Working Paper Series

#2025-024

Human Capital and Occupational Mobility: The Case of Brazil

Renata Mayer Gukovas

Published 28 October 2025

DOI: https://www.doi.org/10.53330/MYOQ8549

Maastricht Economic and social Research institute on Innovation and Technology (UNU-MERIT)
email: **info@merit.unu.edu** | website: **http://www.merit.unu.edu**

Boschstraat 24, 6211 AX Maastricht, The Netherlands
Tel: (31) (43) 388 44 00

UNU-MERIT Working Papers
ISSN 1871-9872

Maastricht Economic and social Research Institute on Innovation and Technology UNU-MERIT

UNU-MERIT Working Papers intend to disseminate preliminary results of research carried out at UNU-MERIT to stimulate discussion on the issues raised.

Human Capital and Occupational Mobility: The Case of Brazil

Renata Mayer Gukovas *[1,2]

[1]UNU-MERIT, The Netherlands
[2]Maastricht University, The Netherlands

Abstract

Skill transferability is essential for occupational mobility and adapting to external shocks, yet detailed data on workers' skills is often scarce and costly to collect, specially in developing countries. This paper explores whether the American O*NET database, one of the most popular databases used in studies on occupational skills, can provide relevant insights in the Brazilian context. By complementing it with unique longitudinal administrative Brazilian employer-employee data, this paper validates the application of skill similarity measures in a country lacking data on workers' skills. Applying the same methodology produces results in Brazil comparable to those in the American literature, and further refining the similarity measure increases its power to explain occupational mobility in the country. Then, this paper uses the validated measure to analyze the role of skill transferability in explaining Brazilian mobility patterns, exploring heterogeneity across different genders, age groups, and education levels.

JEL: J24, J31, J62, O15, O54

Keywords: Skills transferability; Occupational mobility

*Corresponding author: mayer@merit.unu.edu. This paper benefited from comments from Fabiana Visentin, Didier Fourage, Jessica Menezes, Alexandre Pecora, and the participants of the Learning & Work seminar at Maastricht University, the UNU-MERIT 2021 internal conference.

1 Introduction

Human capital, or the skills, knowledge, and abilities acquired through education, training, and work experience, plays a crucial role in determining labor market outcomes. Occupational mobility, or the ability of workers to transition between jobs or industries, depends on the extent to which their skills and knowledge are similar to the requirements of new positions. In other words, workers with transferable human capital are better positioned to adapt to labor market shocks (Nedelkoska et al., 2015), while skill mismatches are often linked to reduced productivity and wage losses (Neffke et al., 2024). In economies with high turnover rates like Brazil, where the average formal employment turnover reaches over six percent per month some sectors, understanding the role of human capital in enabling occupational mobility, potentially facilitating labor intermediation, is especially critical.

However, while the relationship between human capital and occupational mobility is well-established in theory (Ng et al., 2007), empirical evidence is generally limited to data-rich contexts. Measuring occupations relatedness relies on rare and complex data such as occupational descriptions (Poletaev and Robinson, 2008), direct skills test (Desjardins and Rubenson, 2011), or data mining in job postings (Anderson, 2017). Alternatively, co-occurrence of occupations (Penrose, 2009) or excess labor flows (Neffke and Henning, 2013; Neffke et al., 2017) utilize more widely available data such as labor force surveys, however only capture the current conditions of the labor market and provide limited insights into the dimension or nature of skill gaps.

This paper investigates whether human capital, as proxied by skill similarity measure derived from the U.S.-based O*NET database, can explain labor mobility patterns in Brazil. Using a gravitational model, it combines the O*NET dataset, which characterizes occupations across dimensions such as knowledge, skills, and abilities, with the RAIS administrative database, which contains all formal employment contracts in Brazil. This approach allows for a comprehensive and generalizable analysis of occupational mobility in a developing economy.

This paper makes three different contributions to the literature of human capital and labor mobility. First, it confirms the applicability of U.S.-based skill similarity measures in other countries by replicating methodologies originally developed and tested with American labor data (Nedelkoska et al., 2018). Using the Brazilian labor market as a case study, the findings show consistent results, suggesting that the measure of proximity between occupations derived from American occupation descriptions can serve as a good proxy for occupational similarity in other countries. This result validates a methodology with very low data requirements, that can provide at least initial insights to labor markets in countries where detailed skills data is not readily available. Second, this study also refines the measure of skills similarity to include the complexity of skills, achieving better results in predicting mobility patterns than the original metric. Lastly, it increases the understanding of the heterogeneous impact of human capital in occupational mobility by exploring how it explains the transitions among workers of different profiles.

The results corroborate that stronger similarity in skill content between occupations significantly explains workers transitions. Using a Jackard Similarity Index (JSI) based on knowledge, skills, and abilities to measure occupations similarity, this study finds that workers are 1.94 times more likely to move to occupations in the 95th percentile of the similarity compared to those in the 5th percentile. Refining the measures to consider skill complexity, the relationship is even stronger relationship, with workers being 4.98 times more likely to move between more similar occupations. Finally, when analyzing the results by demographic groups and educational level, the likelihood to transition between similar occupations is particularly strong among men, workers with higher levels of skills, and those between the ages of 25 and 40.

This paper is divided into five sections as follows. After this introduction, section 2 provides a summary of the literature and section 3 describes the two databases used in the analysis. section 4 explains in more detail the methodology and section 5 show the results. Finally, section 6 concludes.

2 Literature review

The study of human capital and labor market outcomes places significant emphasis on the transferability of skills between jobs. Workers who are able to leverage their prior experiences and transfer their skills can boost their productivity in subsequent positions, or attenuate the negative consequences of displacements. On the other hand, skills mismatch are linked with expected lower productivity, resulting in forgone wages. Previous studies have shown that workers are more likely to move between occupations with similar task content, and the distance between occupations declines with labor market experience (Nedelkoska et al., 2018; ?). For instance, Neffke et al. (2024) found that despite an initial reduction in wages, workers who move to a similar occupation after displacement recover the losses in about four years, as opposed to those who move to a more distant occupation. In addition, wage losses demonstrate a stronger correlation with the transition of skill portfolios rather than the transition of industry or occupation codes themselves (Poletaev and Robinson, 2008).

Measuring the transferability of skills between occupations is a complex task that often requires data that is not readily available for every country and is usually expensive to collect (Nedelkoska et al., 2019). Occupational descriptions (Nedelkoska et al., 2018; Poletaev and Robinson, 2008; ?), direct skills test (Desjardins and Rubenson, 2011), or data mining in job postings (Anderson, 2017; Atalay et al., 2020; Markow et al., 2017; Turrell et al., 2018) allow researchers to calculate the direct skills' overlap or mismatch between any pair of occupations. These methods can provide valuable insights into workers' potential career transitions and their vulnerability in the labor market (Nedelkoska et al., 2018). However, these methods often require extensive surveys or rely on the availability and accuracy of job postings.

Alternatively, economists have used co-occurrence of occupations (Penrose, 2009; Robins and Wiersema, 1995; Wernerfelt, 1984) or excess labor flows (Neffke and Henning, 2013; Neffke et al., 2017; Shaw, 1987) as measures of skill transferability. These methods utilize administrative or survey-based labor market data, which are widely available. These approaches nonetheless have the drawback of only capturing the current conditions of the labor market and provide limited insights into the dimension or nature of skill gaps.

