capacities used to allocate resources efficiently.	Management of Financial Resources Management of Material Resources Management of Personnel Resources Time Management
	Social Skills Developed capacities used to work with people to achieve goals.	Coordination Instructing Negotiation Persuasion Service Orientation Social Perceptiveness
	Systems Skills Developed capacities used to understand, monitor, and improve socio-technical systems.	Judgment and Decision Making Systems Analysis Systems Evaluation
	Technical Skills Developed capacities used to design, set-up, operate, and correct malfunctions involving application of machines or technological systems.	Equipment Maintenance Equipment Selection Installation Operation and Control Operations Analysis Operations Monitoring Programming Quality Control Analysis Repairing Technology Design Troubleshooting

Notes. Table created by the authors. Information source: The O*NET content model; worker requirements (Basic and Cross-functional skills).

Table B2

Knowledge Groups and Specific Knowledge

Knowledge groups	Specific Knowledge
Arts and Humanities Knowledge of facts and principles related to the branches of learning concerned with human thought, language, and the arts.	English Language Fine Arts Foreign Language History and Archeology Philosophy and Theology
Business and Management Knowledge of principles and facts related to business administration and accounting, human and material resource management in organizations, sales and marketing, economics, and office information and organizing systems	Administration and Management Administrative Customer and Personal Service Economics and Accounting Personnel and Human Resources Sales and Marketing
Communications Knowledge of the science and art of delivering information.	Communications and Media Telecommunications
Education and Training Knowledge of principles and methods for curriculum and training design, teaching and instruction for individuals and groups, and the measurement of training effects.	
Engineering and Technology Knowledge of the design, development, and application of technology for specific purposes	Building and Construction Computers and Electronics Design Engineering and Technology Mechanical
Health Services Knowledge of principles and facts regarding diagnosing, curing, and preventing disease, and improving and preserving physical and mental health and well-being	Medicine and Dentistry Therapy and Counseling
Law and Public Safety Knowledge of regulations and methods for maintaining people and property free from danger, injury, or damage; the rules of public conduct established and enforced by legislation, and the political process establishing such rules	Law and Government Public Safety and Security
Manufacturing and Production Knowledge of principles and facts related to the production, processing, storage, and distribution of manufactured and agricultural goods.	Food Production Production and Processing
Mathematics and Science Knowledge of the history, theories, methods, and applications of the physical, biological, social, mathematical, and geography.	Biology Chemistry Geography Mathematics Physics Psychology Sociology and Anthropology
Transportation Knowledge of principles and methods for moving people or goods by air, rail, sea, or road, including the relative costs and benefits.	

Notes. Table created by the authors. Information source: The O*NET content model; worker requirements (Knowledge´s).

Appendix C.
Alternative Specifications: Heterogeneity

Table C1

Heterogeneity: Skills Returns Regressions

Variables	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)
			Education				***Experience (years)***	
	None	***High School***	***Technical***	***College***	***Post graduate***	***None***	***0 - 2***	***2 +***
Skills								
1st skill								
Content	-0.04	0.05***	0.02*	0.12***	0.97	0.13***	0.03***	0.01
	(0.06)	(0.01)	(0.01)	(0.04)	(0.90)	(0.02)	(0.01)	(0.11)
Process	-0.14	0.12***	0.02	0.09*	0.45	0.08	0.06**	0.20
	(0.09)	(0.03)	(0.02)	(0.05)	(0.90)	(0.05)	(0.02)	(0.14)
Complex problem-solving skills	-0.07	0.07*	0.22***	0.36**		0.12	0.20***	0.33
	(0.16)	(0.04)	(0.06)	(0.15)		(0.17)	(0.06)	(0.25)
Resource management skills	0.01	0.04***	0.03***	0.12***	0.09	0.13***	0.03***	0.09
	(0.04)	(0.01)	(0.01)	(0.03)	(0.51)	(0.02)	(0.01)	(0.06)
Systems skills	0.35	0.05	0.42***	0.46***		0.17	0.21***	1.01***
	(0.52)	(0.04)	(0.10)	(0.12)		(0.11)	(0.07)	(0.22)
Technical skills	0.17***	0.05***	0.11***	0.22***	0.35	0.18***	0.10***	0.19***
	(0.04)	(0.01)	(0.01)	(0.04)	(0.60)	(0.03)	(0.01)	(0.06)
2nd skill								
Content	0.03	0.07***	0.03*	0.16***	0.54	0.16***	0.03***	0.19*
	(0.05)	(0.01)	(0.01)	(0.04)	(0.79)	(0.03)	(0.01)	(0.10)
Process	-0.05	0.04**	0.08***	0.09*	-0.57	0.08**	0.07***	-0.07
	(0.09)	(0.02)	(0.02)	(0.05)	(0.90)	(0.04)	(0.02)	(0.12)
Complex problem-solving skills	-0.04	0.05	0.00	-0.05		-0.03	0.07**	-0.17
	(0.19)	(0.03)	(0.04)	(0.10)		(0.11)	(0.03)	(0.26)
Resource management skills	0.08	0.06***	0.03***	0.12***	0.05	0.13***	0.05***	0.00
	(0.05)	(0.01)	(0.01)	(0.03)	(0.62)	(0.02)	(0.01)	(0.07)
Systems skills	0.12	0.07**	0.08*	0.37***		0.22**	0.15***	0.30
	(0.21)	(0.03)	(0.05)	(0.11)		(0.11)	(0.05)	(0.20)
Technical skills	-0.01	0.04***	0.07***	0.09***	0.30	0.07***	0.05***	-0.04
	(0.04)	(0.01)	(0.01)	(0.04)	(0.59)	(0.02)	(0.01)	(0.06)
3rd skill								
Content	0.24***	0.04***	-0.02	-0.00	0.06	0.12***	0.03**	-0.20**
	(0.05)	(0.01)	(0.01)	(0.05)	(0.79)	(0.03)	(0.01)	(0.09)
Process	-0.11	0.10***	0.00	-0.03	-0.15	-0.03	0.00	-0.09
	(0.11)	(0.02)	(0.02)	(0.05)	(1.12)	(0.04)	(0.02)	(0.12)
Complex problem-solving skills	-0.23	-0.06**	0.05	0.01	-0.19	-0.01	0.02	-0.27
	(0.15)	(0.03)	(0.05)	(0.11)	(0.62)	(0.12)	(0.03)	(0.24)
Resource management skills	0.10**	0.06***	-0.01	-0.00	0.02	0.09***	0.01	-0.05

	(0.04)	(0.01)	(0.01)	(0.04)	(0.35)	(0.02)	(0.01)	(0.07)
Systems skills	0.22	0.06**	0.21***	0.11	-0.01	0.28***	0.10***	-0.06
	(0.21)	(0.03)	(0.05)	(0.07)	(0.92)	(0.08)	(0.03)	(0.16)
Technical skills	0.04	0.05***	-0.00	-0.00	-0.10	0.07***	0.02*	-0.07
	(0.04)	(0.01)	(0.01)	(0.04)	(0.39)	(0.02)	(0.01)	(0.07)
Education								
None						0.29	-0.17	0.67***
						(0.19)	(0.12)	(0.25)
High School						0.24	-0.36***	-0.14
						(0.19)	(0.12)	(0.26)
Technical						0.18	-0.32***	-0.20
						(0.19)	(0.12)	(0.26)
College						0.82***	0.39***	0.88***
						(0.19)	(0.13)	(0.27)
Years work experience	0.27***	0.08***	0.24***	0.52***	0.00		0.25***	0.46***
	(0.05)	(0.01)	(0.01)	(0.01)	(0.02)		(0.01)	(0.02)
Constant	1.70***	1.56***	1.42***	2.22***	2.36***	1.51***	1.71***	1.75***
	(0.06)	(0.01)	(0.02)	(0.05)	(0.55)	(0.19)	(0.13)	(0.28)
Observations	5,693	26,627	37,124	26,739	159	18,320	65,666	12,731
R-squared	0.49	0.57	0.39	0.52	0.81	0.45	0.44	0.58
City-FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time-FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation-FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm-FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses. This table shows the estimation of Equation 5.1. with grouping constraints by categories according to educational level (None, High School, Technical – Technological, and College) and required experience (None, 0-2, and 2 or more years). All regressions account for fixed effects of city, period, occupation, and firm. Regressions in columns 1 to 4 correspond to the grouping by educational level, with years of experience as a control variable. The remaining regressions (columns 5 to 7) represent the grouping by years of experience, controlling for educational level.

Table C2

Heterogeneity: Knowledge Returns Regressions

	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)
Variables			***Education***				***Experience(years)***	
	None	***High School***	***Technical***	***College***	***Post graduate***	***None***	***0 - 2***	***2 +***
Knowledge								
1st knowledge								
Arts and Humanities	0.06	0.12***	0.15***	0.13		0.01	0.11***	0.39
	(0.13)	(0.03)	(0.04)	(0.10)		(0.08)	(0.04)	(0.25)
Communications	0.24**	-0.01	0.04*	0.26***	0.95	0.12*	0.04**	0.50***
	(0.09)	(0.03)	(0.02)	(0.07)	(0.83)	(0.07)	(0.02)	(0.13)
Education and Training	-0.01	0.06*	0.12***	-0.11**	2.82**	-0.08	0.05*	-0.06
	(0.10)	(0.03)	(0.04)	(0.06)	(1.17)	(0.06)	(0.03)	(0.15)
Engineering and Technology	0.29***	0.06***	0.07***	0.27***	0.12	0.14***	0.15***	0.25***
	(0.07)	(0.01)	(0.02)	(0.04)	(0.58)	(0.03)	(0.01)	(0.06)
Health Services	0.36***	-0.06***	0.02*	0.14***	0.41	0.07	0.04***	0.08
	(0.07)	(0.02)	(0.01)	(0.04)	(0.43)	(0.05)	(0.02)	(0.08)
Law and Public Safety	0.05	0.00	0.08***	0.21***	0.22	0.05*	0.09***	0.16
	(0.04)	(0.01)	(0.02)	(0.05)	(0.44)	(0.03)	(0.01)	(0.10)
Manufacturing and Production	0.03	-0.02***	0.02	0.19***	0.49	-0.09***	0.03***	0.17*
	(0.05)	(0.01)	(0.02)	(0.05)	(0.44)	(0.02)	(0.01)	(0.09)
Mathematics and Science	-0.01	0.00	0.02	0.04	0.04	-0.01	-0.01	0.17**
	(0.05)	(0.01)	(0.01)	(0.04)	(0.96)	(0.02)	(0.01)	(0.07)
Transportation	0.04	0.05***	0.08*	-0.10		0.22***	-0.02	-0.25
	(0.06)	(0.02)	(0.04)	(0.14)		(0.06)	(0.02)	(0.16)
2nd knowledge								
Arts and Humanities	0.50	0.21***	0.10**	-0.07		-0.02	0.04	0.29
	(0.36)	(0.04)	(0.04)	(0.08)		(0.05)	(0.03)	(0.20)
Communications	0.13	0.01	0.06***	0.01	-1.09	-0.03	0.02	0.14
	(0.12)	(0.01)	(0.02)	(0.05)	(1.15)	(0.04)	(0.01)	(0.09)
Education and Training	-0.04	0.14***	0.01	-0.07	-1.23**	-0.08**	0.12***	-0.11
	(0.06)	(0.02)	(0.03)	(0.06)	(0.57)	(0.03)	(0.02)	(0.10)
Engineering and Technology	-0.03	0.07***	0.07***	-0.06*	-0.62	-0.03	0.06***	-0.04
	(0.04)	(0.01)	(0.01)	(0.03)	(0.58)	(0.03)	(0.01)	(0.06)
Health Services	-0.05	0.07*	0.07***	0.02	-0.50	0.10*	0.05***	-0.05
	(0.07)	(0.03)	(0.02)	(0.04)	(0.71)	(0.05)	(0.02)	(0.08)
Law and Public Safety	0.12**	0.04***	0.05***	-0.05	0.24	0.05	0.04***	-0.00
	(0.06)	(0.01)	(0.01)	(0.05)	(0.33)	(0.03)	(0.01)	(0.09)
Manufacturing and Production	-0.01	0.03***	-0.00	0.14***	-0.16	0.05	0.04***	0.09
	(0.07)	(0.01)	(0.01)	(0.05)	(0.26)	(0.04)	(0.01)	(0.09)
Mathematics and Science	-0.02	0.02**	-0.01	-0.14***	0.36	-0.03	-0.01	-0.23***
	(0.04)	(0.01)	(0.01)	(0.04)	(0.29)	(0.02)	(0.01)	(0.07)
Transportation	0.00	0.09***	0.12***	-0.01	-1.64**	0.13***	0.01	0.08