A combined use of two alternative approaches of occupation relatedness – skill similarity and excess labor flows - to validate the measure of skill relatedness was first proposed by Nedelkoska et al. (2018). Their objective was to utilize the results to analyze workers' vulnerability to technological change. In their paper, the authors initially examine if a measure of similarity, a Jackard Similarity Index based either on tasks, tools, or knowledge, skills, and abilities (KSA) explains the excess labor flows observed in the U.S. Current Population Survey using a gravitation model. Their findings indicate that similarities in job tasks, KSAs, and tools among occupations exhibit a strong predictive capacity for occupational mobility. This paper makes use of this same methodology, both in terms of the measure of skill similarity and the use of the gravitation model, to validate the use of the similarity measure built using O*NET to a country where detailed information on skills or alternative measures of occupation similarity is not available. [1]

After having validated the use of the similarity measure to the Brazilian context, this paper also investigates the heterogeneity of the relation between occupation similarity and mobility across the workers' characteristics. Prior research has shown that mobility patterns can differ significantly by age, gender, and education level. For example, **?** observed that the relevance of task similarity in occupational transitions increases with labor market experience, while Nedelkoska et al. (2018) found that skill relatedness plays a greater role for workers with higher education levels. By examining these dynamics in the Brazilian labor market, this paper provides insights into how human capital influences mobility in diverse demographic contexts.

[1] The O*NET database has already been used to describe the Brazilian labor market, for instance in **?**. However, it simply applies the data after using a crossover. Maciente (2012) builds the crossover, verifying carefully if the skills reflect the content. This paper adds an additional validation by linking the characteristics with occupational transitions.

3 Description of the databases

This paper uses two databases: the American O*NET database and the Brazilian RAIS, to investigate the role of human capital in occupation mobility. The O*NET Database provides a detailed description of various occupations. This study uses the skills, knowledge, and abilities associated with each occupation as a proxy for workers' human capital. These descriptions enables the construction of a measure of similarity between occupations. RAIS, on the other hand, is a matched employer-employee administrative dataset from Brazil, that allows us to follow workers through different contracts, therefore observing their occupational mobility. Both datasets are matched using a crosswalk between the occupational codes. This section provides a comprehensive overview of each database and the matching process employed.

3.1 O*NET

To characterize the skills contents of the worker's current occupation, this paper uses O*NET (USDOL/ETA 2019) (Hilton and Tippins, 2010). The O*NET database is an American database that aims to explain the anatomy of 970 occupations. It associates in detail a series of knowledge topics, skills, abilities, tasks, tools, work styles, etc, with each occupation in the Standard Occupation Code (SOC). It was first built based on the Dictionary of Occupational Titles (DOT) description and is updated through surveys periodically, with about 100 occupations updated per year. It is currently in its 24th version. In each round of updates, researchers interview workers of a given occupation, supervisors, and occupational experts and ask them to describe what they do and how often, what they need to know, and how crucial it is for their job. This database aims to inform workers aiming at labor market transitions (Peterson et al., 1999) and is shown to help employers identify useful employee selection criteria (Jeanneret and Strong, 2003); therefore, it is an excellent instrument for the purpose of this study.

This study focuses on the knowledge, skills, and abilities (KSA) dimension of the O*NET database, which Nedelkoska et al. (2018) classify as the basis of the production process – a worker applies KSA to execute job tasks, using tools or not, in order to produce products and services. These tend to be more general and applicable to different contexts. Each of the three dimensions is broken down into items: knowledge (33), skills (35), and abilities (52), adding up to 120 items, such as economics and accounting (knowledge), active learning (skill), and finger dexterity (ability). [2] For each of the items, specialists are asked to specify what complexity level is required in the occupation, on a scale ranging from 0 to 7. [3] Finally, they are asked to say with what importance they are used on the job in a scale from 1 to 5 (range from "Not Important" (1) to "Extremely Important" (5)). The answers are compiled, and for each occupation, the database contains the average level and importance for each item, and their standard deviation. Annex 1 shows the list of KSA, and lists those that are most and least used across the different occupations.

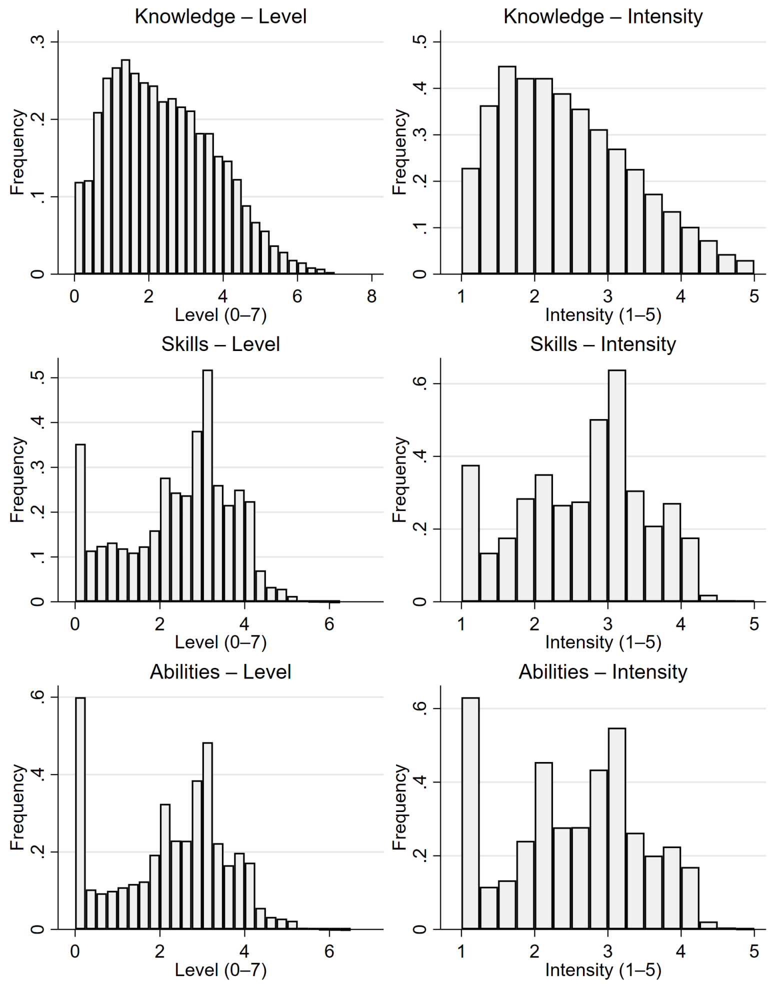
Figure 1 displays the distribution of the level and the importance in which the different items of knowledge, skills and abilities are used across different occupations. The distribution of the importance and the level of how knowledge, skills and abilities are employed follow a similar pattern, as it can be seen compared the graphs on the top and the bottom for each subset, and both measures are highly correlated. Taking abilities, for instance, there are several pairs occupation and ability that display both the level and importance zero - meaning that the specific ability is not used in that given occupation. Disregarding the values zero, the distribution of the level and importance the different ability items are employed at different occupations follows a bell shaped curve, with the second most common value being 3 (mid point of the alternatives). This means that the abilities are often required at an intermediate level. There are fewer abilities that are required rarely or all the time, also they are less required at a very basic or very advanced level. Skills follow a similar distribution. For the items in Knowledge, on the other hand, it is more common that they are required

[2] More examples in Appendix A.

[3] There are very few occupations that use level 7, as shown in **??**

at a basic level, and rarely at occupations, while they are required less frequently at a more advanced level or in most of the time individuals in those occupations are at work. [4]

Figure 1: Distribution of the level and importance of Knowledge, Skill, and Abilities across occupations



The figure shows the distribution of the values of importance and level that different dimensions of Knowledge, Skills, and Abilities are used across occupations.

[4]In addition, some items across knowledge, skills and abilities are highly correlated between themselves - they usually appear together at the same level or importance in different occupations. For instance, "Critical Thinking" and "Complex Problem Solving." Using both items to compute a measure of similarity between occupations would inflate the measure, so only one of each is considered. The results don't vary according to which item is chosen.