	(0.07)	(0.01)	(0.04)	(0.24)	(0.70)	(0.03)	(0.01)	(0.24)
3rd knowledge								
Arts and Humanities	-0.13	0.29***	0.11***	0.08		0.02	0.08***	0.55**
	(0.17)	(0.04)	(0.04)	(0.09)		(0.06)	(0.03)	(0.22)
Communications	0.13	0.01	0.06***	-0.04		0.03	0.03**	0.01
	(0.12)	(0.01)	(0.02)	(0.05)		(0.05)	(0.01)	(0.09)
Education and Training	-0.06	0.05**	0.14***	-0.01	0.64	0.00	0.08***	-0.00
	(0.09)	(0.02)	(0.03)	(0.06)	(0.39)	(0.04)	(0.02)	(0.11)
Engineering and Technology	0.05	0.09***	0.03**	0.04	-0.14	0.01	0.05***	-0.02
	(0.04)	(0.01)	(0.01)	(0.03)	(0.25)	(0.02)	(0.01)	(0.06)
Health Services	0.10*	0.07***	0.06***	-0.08**	-1.98***	-0.02	0.05***	-0.16**
	(0.06)	(0.02)	(0.02)	(0.04)	(0.64)	(0.04)	(0.02)	(0.07)
Law and Public Safety	0.05	0.01	0.03*	-0.00	-0.64	0.16***	-0.00	-0.09
	(0.05)	(0.01)	(0.02)	(0.05)	(0.48)	(0.04)	(0.01)	(0.08)
Manufacturing and Production	-0.02	0.01	0.07***	0.08	-0.91	0.00	0.01	0.13
	(0.07)	(0.01)	(0.02)	(0.05)	(0.60)	(0.03)	(0.01)	(0.09)
Transportation	-0.06	0.01*	-0.00	-0.12***	-0.30	-0.08***	-0.01	-0.10
	(0.04)	(0.01)	(0.01)	(0.04)	(0.28)	(0.02)	(0.01)	(0.07)
Education								
None						0.22	-0.21*	0.38*
						(0.19)	(0.11)	(0.21)
High School						0.25	-0.41***	-0.43*
						(0.19)	(0.11)	(0.23)
Technical						0.17	-0.37***	-0.41*
						(0.19)	(0.11)	(0.23)
College						0.87***	0.35***	0.69***
						(0.19)	(0.11)	(0.23)
Years Work Experience	0.28***	0.07***	0.23***	0.52***	0.02		0.25***	0.48***
	(0.04)	(0.01)	(0.01)	(0.01)	(0.02)		(0.01)	(0.02)
Constant	1.75***	1.59***	1.45***	2.41***	2.92***	1.75***	1.79***	1.89***
	(0.06)	(0.01)	(0.01)	(0.04)	(0.25)	(0.19)	(0.11)	(0.24)
Observations	6,235	29,736	39,314	29,318	218	20,025	71,187	13,996
R-squared	0.51	0.56	0.39	0.52	0.90	0.44	0.45	0.58
City-FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time-FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation-FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm-FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses. This table shows the estimation of Equation 5.2. with grouping constraints by categories according to educational level (None, High School, Technical – Technological, and College) and required experience (None, 0-2, and 2 or more years). All regressions account for fixed effects of city, period, occupation, and firm. Regressions in columns 1 to 4 correspond to the grouping by educational level, with years of experience as a control variable. The remaining regressions (columns 5 to 7) represent the grouping by years of experience, controlling for educational level.

Appendix D.
Robustness Check

Table D1

Robustness Check: Skills Returns Regressions (Lemmatization)

Variables	(1)	(2)	(3)	(4)	(5)	(6)
			Wage			
Skills						
1st skill						
Content	0.38***	0.21***	0.22***	0.22***	0.16***	0.11***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Process	0.70***	0.25***	0.23***	0.22***	0.16***	0.07**
	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Complex problem solving skills	0.63***	0.27***	0.31***	0.30***	0.19***	0.12**
	(0.08)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)
Resource management skills	0.66***	0.18***	0.16***	0.15***	0.08***	0.04***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Systems skills	1.19***	0.58***	0.60***	0.58***	0.40***	0.30***
	(0.09)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Technical skills	0.60***	0.26***	0.24***	0.24***	0.12***	0.08***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
2nd skill						
Content	0.36***	0.22***	0.22***	0.21***	0.12***	0.09***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Process	0.48***	0.16***	0.16***	0.15***	0.12***	0.09***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
Complex problem solving skills	0.35***	0.15***	0.19***	0.19***	0.12***	0.13***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Resource management skills	0.51***	0.17***	0.17***	0.16***	0.10***	0.07***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Systems skills	0.79***	0.47***	0.47***	0.46***	0.30***	0.23***
	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Technical skills	0.41***	0.18***	0.18***	0.18***	0.09***	0.07***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
3rd skill						
Content	0.28***	0.17***	0.16***	0.15***	0.07***	0.03***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Process	0.38***	0.09***	0.07***	0.07***	0.04**	-0.00

	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Complex problem solving skills	0.23***	0.10***	0.12***	0.11***	0.01	0.03
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Resource management skills	0.38***	0.13***	0.12***	0.11***	0.08***	0.06***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Systems skills	0.54***	0.24***	0.23***	0.22***	0.08**	0.04
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Technical skills	0.29***	0.09***	0.09***	0.09***	0.04***	0.03***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Education						
None		0.28***	0.16	0.19*	0.23**	0.15
		(0.08)	(0.11)	(0.11)	(0.11)	(0.10)
High School		-0.11	-0.28**	-0.24**	-0.15	-0.12
		(0.08)	(0.11)	(0.11)	(0.11)	(0.10)
Technical		-0.18**	-0.30***	-0.26**	-0.24**	-0.21**
		(0.08)	(0.11)	(0.11)	(0.11)	(0.10)
College		1.13***	0.99***	1.03***	0.60***	0.57***
		(0.08)	(0.11)	(0.11)	(0.11)	(0.10)
Years work experience		0.46***	0.47***	0.47***	0.43***	0.40***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	1.23***	0.99***	1.14***	1.11***	1.41***	1.53***
	(0.02)	(0.09)	(0.11)	(0.11)	(0.11)	(0.10)
Observations	107,346	107,346	107,346	107,342	107,282	106,703
R-squared	0.05	0.40	0.42	0.42	0.49	0.56
City-FE	No	No	Yes	Yes	Yes	Yes
Time-FE	No	No	No	Yes	Yes	Yes
Occupation-FE	No	No	No	No	Yes	Yes
Firm-FE	No	No	No	No	No	Yes

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses This table shows the estimation of Equation 5.1. The regressions are estimated based on the Jaccard index obtained from the lemmatized text. The dependent variable is the wage expressed in terms of minimum wages. Regression in column 1 does not consider control variables or fixed effects, while column 2 includes control variables. From columns 3 to 6, we progressively incorporate fixed effects, culminating in regression 6, the most comprehensive specification, including city, time, occupation, and firm.

Table D2

Robustness Check: Knowledge Returns Regressions (Lemmatization)

Variables	(1)	(2)	(3)	(4)	(5)	(6)
			Wage			
Knowledge						
1st knowledge						
Arts and Humanities	0.11**	0.07	0.04	0.04	0.12***	0.07**
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Communications	0.30***	0.05**	0.06***	0.06***	0.02	0.03
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Education and Training	-0.11***	-0.22***	-0.24***	-0.25***	-0.01	-0.03
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Engineering and Technology	0.65***	0.37***	0.33***	0.32***	0.11***	0.10***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Health Services	0.62***	0.28***	0.28***	0.27***	0.19***	0.15***
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Law and Public Safety	0.23***	0.21***	0.16***	0.16***	0.17***	0.10***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
Manufacturing and Production	-0.02	0.06***	0.06***	0.05***	0.04***	0.02
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Mathematics and Science	-0.04**	0.08***	0.04***	0.04***	0.04***	0.04***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Transportation	-0.35***	-0.09***	-0.12***	-0.12***	-0.05***	-0.03*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
2nd knowledge						
Arts and Humanities	0.08	0.10**	0.07*	0.07*	0.11***	0.08**
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Communications	0.08***	0.04**	0.07***	0.07***	0.07***	0.08***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Education and Training	-0.01	0.01	0.00	-0.01	0.08***	0.04**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
Engineering and Technology	0.28***	0.16***	0.14***	0.14***	0.05***	0.05***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Health Services	0.61***	0.14***	0.14***	0.13***	0.09***	0.06***
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Law and Public Safety	0.21***	0.10***	0.09***	0.08***	0.08***	0.05***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Manufacturing and Production	0.13***	0.08***	0.06***	0.06***	0.05***	0.04**