3.2 RAIS

Relação Anual de Informaços Sociais (RAIS - Annual Register of Social Information) is a yearly database that matches information about employees and employers, containing all formal contracts in Brazil between the years 1985 and 2019 [5]. It is collected by the Secretary of Labor under the Ministry of Economy (former Ministry of Labor and Employment – MTE), based on a form filled by formal companies [6]. Its purpose is threefold: to provide information to manage labor activities in the country, such as paying labor benefits to workers or monitoring firm compliance with labor regulation; to provide information for official labor statistics; and to provide information on the labor market for government institutions. Although its structure has changed over time, since 2003, it has maintained mostly the same variables[7]. It consists of two databases per year – one containing all the labor contracts, and the second with information about the firms and has unique identifiers for firms, establishments, and workers, allowing them to be matched and followed over time.

RAIS contains information about the workers, the contracts, and the companies. For the companies, this database contains their fiscal identification number (which can identify the company and the establishment), sector, legal nature, and size. The information about the worker includes their fiscal identification number, gender, date of birth, school level, race, and disability status. The information about the contracts is more comprehensive, including the municipality of the contract, whether the contract was active on December 31st, the type of contract (temporary, statutory, regular formal contract, etc), how the worker was hired, the date they were hired if they were hired that year, months in the current contract, occupation code, average wage, wage in December, contractual wage, weekly hours in the contract, number of days the worker was on non-annual leave, the reason for the leave, the month the contract was terminated, and reason for the termination.

[5] Including all workers hired under the Consolidated Labor Laws (Consolidação das Leis do Trabalho - CLT), civil servants, and casual workers with contracts accorting to the Law 12.815 from June 5th, 2013.

[6] All companies with a fiscal number registered in the national registry of companies - Cadastro Nacional da Pessoa Jurídica - CNPJ

[7] RAIS data prior to 2003 is excluded due to differences in occupational classification systems.

To observe the occupational mobility, first I build a panel of occupation transitions - considering for each worker, the pair of occupations in subsequent contracts [8]. In order to identify the sequence of contracts held by workers in the RAIS dataset, firstly, two distinct periods are considered: 2004-2008 and 2009-2012 [9]. Only workers who have been in a particular occupation for more than 6 months and have been hired for more than 30 hours per week are included in the analysis (assuming that workers need to spend a certain amount of time in that occupation to have the skills associated with the occupation). Transitions into public sector employment are disregarded, as the selection process for such positions is based on public competition[10] and not necessarily linked to specific occupational skills (although transitions from public to private employment are considered). In cases where workers have multiple simultaneous contracts, only the one with the highest pay is taken into account. To construct the transition dyads (pairs of origin and destination occupations), contracts are ordered based on the date of admission, and a transition is considered to occur when a worker starts a new contract, even if there are overlaps in time. As a robustness check, transitions involving workers in establishments with fewer than 30 employees are excluded, as workers in small companies and establishments may have responsibilities beyond those typical of their occupation. A transition is considered to have taken place in the year of hiring, and if a worker remains out of the formal labor market for more than 2 years, that transition is disregarded. Based on these criteria, a total of 184,039 dyads were observed in the first period (2004-2008), and 207,426 dyads in the second period (2009-2012). Details are reported in Appendix B. Approximately 20 percent of the dyads in both periods involve only one worker transitioning between them. Finally, 33.8% of the dyads (70,269) account for 99% of the observed transitions.

[8] Or registered in subsequent years, in case the worker is registered with a different occupation within the same contract.

[9] The sample is broken down for a few reasons: (i) to reflect new updates of the O*NET database to be used, (ii) to measure the evolution of the relation between occupations, and (iii) for computational purposes.

[10] In the format of exams

3.3 Matching O*NET database and RAIS

To match RAIS with the O*NET database, I use the fact that both the Brazilian Code of Occupations (CBO) and SOC, used in the O*NET database, were built using a similar pyramid principle (Maciente, 2013). [11] Table 1 details the results of the crosswalk between the two occupation codes. 234 SOC occupations (or 24.2% of the SOC codes) don't have an equivalent in the Brazilian labor market. On the other hand, some of the occupations in the Brazilian context are not available in the O*NET database, such as senators, political party managers, or tribe leaders (218 in total, which represents 9% of the number of occupations form CBO at 6-digits)[12]. CBO is also significantly more disaggregated than SOC when using all 6 digits (O*NET has 968 SOC occupations, while CBO has 2,422 occupations at 6 digits). Besides the occupations that are not available in O*NET, there are many CBO occupations mapped to repeated SOC codes (one SOC for multiple CBO) [13]. Because of that, using the data disaggregated at the 6-digit level would result in several occupations having the exact same characteristics, therefore, after merging both codes at the 6-digit level, the analysis is conducted at the 4 digit level taking an average of the level and importance of each dimension within the same general occupation (with 596 total occupations, and 583 occupations, or 97.8% of the occupations at 4-digits with matches to SOC). On the other hand, even at the 6-digits, 459 CBO occupations can be matched to more than one occupation in SOC (representing 18.9% of the CBO codes at 6-digits). For those, for each dimension of KSA, I take an average of the level and importance each dimension is used. The final sample consists of 583 matched occupations, and 339,889 potential dyads (pairs of

[11] CBO was built using the same pyramid-like principle as SOC, with the first digit representing the broad group (e.g. Professionals of science and arts), the second the main subgroup (eg. Professionals of exact sciences, physics, and engineers), the third digit defines the subgroup (e.g. Engineers, architects, and similar occupations), the fourth the general occupation (e.g. Mechanical engineer), and the fifth and sixth the specific occupation (e.g. Automotive mechanical engineer). In addition, the broad groups in both codes are generally the same, allowing for the construction of a simple crosswalk (Maciente 2013)

[12] About 20% of the transitions between occupations are either to or from one of the unmatched occupations. This is one limitation of this method

[13] For example, CBO 2144 groups all mechanical engineers, which encompasses 7 different occupations, including automotive mechanical engineer, industrial mechanical engineer, aeronautic engineer.

occupations).

Table 1: Crosswalk between occupational codes in O*NET and RAIS

Number of CBO at 6 digits	2,422
Number of CBO at 4 digits	596
Number of SOC occupations in O*NET	968
Number of CBO matched to more than one SOC	459
Share of CBO matched to more than one SOC (on total number of CBO)	18.9%
Number of unmatched CBO (at 6-digit level)	218
Share of unmatched CBO (at 6-digit level) (on total number of CBO)	9.0%
Number of unmatched SOC occupations	234
Share of unmatched SOC occupations (on total number of SOC)	24.2%
Number of matched CBO (at 6-digit level)	2,204
Share of matched CBO (at 6-digit level) (on total number of CBO)	91.0%
Number of matched CBO (at 4-digit level)	583
Share of matched CBO (at 4-digit level) (on total number of CBO)	97.8%
Final sample (number of dyads of CBO at 4 digits matched to SOC in O*NET)	339,889

Table shows the number of occupations matched and unmatched when using the crosswalk between CBO (RAIS) and SOC (O*NET).

4 Methodology

In order to understand if workers are more likely to transition between similar occupations, I need to have a measure of similarity. There are several methods to measure the similarity between two objects, and the Jaccard Similarity Index (JSI) is among the best measures of similarity when applied to the O*NET data, with the additional advantage of having a simple interpretation (Yildirim and Coscia, 2014). [14] Nedelkoska et al. (2018) propose three alternatives for using the JSI: build the index based on tasks, on tools, or on knowledge, skills, and abilities (KSA). In each of the approaches, for each pair of occupations j and k (dyads), they take the set of tasks (or tools, or KSA) required in both occupations ($HC_j \cap HC_k$) and divide by the number of those required in at least one of them ($HC_j \cup HC_k$).