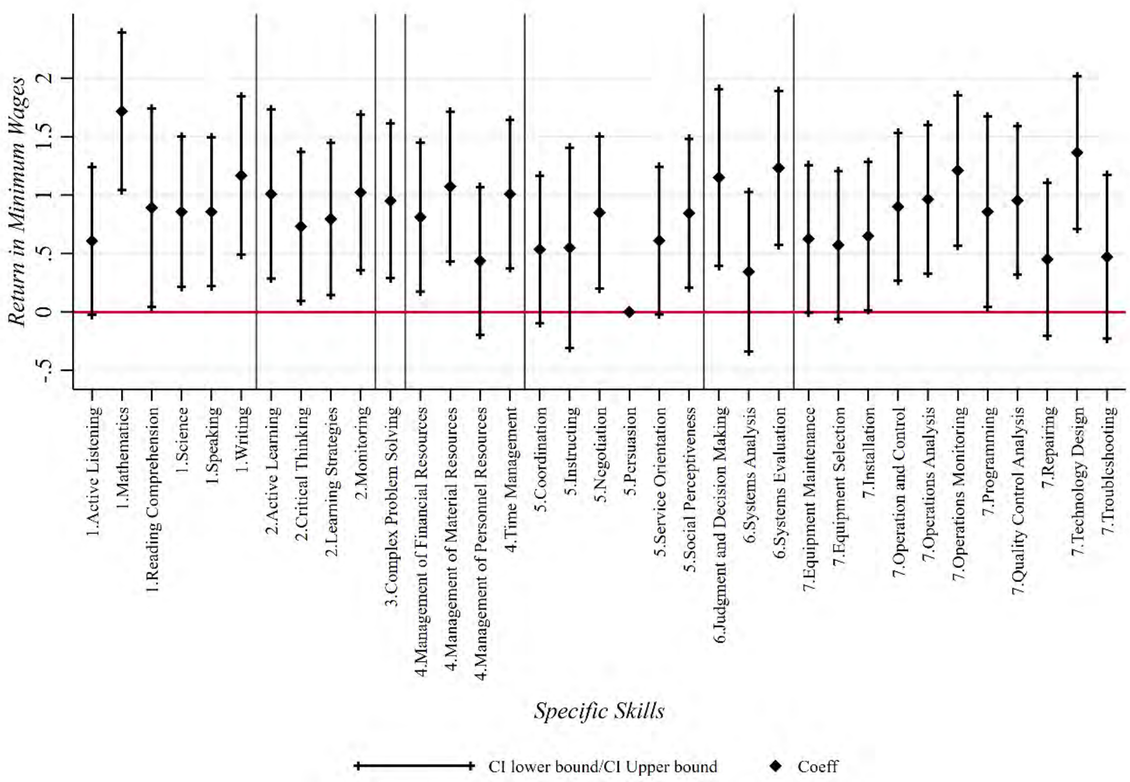
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Mathematics and Science	-0.14***	-0.03**	-0.05***	-0.04***	-0.03**	-0.02**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Transportation	-0.29***	-0.16***	-0.17***	-0.16***	-0.11***	-0.05***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
3rd skill						
Arts and Humanities	0.33***	0.22***	0.20***	0.20***	0.25***	0.17***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
Communications	0.05**	-0.02	0.00	0.00	0.02	0.04**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Education and Training	0.10***	0.03	0.04	0.04	0.09***	0.08***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
Engineering and Technology	0.23***	0.13***	0.10***	0.10***	0.04***	0.05***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Health Services	0.39***	0.04**	0.04**	0.04**	-0.02	-0.04**
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Law and Public Safety	0.21***	0.07***	0.06***	0.06***	0.04***	0.04**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Manufacturing and Production	0.13***	0.08***	0.07***	0.07***	0.04**	0.03*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Mathematics and Science	-0.10***	-0.03***	-0.05***	-0.05***	-0.04***	-0.03***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Transportation	-0.19***	-0.02	-0.06***	-0.06***	-0.02	-0.02
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Education						
None		0.22***	0.09	0.13	0.23**	0.16*
		(0.08)	(0.10)	(0.10)	(0.10)	(0.10)
High School		-0.14*	-0.32***	-0.28***	-0.14	-0.11
		(0.08)	(0.10)	(0.10)	(0.10)	(0.10)
Technical		-0.20**	-0.34***	-0.30***	-0.23**	-0.19**
		(0.08)	(0.10)	(0.10)	(0.10)	(0.10)
College		1.12***	0.96***	1.00***	0.63***	0.59***
		(0.08)	(0.10)	(0.10)	(0.10)	(0.10)
Years work experience		0.46***	0.47***	0.47***	0.44***	0.40***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	1.99***	1.25***	1.42***	1.39***	1.52***	1.58***
	(0.01)	(0.08)	(0.10)	(0.10)	(0.10)	(0.10)

Observations	111,679	111,679	111,679	111,674	111,614	111,028
R-squared	0.05	0.41	0.43	0.43	0.49	0.56
City-FE	No	No	Yes	Yes	Yes	Yes
Time-FE	No	No	No	Yes	Yes	Yes
Occupation-FE	No	No	No	No	Yes	Yes
Firm-FE	No	No	No	No	No	Yes

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses This table shows the estimation of Equation 5.2. The regressions are estimated based on the Jaccard index obtained from the lemmatized text. Regression in column 1 does not consider control variables or fixed effects, while column 2 includes control variables. From columns 3 to 6, we progressively incorporate fixed effects, culminating in regression 6, the most comprehensive specification, including city, time, occupation, and firm.