[14] To determine the performance of different similarity measures, they calculate receiver operating characteristics (ROC) curves for each of the measures, and AUC values, which are the integrals below the curves. A random prediction has an AUC value of 0.5, and values closer to 1 indicate a high performance. JSI has the fourth highest AUC among the measures they observe (0.80507), much higher than the Euclidian distance (0.68452), and the top three are probability-based methods with a more complex interpretation.

$$JSI_{jk} = \frac{HC_j \cap HC_k}{HC_j \cup HC_k} \tag{1}$$

This paper begins using the same approach used in Nedelkoska et al. (2018) to measure the JSI based on KSA. As in Nedelkoska et al. (2018), to calculate JSI_1 and JSI_2 it considers each dimension of KSA as a different unit, meaning that each occupation is characterized by a KSA vector containing 123 dimensions. Each dimension gets a dummy value of 1 if enough respondents say workers in that occupation apply it in their daily activities. [15] [16] While JSI_1 uses all the dimensions listed as used by each occupation, JSI_2 disregards those used with an importance lower than 2,5, similarly to Nedelkoska et al. (2018). The rationale is that when the skill is used at a lower importance in the occupation, the worker's might not be so much familiar with it, and may not be able to apply it to other contexts. By doing this, the JSI average decreases from 0.6904 to 0.4446, and the variance decreases from 0.2023 to 0.1705 (Table 2).

However, in addition to those measures, this paper also proposes including the other dimension associated with each skill in the O*NET database - the level in which each skill is used in each occupation. To facilitate the answers and anchor the levels associated with each KSA, the O*NET survey questionnaire provides examples of activities performed at different levels of complexity. For example, respondents are asked to consider 'Mathematical reasoning' at level 1 if the worker can 'Determine how much 10 oranges will cost when they are priced at 2 for 20 cents', at level 4 if they can 'Decide how to calculate profits to determine the amounts of yearly bonuses' and at level 7 if they can "Determine the mathematics required to simulate a spacecraft landing on the moon'. This scale provides a clear anchor that practically distinguishes the skills at different levels. This papers propose to break each skill according to their level based on the assumption that considering occupations that employ KSA at different levels to be similar would underestimate the real distance between them.

[15] O*NET recommends dropping dimensions that very few respondents say it is applied to the job.

[16] In addition, we disregard KSA used by more than 95 percent of the occupations, as Nedelkoska proposed, to reduce the noise caused by frequently used KSA.

In practice, the paper explores breaking down the levels of each KSA into three or six groups to calculate the JSI [17]. The additional JSI_3 and JSI_4 break each dimension in three, adding to a total of 360 subdimensions. They are calculated as follows: if it is used at levels between 1 and 2, I consider that only the first subdimension is one; if it's used at levels 3 and 4, I consider that both first and second subdimensions are one; and finally, if it's used at levels 5, 6 or 7, all subdimensions are one. The difference between JSI_3 and JSI_4 is similar to JSI_1 and JSI_2, where JSI_3 considers all subdimensions, and JSI_4 only those that are used with an importance higher than 2,5. [18] Finally, JSI_5 and JSI_6 follow a similar approach, considering the following breakdowns for the level at which the skill is used, 1, 2, 3, 4, 5, and 6-7 (there are very few skills used at level 7). Table 2 describes the different JSI using each of the definitions.

Table 2: Description of the different measures of similarity across occupations - Jackard Similarity Indexes (JSI)

	JSI 1	JSI 2	JSI 3	JSI 4	JSI 5	JSI 6
Number of dimensions	123	123	369	369	738	738
Average	0.6904	0.4446	0.4868	0.4401	0.4540	0.4338
SD	0.2023	0.1705	0.1457	0.1588	0.1457	0.1536
5th percentile	0.3158	0.1500	0.2366	0.1667	0.2021	0.1702
Median	0.7059	0.4507	0.4914	0.4437	0.4620	0.4405
95th percentile	1.0000	0.7097	0.7128	0.6887	0.6756	0.6696

Table shows the main attributes of the different JSI calculated.

By construction, measures of similarity range from zero to one, where zero would mean that the occupations have no skills overlap, and one means that both occupations use exactly the same KSA. When considering all KSA available in O*NET, regardless of their importance or level (JSI_1), [19] the average similarity between occupations is 0.6904, and the standard deviation is 0.2023. For that measure, over 5% of occupations employ exactly the same KSA. When discarding KSA that are used with an importance lower than 2.5 (JSI_2), the

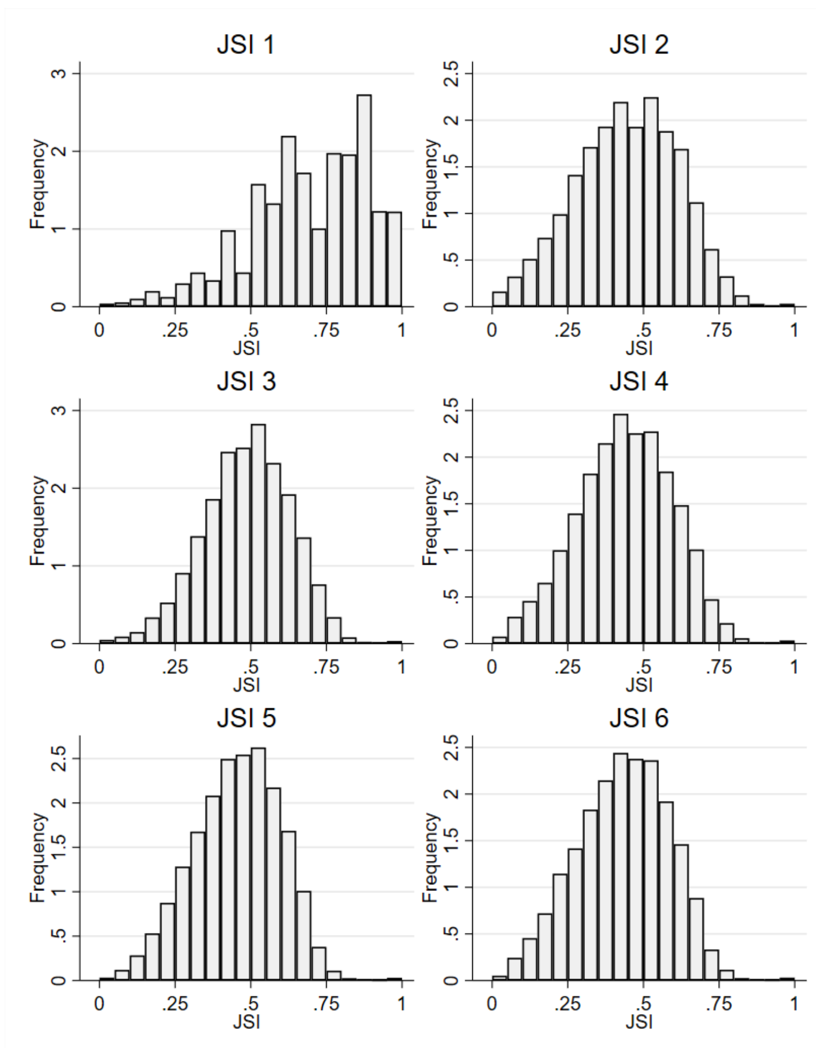
[17] Each KSA in the O*NET database is characterized by a level that ranges from 1 to 7.

[18] In case a dimension is used at a higher level of complexity (between 3 and 7), but at a lower importance, I still consider the lower subdimension dummy as one, assuming that workers have the skill at a lower level even if they only use it less frequently at a more complex level.

[19] Except for those that are used by more than 95 percent of the occupations.

average similarity falls to 0.4446, and there are fewer occupations with a complete overlap of skills. If we break down KSA into 3 groups according to the level, the average is 0.4868, and into 6 groups, the average is 0.4540. Figure 2 shows the distribution of the different JSI measures. JSI_1 shows a different pattern due to the large number of skills that are used by most occupations with a low level and importance that are not considered in the remaining measures; however, the others show a bell-shaped distribution.

Figure 2: Distribution of the six proposed JSI for KSA



The figure shows the distribution of the values of the six proposed Jackard Similarity Indexes calculated for the Brazilian occupations at four digits.

To answer the question of how human capital influences labor mobility in Brazil, it's necessary to measure how JSI_{jk} explains the number of workers transitioning between occupations j and k, T_{jk}. The assumption is that the more similar two occupations are in terms of human capital required, the number of workers transitioning between one and the other would be higher than what would be expected based on occupational growth patterns. The literature has used a gravitational model taken from international trade (Baier et al., 2018; Mátyás, 1997; Shepherd et al., 2013) to address this question (Anderson, 2017; Nedelkoska et al., 2018). The construction of the model allows it to consider occupation-specific characteristics and trends. For instance, similarly to how the size of planets influence gravity, some occupations might have a pull effect during periods of growth, attracting workers from the entire labor market, and push effect when the overall demand for that occupation decreases. By taking that into the consideration, the model isolates the effect of the distance between two points to the aggregate flow of workers between them. This characteristic of the model makes it ideal for the purpose of this paper, allowing to measure if the similarity between their skill content increases the flow of workers between them while taking into account general labor market trends. The model takes shape as Equation 2, and it is estimated at the level of pairs of occupations, k and j.

$$T_{jkt} = G \frac{O_{kt}^{\beta_2} O_{jt}^{\beta_3}}{JSI_{kj}^{\beta_1}} \nu_{kjt} \tag{2}$$

Where G is a constant, O_{kt} and O_{jt} are economic factors associated, respectively, with occupation k and j in time t, and υ_{kjt} an error term with mean 1. The idea is that, similarly to gravity, economic factors in occupations k and j would pull (or push) workers based on the economic factors associated with the occupations in period t, O_{kt} and O_{jt}. The value of β_1 would indicate if the measure of similarity is relevant to explain these transitions. If β_1 equals zero, the denominator turns to one, regardless of the value of JSI_{kj}, meaning that the degree of similarity between occupations actually does not influence the number of transitions between occupations k and j. By definition, JSI_{kj} is a positive value between 0 and 1, so the higher β_1 is, the smaller the denominator, meaning that the number of workers transitioning between k and j would be higher than what would be expected given the occupation specific characteristics O_{kt} and O_{jt}, and the higher the JSI_{kj}, the higher the number of transitions.

Taking the log on both sides, the equation above translates into Equation 3 which can be estimated using a simple linear regression .

$$\ln(T_{jkt}) = \alpha + \beta_1 \ln(JSI_{kj}) + \beta_2 O_{kt} + \beta_3 O_{jt} + \varepsilon_{kjt} \tag{3}$$

$\beta_2 O_{kt}$ and $\beta_3 O_{jt}$ can be translated into occupation time dummies that control for the economic factors associated with the occupations in that period, such as occupational growth, wage composition, etc. As described in the previous paragraph, the higher β_1 is, the stronger the effect of the similarity between occupations. To interpret the coefficients, a worker is $(e^{\beta} - 1)$ times more likely to move between two occupations in the 95^{th} percentile of the JSI distribution than between two occupations in the 5^{th} percentile.[20]

5 Results

The results of the study demonstrate a strong correlation between the various Jaccard Similarity Index (JSI) measures, constructed using the Knowledge, Skills, and Abilities (KSA) data from the O*NET database, and the occupational mobility of workers in the RAIS dataset. The regression results are presented in Table 3, while Table 4 displays the transformed coefficients. The findings reveal that workers in the 95^{th} percentile of KSA similarity, as measured by JSI 1, are 1.37 times more likely to transition between occupations compared to those in the 5^{th} percentile. Moreover, when excluding KSA items with an importance lower than 2.5, this likelihood increases to 1.94, which closely aligns with the value observed in the US labor market (1.911) according to the study by Nedelkoska et al. (2018). This confirms the relevance of the measure of skill similarity in the Brazilian context [21].

[20] Therefore, the expression to interpret the results of the $(e^{\beta} - 1) * \Delta JSI_{95^{th}-5^{th}}$

[21] It is important to note that the skills used on each occupation could still be different, but the measure of how they relate to each other in terms of skills is likely to be similar

Table 5 breaks down the transitions into two periods (between 2004-2008 and 2009-2012), and for all the JSI, the results are stronger in the second period [22]. A similar pattern is also observed in the US labor market.

Table 3: Human Capital and Occupation mobility

	JSI 1	JSI 2	JSI 3	JSI 4	JSI 5	JSI 6
Ln(JSI-KSA)	1.098***	1.495***	2.438***	1.761***	2.220***	1.861***
	(36.98)	(115.61)	(120.99)	(120.51)	(140.33)	(134.02)
N	1.52e+05	1.52e+05	1.52e+05	1.52e+05	1.52e+05	1.52e+05
R2	0.462	0.559	0.576	0.573	0.592	0.586
Adjusted R2	0.453	0.552	0.569	0.567	0.586	0.580

Standard errors are clustered by occupation of origin and occupation of destination. Models include occupation year dummies O_jt. t statistics in parentheses. Significant at $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$

Table 4: Human Capital and Occupation mobility - transformed betas

	JSI 1	JSI 2	JSI 3	JSI 4	JSI 5	JSI 6
Ln(JSI-KSA)	1.37	1.94	4.98	2.52	3.89	2.71

Table shows the transformed betas using the equation $(e^{\beta} - 1) * \Delta JSI_{95^{th}-5^{th}}$.

However, when further refining the analysis by incorporating information on the level at which each Knowledge, Skills, and Abilities (KSA) is applied to each occupation, the results become even more robust. Using the JSI measure that breaks down KSA into three different levels (JSI 3), the findings indicate that workers in the 95^{th} percentile of the similarity distribution are nearly 5 times more likely to transition between occupations compared to those in the 5^{th} percentile. However, when breaking down the levels of KSA into 6 groups, the relationship weakens, reducing the likelihood to 3.89 times. This suggests that it may be more relevant to consider whether an occupation utilizes a KSA item in the middle or at the extreme ends of the distribution rather than the exact point within the distribution. Additionally, when the similarity measure considered the importance of each KSA (JSI 4 and 6), disregarding items that were used with an importance lower than 2,5, the results are

[22]To allow the model to capture different trends in the labor market as in Nedelkoska et al. (2018)

not as strong, potentially indicating a strong correlation between the level and importance of each KSA.

As a first robustness check, the transitions were broken down into two different periods. Table 5 show that the results still hold when dividing the sample into two different periods. Using all the different JSI, the coefficients were slightly larger for the second period. One possible explanation is that the relation between occupations in Brazil and in the United States became more similar over time, making the description of the occupations in the US more relevant to predicting workers' movements in Brazil. In addition, Table 5 indicates that there is still some gain in breaking down the JSI into 3 different levels of complexity (JSI 3 and JSI4), with JSI 3 showing the highest predictive power. JSI5 and JSI 6 are still better than JSI 1 and JSI2; however, they explain less the differences in the number of transitions between occupations.