Figure D1

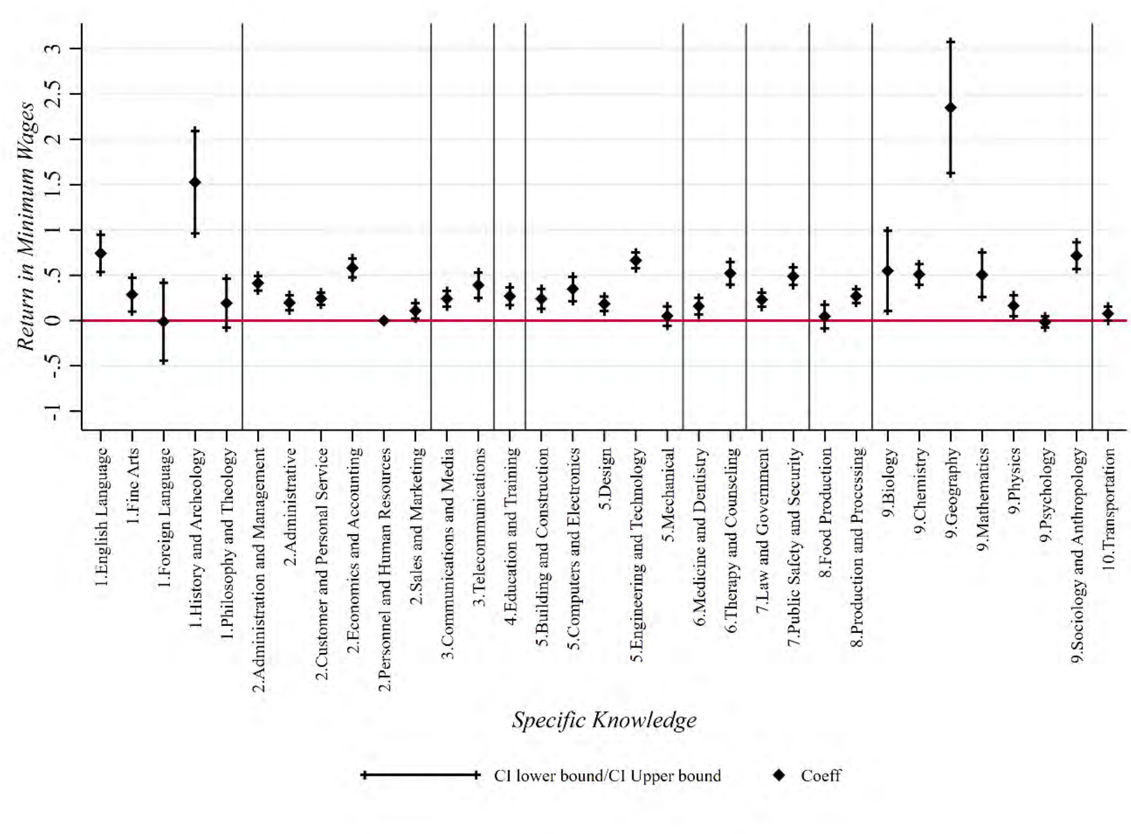
Robustness Check: Returns to Specific Skills by Groups (Trigrams)



Notes. Figure D1 shows the graphical representation of the β coefficients for each of the 35 skills, obtained from estimating an equation like Equation 5.1 (instead of using the vector of aggregated skills, we employ the vector of specific skills). In this regression, we control by education level, years of experience, and fixed effects of city, period, occupation, and firm. Confidence intervals are constructed at 95%.

Figure D2

Robustness Check: Returns to Specific Knowledge by Groups (Trigrams)



Notes. Figure D2 shows the graphical representation of the β coefficients for each of the 33 knowledge categories, obtained from estimating an equation like Equation 5.2 (instead of using the vector of aggregated knowledge, we employ the vector of specific categories). In this regression, we control by education level, years of experience, and fixed effects of city, period, occupation, and firm. Confidence intervals are constructed at 95%.

Appendix E.
Alternative skills and knowledge classifications (CUOC)

Table E1

Skills CUOC

Skills subgroups	Specific skills	CUOC Skills
Content	Active Listening	Active Listening
	Mathematics	Mathematics
	Reading Comprehension	Reading Comprehension
	Science	Science
	Speaking	Assertive communication
	Writing	Writing
Process	Active Learning	Applied learning, Flexibility and adaptability
	Critical Thinking	Critical Thinking
	Learning Strategies	Learning Strategies
	Monitoring	Activity evaluation and control
Complex Problem-Solving Skills	Complex Problem Solving	Complex Problem Solving
Resource Management Skills	Management of Financial Resources	Management of Financial Resources
	Management of Material Resources	Management of Material Resources
	Management of Personnel Resources	Management of Personnel Resources
	Time Management	Time Management
Social Skills	Coordination	Leadership, Teamwork.
	Instructing	Knowledge transmission
	Negotiation	Conciliation
	Persuasion	Persuasion
	Service Orientation	Service Orientation
	Social Perceptiveness	Interpersonal relationships.
Systems Skills	Judgment and Decision Making	Judgment and Decision Making
	Systems Analysis	Systems Analysis, Systemic thinking.
	Systems Evaluation	Systems Evaluation
Technical Skills	Equipment Maintenance	Equipment Maintenance
	Equipment Selection	Equipment Selection
	Installation	Assembly
	Operation and Control	Operation and Control
	Operations Analysis	Operations Analysis
	Operations Monitoring	Operations Monitoring
	Programming	Programming
	Quality Control Analysis	Quality Control Analysis
	Repairing	Repairing
	Technology Design	Technology Design
	Troubleshooting	Handling unexpected situations.
*CUOC skills not classified under O*NET.*		Following instructions, Proactivity, Creativity, Working under pressure, Interpretation.