Table 5: Human Capital and Occupation mobility - Two periods

Ln. (Total numb. of transitions)	JSI 1	JSI 2	JSI 3	JSI 4	JSI 5	JSI 6
Period 1 * Ln(JSI-KSA)	1.088***	1.469***	2.399***	1.732***	2.185***	1.832***
	(35.73)	(110.04)	(116.11)	(116.15)	(134.35)	(128.94)
Period 2 * Ln(JSI-KSA)	1.108***	1.522***	2.478***	1.792***	2.256***	1.892***
	(36.00)	(110.39)	(116.52)	(115.88)	(134.66)	(128.49)
N	1.52e+05	1.52e+05	1.52e+05	1.52e+05	1.52e+05	1.52e+05
R2	0.462	0.559	0.576	0.573	0.592	0.586
Adjusted R2	0.453	0.552	0.569	0.567	0.586	0.580

Standard errors are clustered by occupation of origin and occupation of destination. Models include occupation year dummies $O_j t$. Significant at $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$

Taking advantage of the extensive data available in RAIS, this study conducts various robustness checks to enhance the reliability of the findings. The outcomes of these checks are presented in Table 6, which examines different types of transitions using only JSI_3. Surprisingly, when excluding transitions into the public sector, the measure of similarity becomes less significant compared to considering all transitions.[23] Notably, workers who voluntarily leave their previous job to a private sector job (Voluntary transitions to private sector) ex-

[23] This is surprising because although there is a value attributed to the analysis of a candidate's CV and experience, the hiring process for public sector jobs mostly relies on broad exams that may not be directly linked to the skills associated with the occupations. It would be expected that transitions only to the private sector would give a higher coefficient to the JSI

hibit a higher tendency to move towards occupations that are more similar in nature. This finding implies that individuals who actively choose to change their occupation are more likely to seek opportunities that align closely with their existing skill set or experience.

Table 6: Human Capital and Occupation mobility - Robustness checks

	All transitions	All transitions to private sector	Voluntary transitions (VT) to private sector
Ln(JSI-KSA)	2.220***	2.205***	2.125***
	(140.33)	(138.67)	(137.33)
N	1.52e+05	1.51e+05	1.43e+05
R2	0.592	0.595	0.633
Adjusted R2	0.586	0.588	0.627

Standard errors are clustered by occupation of origin and occupation of destination. Models include occupation year dummies $O_j t$. Significant at $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$

5.1 Heterogeneity Analysis: age and gender

Table 7 presents the findings obtained by disaggregating workers based on their age and gender. When examining the breakdown by age, it is observed that the measure of similarity becomes less relevant for young workers, specifically those aged up to 25 years old, compared to workers between the ages of 25 and 40. Within this age range, workers in the 95^{th} percentile are 3.44 times more likely to transition between occupations than those in the 5^{th} percentile (against 2.26 for younger workers). However, this relationship weakens for older workers aged over 40 (3.15). Furthermore, the analysis reveals that skill similarity holds greater significance for men than for women. This gender disparity in the importance of skill similarity could be attributed to various factors, such as differences in occupational segregation, societal expectations, or variations in career aspirations. Understanding these gender-specific dynamics can provide valuable insights for designing targeted workforce development and training programs that address the unique needs and preferences of both men and women in the labor market.

These findings highlight the importance of age as a factor influencing the relevance of skill similarity in career transitions. It suggests that for young workers, other factors may play a more significant role in shaping their occupational choices, while for workers between

Table 7: Human Capital and Occupation Mobility - Age and gender

	VT - age < 25	VT - age $25 \leq . < 40$	VT - age ≥ 40	VT - male	VT - female
Ln(JSI-KSA)	1.753***	2.111***	2.036***	2.075***	1.862***
	(116.64)	(136.74)	(119.95)	(140.24)	(100.98)
N	1.11e+05	1.42e+05	1.22e+05	1.45e+05	78659
R2	0.698	0.621	0.558	0.632	0.643
Adjusted R2	0.692	0.615	0.550	0.626	0.633

Standard errors are clustered by occupation of origin and occupation of destination. Models include occupation year dummies O_jt. Significant at $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$

25 and 40, skill similarity becomes a more influential factor. This could be potentially linked to inertia, either because prime-age workers have more work experience, giving it more value during transitions vis a vis their formation, or because younger workers adapt more easily.

5.2 Heterogeneity Analysis: education and wage differences

Table 8 presents the findings obtained by dis-aggregating workers based on their education level and on whether there was a wage increase or decrease in the transition. The relevance of similarity between occupations is more pronounced for individuals who have completed education beyond the secondary level, compared to those with lower educational attainment. This observation suggests that workers with higher levels of education tend to possess more specialized skills in their respective occupations, which may require extensive training and are not easily acquired on the job in a new position. Consequently, the degree of skill similarity becomes a crucial factor in determining the career transitions of highly educated individuals. Additionally, the analysis reveals that there is no significant difference in the measure of similarity when considering transitions where workers experienced an increase in their wages versus those who observed a wage decrease. This suggests that wage changes do not significantly impact the relevance of skill similarity in determining occupational transitions.

Table 8: Human Capital and Occupation Mobility - Education and Wage Progression

	VT - unskilled	VT – skilled	VT - increase wage	VT - decrease wage
Ln(JSI-KSA)	1.421***	2.123***	2.054***	2.070***
	(54.88)	(137.95)	(127.91)	(135.67)
N	48941	1.44e+05	1.34e+05	1.37e+05
R2	0.596	0.634	0.619	0.627
Adjusted R2	0.580	0.629	0.612	0.621

Standard errors are clustered by occupation of origin and occupation of destination. Models include occupation year dummies O_jt. Significant at $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$

6 Conclusion

The main findings of this paper obtained by applying a skill similarity measure developed using data from the US labor market to the Brazilian context suggest that it is feasible to generalize occupation descriptions across different contexts. This generalizability is particularly significant considering the limited availability and high cost of country-specific skills surveys. By utilizing skill similarity between occupations, valuable insights can be derived to inform public policy decisions. The application of skill similarity analysis holds a large potential in predicting labor market dynamics following various shocks, evaluating the vulnerability of different segments of the workforce, and designing tailored public policies. These insights can contribute to the development of effective strategies that address the specific needs and challenges of the labor market, ultimately fostering economic resilience and inclusive growth.

In addition, it is evident that a more granular measure of skills, which takes into account the specific level at which they are applied, provides a more comprehensive understanding of workers' transitions in the labor market. By refining the similarity measure to incorporate the level of each Knowledge, Skills, and Abilities (KSA), the relevance of the similarity index is significantly enhanced. The inclusion of the level at which KSAs are utilized offers valuable insights into the intricacies of occupational transitions. This refined measure captures the nuanced differences in skill requirements across occupations, allowing for a more accurate assessment of the transferability and compatibility of skills between different job roles. Consequently, it enables a more precise prediction of workers' mobility patterns and facilitates the identification of potential career pathways.

Finally, the findings of this paper show that there is heterogeneity in how important the measure of distance is across individual characteristics. Men are more likely to transition to more similar occupations, as are workers between the ages of 25 and 40 and with higher levels of education. Surprisingly, the results don't differ significantly when looking into transitions where workers suffered a wage loss or gain.

Despite a few limitations, the results of this paper provide ample ground for policy discussions and research. The main limitation of this paper is that RAIS only covers the formal labor market in Brazil, a country where about 40 percent of workers work in the informal sector. Because of this, it ignores skills acquired in positions that are not observed in RAIS. However, this is similar to ignoring skills that work potentially acquired in training, which is also left out of other similar studies. One way this paper tries to address this limitation is to consider only highly skilled workers, with at least a college degree. These are less likely to work in the informal sector, with an informality rate of only 20 percent, and probably acquire most of their training in university. The results confirms that the skill content of occupations play a significant larger role in explaining their occupational mobility.

Nevertheless, shedding light on the movements in the labor market opens the possibility of conducting more research on the structure of the labor market and provides evidence for public policy at a very low cost. Results equivalent to those in this paper point to which segments of the population are better able to use their current skills to reposition themselves in the labor market. Using a validated distance metric, it is possible to build a network of occupations, identifying particular occupations that are far removed from the rest of the labor market, and that could be at greater risk if displaced. It makes it possible to identify workers who are particularly vulnerable to individual or structural shocks in the labor market. Governments can play also a role in fixing potential inefficiencies in public policy, identifying workers who are ready to reposition themselves in the labor market after displacement, and focusing on providing tools to the remaining workers to adapt to another career.

References

Anderson, K. A. (2017). Skill networks and measures of complex human capital. *Proceedings of the National Academy of Sciences 114*(48), 12720–12724.

Atalay, E., P. Phongthiengtham, S. Sotelo, and D. Tannenbaum (2020). The evolution of work in the united states. *American Economic Journal: Applied Economics 12*(2), 1–34.

Baier, S. L., A. Kerr, and Y. V. Yotov (2018). Gravity, distance, and international trade. In *Handbook of international trade and transportation*, pp. 15–78. Edward Elgar Publishing.

Desjardins, R. and K. Rubenson (2011). An analysis of skill mismatch using direct measures of skills. *OECD Education Working Papers* (63).

Hilton, M. L. and N. T. Tippins (2010). A database for a changing economy: Review of the occupational information network (o* net).

Jeanneret, P. R. and M. H. Strong (2003). Linking o* net job analysis information to job requirement predictors: An o* net application. *Personnel Psychology 56*(2), 465–492.

Maciente, A. N. (2012). A mensuração das competências cognitivas e técnicas das ocupações brasileiras.

Maciente, A. N. (2013). *The determinants of agglomeration in Brazil: input-output, labor and knowledge externalities*. University of Illinois at Urbana-Champaign.

Markow, W., S. Braganza, B. Taska, S. Miller, and D. Hughes (2017). The quant crunch: How the demand for data science skills is disrupting the job market. burning glass technologies.

Mátyás, L. (1997). Proper econometric specification of the gravity model. *The world economy 20*(3), 363–368.

Nedelkoska, L., D. Diodato, and F. Neffke (2018). Is our human capital general enough to withstand the current wave of technological change? *CID Research Fellow and Graduate Student Working Paper Series*.

Nedelkoska, L., F. Neffke, et al. (2019). Skill mismatch and skill transferability: review of concepts and measurements. *Papers in Evolutionary Economic Geography*, 19–21.

Nedelkoska, L., F. Neffke, and S. Wiederhold (2015). Skill mismatch and the costs of job displacement. In *Annual Meeting of the American Economic Association*.

Neffke, F. and M. Henning (2013). Skill relatedness and firm diversification. *Strategic Management Journal 34*(3), 297–316.

Neffke, F., L. Nedelkoska, and S. Wiederhold (2024). Skill mismatch and the costs of job displacement. *Research Policy 53*(2), 104933.

Neffke, F. M., A. Otto, and A. Weyh (2017). Inter-industry labor flows. *Journal of Economic Behavior & Organization 142*, 275–292.

Ng, T. W., K. L. Sorensen, L. T. Eby, and D. C. Feldman (2007). Determinants of job mobility: A theoretical integration and extension. *Journal of Occupational and Organizational Psychology 80*(3), 363–386.

Penrose, E. T. (2009). *The Theory of the Growth of the Firm*. Oxford university press.

Peterson, N. G., M. D. Mumford, W. C. Borman, P. Jeanneret, and E. A. Fleishman (1999). *An occupational information system for the 21st century: The development of O* NET.* American Psychological Association.

Poletaev, M. and C. Robinson (2008). Human capital specificity: Evidence from the dictionary of occupational titles and displaced worker surveys, 1984–2000. *Journal of Labor Economics 26*(3), 387–420.

Robins, J. and M. F. Wiersema (1995). A resource-based approach to the multibusiness firm: Empirical analysis of portfolio interrelationships and corporate financial performance. *Strategic management journal 16*(4), 277–299.

Shaw, K. L. (1987). Occupational change, employer change, and the transferability of skills. *Southern Economic Journal*, 702–719.

Shepherd, B. et al. (2013). The gravity model of international trade: A user guide.

Turrell, A., B. Speigner, J. Djumalieva, D. Copple, and J. Thurgood (2018). Using job vacancies to understand the effects of labour market mismatch on uk output and productivity.

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic management journal 5*(2), 171–180.

Yildirim, M. A. and M. Coscia (2014). Using random walks to generate associations between objects. *PloS one 9*(8), e104813.

Appendix A O*NET Description

This annex provides additional detail on the O*NET database. Table 9 presents the list of Knowledge, Skills, and Abilities (KSA) items used to describe each occupation. Although not exhaustive, this list—when combined with the information on the level and importance of each item across occupations—captures meaningful variation between them. This is reflected in the distribution of the JSI scores which account for both presence and intensity of shared characteristics. On average, occupations share only about 50% of their KSA items, with few pairs exhibiting high similarity (i.e., JSI values approaching 1).

Table 9: List of KSA

Knowledge	**Skills**	**Abilities**
Administration and Management	Active Learning	Arm-Hand Steadiness
Biology	Active Listening	Auditory Attention
Building and Construction	Complex Problem Solving	Control Precision
Chemistry	Coordination	Dynamic Flexibility
Clerical	Critical Thinking	Dynamic Strength
Communications and Media	Equipment Maintenance	Explosive Strength
Computers and Electronics	Equipment Selection	Extent Flexibility
Customer and Personal Service	Installation	Finger Dexterity
Design	Instructing	Glare Sensitivity
Economics and Accounting	Judgment and Decision Making	Gross Body Equilibrium
Education and Training	Learning Strategies	Hearing Sensitivity
Engineering and Technology	Management of Financial Resources	Inductive Reasoning
English Language	Management of Material Resources	Information Ordering
Fine Arts	Management of Personnel Resources	Manual Dexterity
Food Production	Mathematics	Memorization
Foreign Language	Monitoring	Multilimb Coordination
Geography	Negotiation	Near Vision
History and Archeology	Operation and Control	Night Vision
Law and Government	Operation Monitoring	Number Facility
Mathematics	Operations Analysis	Perceptual Speed
Mechanical	Persuasion	Rate Control
Medicine and Dentistry	Programming	Reaction Time
Personnel and Human Resources	Quality Control Analysis	Response Orientation
Philosophy and Theology	Reading Comprehension	Sound Localization
Physics	Repairing	Spatial Orientation
Production and Processing	Science	Speed of Closure
Psychology	Service Orientation	Speed of Limb Movement
Public Safety and Security	Social Perceptiveness	Static Strength
Sales and Marketing	Speaking	Visual Color Discrimination
Sociology and Anthropology	Systems Analysis	Visualization
Telecommunications	Systems Evaluation	Wrist-Finger Speed
Therapy and Counseling	Technology Design	
Transportation	Time Management	
	Troubleshooting	
	Writing	

To provide a sense of how often the different items of KSA are used to describe different occupations, Table 10, Table 11, and Table 12 display the items that are mostly used in each occupation, those that are mostly used at an advanced level (higher than 4), and those that are the least used. If considered at any level, table Table 10 show that there are several Skills and Abilities that are used by all the occupations. Examples are "Active listening", "Critical Thinking", "Speaking" (Skills), "Oral Comprehension", "Near Vision", or "Oral Expresion" (Abilities). This results in several high values of JSI 1, that doesn't consider the level the occupations are employed. Nedelkoska et al. (2015) proposes to consider the intensity and eliminate items that are used by more than 95% of the occupations to reduce the noise caused by them. JSI 2 follows the same approach.