Notes: This classification of skills areas was developed using the CUOC (Clasificación Única de Ocupaciones de Colombia) system, adapted to the categories of the O*NET system. Due to certain skills areas not being directly classifiable under O*NET, they are included in a separate group. *Source*: Classification of knowledge areas adapted from the CUOC (Clasificación Única de Ocupaciones de Colombia) system to the O*NET system, developed by the authors.

Table E2

Classification knowledge CUOC

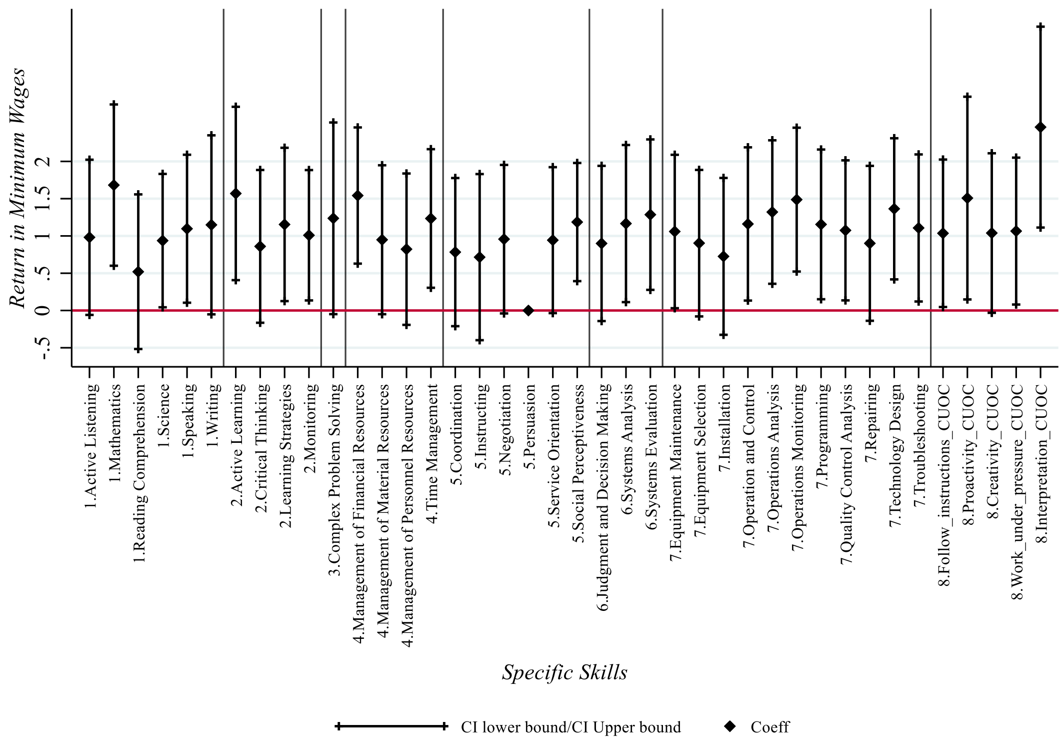
Knowledge groups	Specific Knowledge	CUOC Knowlege
Arts and Humanities	English Language	English Language
	Fine Arts	Fine Arts, Literature and Linguistics, Music and Performing Arts, Harmony and Composition, Instrumental Performance, Musical Instruments, Handicrafts
	Foreign Language	Foreign Language, Foreign Language
	History and Archeology	History and Archaeology, Cultural Heritage
	Philosophy and Theology	Ethical Principles, Philosophy and Ethics, Religion and Theology
Business and Management	Administration and Management	Administration and Management, Hospitality, Restaurants, and Banquet Services, Supply Chain, Travel, Tourism, and Recreational Activities
	Administrative	Office and Administrative Services, Occupational Health and Safety
	Customer and Personal Service	Customer Service, Personal Service and Humanization, Hairdressing and Beauty Treatments
	Economics and Accounting	Economics, Accounting and Taxes, Financial Management, Banking Administration, and Insurance
	Personnel and Human Resources	Human Talent and Personnel Management
	Sales and Marketing	Marketing and Advertising, Wholesale and Retail Sales
Communications	Communications and Media	Journalism, Communication, and Reporting, Audiovisual Techniques and Media Production
	Telecommunications	Telecommunications
Education and Training	Education and Training	Education and Training, Library Science, Information, and Archival Studies, Research Methodology, Physical Education
Engineering and Technology	Building and Construction	Construction and Civil Engineering, Blueprint Interpretation
	Computers and Electronics	Computers, Electronics, and Automation; ICT (Information and Communication Technology) Management; Software Development and Application Analysis, Network and Database Design and Administration
	Design	Design, Simulation and Design Tools, Architecture and Urban Planning, Fashion, and Interior Design, Fashion and Beauty
	Engineering and Technology	Engineering and Technology, Environmental Protection and Management Technologies, Chemical Engineering and Processes, Mining

	Mechanical	and Extraction, Materials (Glass, Paper, Plastic, and Wood), Electricity and Energy, Industrial. Mechanics and Metalworking, Maintenance, Mechanics and Engineering of Motor Vehicles, Ships, and Aircraft
Health Services	Medicine and Dentistry	First aid, Medicine, Anatomy, Comprehensive health care, Veterinary medicine, Medical diagnostic and treatment technology, Nursing and midwifery, Alternative and complementary medicines and therapies, Surgical instrumentation, Dentistry and dental studies.
	Therapy and Counseling	Care, protection, and services for children, adolescents, and youth; Social work; Care for adults, elderly people with or without disabilities; Physiotherapy, speech therapy, occupational therapy, nutrition and related fields; Somatic techniques and body management.
Law and Public Safety	Law and Government	Laws and government, Political science and civic education, Disaster risk management, Law, Community sanitation.
	Public Safety and Security	Public security.
Manufacturing and Production	Food Production	Horticulture (techniques for orchards, greenhouses, nurseries, and gardens), Food processing, Agricultural and livestock production, Aquaculture and fishing, Food handling.
	Production and Processing	Production and processing, Textile products (clothing, footwear, and leather goods).
Mathematics and Science	Biology	Environmental sciences, Natural environment and wildlife, Natural heritage, Forestry and forest production.
	Chemistry	Chemistry, Biochemistry, Pharmacy.
	Geography	Geography and territory, Earth sciences.
	Mathematics	Mathematics
	Physics	Physics
	Psychology	Psychology
	Sociology and Anthropology	Sociology, anthropology, and cultural studies.
Transportation	Transportation	Transportation services.
CUOC knowledge not classified under ONET.		Thanatopraxy, Domestic services, Funeral services.

Notes: This classification of knowledge areas was developed using the CUOC (Clasificación Única de Ocupaciones de Colombia) system, adapted to the categories of the ONET system. Due to certain knowledge areas not being directly classifiable under ONET, they are included in a separate group. *Source*: Classification of knowledge areas adapted from the CUOC (Clasificación Única de Ocupaciones de Colombia) system to the O*NET system, developed by the authors.

Figure E1

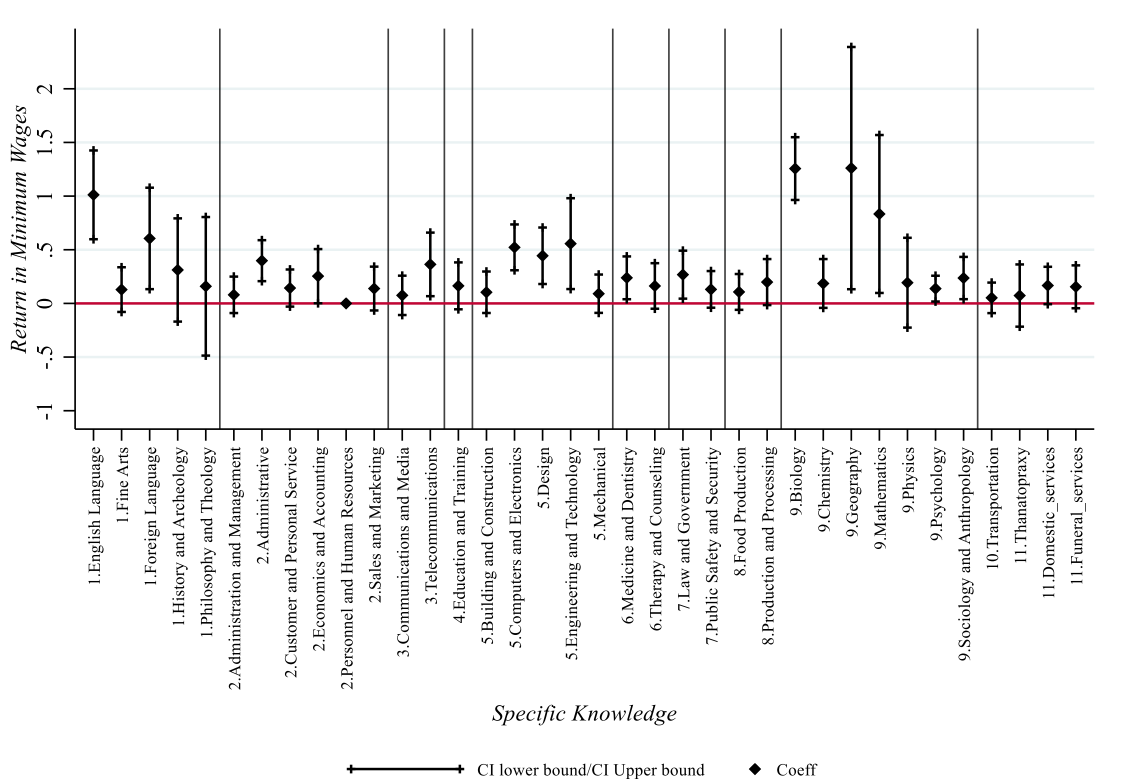
Robustness Check (CUOC): Returns to Specific Skills by Groups



Notes. Figure E1 presents the graphical representation of the β coefficients for each of the 35 O*NET skills and 6 CUOC skills, obtained by estimating an equation similar to Equation 5.1. Instead of using the aggregated skills vector, we employ the specific skills vector. This regression controls for education level, years of experience, and fixed effects for city, period, occupation, and firm. Confidence intervals are constructed at the 95% level.

Figure E2.

Robustness Check (CUOC): Returns to Specific Knowledge by Groups



Notes. Figure E2 shows the graphical representation of the β coefficients for each of the 33 knowledge O*NET and 3 CUOC knowledge categories, obtained from estimating an equation like Equation 5.2 (instead of using the vector of aggregated knowledge, we employ the vector of specific categories). In this regression, we control by education level, years of experience, and fixed effects of city, period, occupation, and firm. Confidence intervals are constructed at 95%.