Table 10: KSA most used at any level

Knowledge		**Skills**		**Abilities**	
English Language	97%	Active Listening	100%	Oral Comprehension	100%
Customer and Personal Service	85%	Critical Thinking	100%	Near Vision	100%
Administration and Management	79%	Speaking	100%	Problem Sensitivity	100%
Mathematics	79%	Monitoring	100%	Oral Expression	100%
Education and Training	78%	Time Management	98%	Speech Recognition	100%

Table 11: KSA most used at an advanced level (> 4)

Knowledge		**Skills**		**Abilities**	
Customer and Personal Service	47%	Reading Comprehension	31%	Oral Comprehension	38%
English Language	38%	Critical Thinking	22%	Oral Expression	36%
Computers and Electronics	30%	Active Listening	18%	Written Comprehension	30%
Education and Training	29%	Speaking	17%	Deductive Reasoning	24%
Mathematics	28%	Writing	17%	Written Expression	23%

Table 12: KSA least used at any level

Knowledge		**Skills**		**Abilities**	
Foreign Language	3%	Installation	5%	Dynamic Flexibility	0%
Fine Arts	5%	Technology Design	6%	Explosive Strength	1%
Food Production	5%	Programming	6%	Night Vision	2%
History and Archeology	5%	Management of Financial Resources	14%	Sound Localization	3%
Philosophy and Theology	9%	Management of Material Resources	14%	Peripheral Vision	4%

The picture changes when we take into account the level the items are used. The item most employed at an advanced level is "Customer and Personal Service", employed by 47%

(Knowledge) of the occupations, followed by "English Language" [24] (Knowledge, 38%), and "Oral Comprehension" (Ability, 38%). While "English Language" is required by 97% of the occupations, only 38% require it at an advanced level. To have an idea, the O*NET database provides anchor descriptions for each of the values - for the case of "English Language", level 2 is equivalent to "Write a thank you note", while level 4 refers to "Edit a feature article in a local newspaper", and level 6, to "Teach a college English class". Although almost all occupations require a basic understanding of the language, less than half require workers to use it in an eloquent way. To have another example of anchor descriptions, "Customer and Personal Service" at level 2 is described as "Process customer dry-cleaning drop off", at level 4 as "Work as a day care aide supervising 10 children", and at level 6 as "Respond to a citizen's request for assistance after a major disaster".

Finally, the items most rarely used (at any level) are described at Table 12. "Dynamic Flexibility" (Ability) is only required by 14 out of the 968 occupations, and O*NET recommends discarding most of them, therefore considered by only 5 in our sample (including athletes, choreographers, and aerobic instructors). Other examples of uncommon items are "Explosive Strength", used by athletes, police officers, baggage porters and bellhops, and psychiatric technician; "Night vision", used by funeral service managers, zoologists, firefighters, and pipelayers; and "Foreign Language", used by interpreters, curators, immigration and customs inspectors, managers, nurse practitioners, and construction laborers [25].

[24] The O*NET database is built based on the American labor market, therefore, English is the most widely spoken language. When applying to other contexts, I assume this would be the most widely spoken language in the country, Portuguese in the case of Brazil.

[25] This case in particular is likely an example where the O*NET database may describe a characteristic of an occupation more commonly used in the American labor market, which might not be the case in other countries. According to the **?**, about 3 million construction workers are foreign-born, representing 25% of the total construction workforce. That might not be the case in other countries. However, for this analysis, the important factor is that the relation between different occupations are similar - when building the JSI, these different characteristics may be diluted across the other characteristics associated with the occupation. In addition, as long as the differences other similar occupations have across countries are similar, the JSI will remain the same.

Appendix B Description of Dyads - Number of transitions and JSI

This annex provides examples of pairs of occupations (dyads), their JSI, the number of workers that transition between them between 2004 and 2008, or between 2008 and 2012, and how much that represents of the total labor mobility in the country, or the number of workers transitioning from the occupation of origin. Table 13 and Table 14 display the dyads with the highest share of transitions compared to the total labor mobility in the country. The most common labor transition in the first period are Retail operators in stores and markets taking on a position of assistant agents and administrative assistants, with 332,334 transitions, representing 0.87% of all transitions between two occupations in the country. The JSI 5 between them is only 0.409, meaning that out of the total number of KSA employed by both occupations, less than half, only 40.9% of them are common to both. Other common dyads show sometimes lower JSI (e.g. "Workers in waste collection services for cleaning and conservation of public areas" and "Workers in building maintenance services", that share only 39% of the KSA), sometimes higher (e.g. "High school teachers in elementary school" and "Higher-level elementary school teachers (first to fourth grades)", with a JSI 5 of 95.6%). However, although not particularly low, the average JSIs are not clearly high for the dyads with the highest number of transitions. This might reflect however, general trends in the labor market, where one occupation attracts more workers, or reduces at a larger pace. This is the reason for the use of the gravitational model. Table 15 and Table 16 display the dyads with the highest share of transitions out of the origin occupation (which takes into account the push factor), and it is already visible that the JSI are consistently higher.

Table 13: JSI 1, 3 and 5 across the dyads with the 15 highest number of transitions (period 1)

Origin CBO		Destination CBO		JSI 1	JSI 3	JSI 5	Total num. of transitions	Share on the the total number of transitions
5211	Retail operators in stores and markets	4110	Assistant agents and administrative assistants	0.526	0.476	0.409	332,334	0.87%
4110	Assistant agents and administrative assistants	5211	Operators in shops and markets	0.526	0.476	0.409	280,633	0.73%
5211	Retail operators in stores and markets	4211	Cashiers and ticket sellers (except bank tellers)	0.474	0.571	0.570	228,459	0.60%
4211	Cashiers and ticket sellers (except bank tellers)	5211	Operators in shops and markets	0.474	0.571	0.570	207,907	0.54%
5143	Workers in building maintenance services	5142	Workers in waste collection services, cleaning and conservation of public areas	0.412	0.313	0.397	189,694	0.50%
5142	Workers in waste collection services for cleaning and conservation of public areas	5143	Workers in building maintenance services	0.412	0.313	0.397	186,882	0.49%
4221	Receptionists	4110	Assistant agents and administrative assistants	0.684	0.717	0.677	166,686	0.44%
5211	Retail operators in stores and markets	3541	Specialists in product promotion and sales	0.737	0.533	0.516	162,247	0.42%
5134	Workers in customer service in food, beverage and hotel establishments	5211	Operators in shops and markets	0.579	0.683	0.587	155,300	0.41%
3541	Specialists in product promotion and sales	5211	Operators in shops and markets	0.737	0.533	0.516	155,283	0.41%
3312	High school teachers in elementary school	2312	Higher-level elementary school teachers (first to fourth grades)	1.000	0.985	0.956	154,666	0.40%
7170	Civil construction assistants	7152	Workers in masonry structures	0.750	0.632	0.673	147,462	0.39%
7823	Drivers of small and medium-sized vehicles	7825	Drivers of general cargo vehicles	0.647	0.700	0.673	147,001	0.38%
2312	Higher education teachers in elementary school (first to fourth grades)	3312	Secondary-level elementary school teachers	1.000	0.985	0.956	135,353	0.35%
4223	Telemarketing operators and similar	4110	Assistant agents and administrative assistants	0.467	0.395	0.388	132,342	0.35%

Table 14: JSI 1, 3 and 5 across the dyads with the 15 highest number of transitions (period 2)

Origin CBO		Destination CBO		JSI 1	JSI 3	JSI 5	Total num. of transitions	Share on the the total number of transitions
5211	Retail operators in stores and markets	4110	Assistant agents and administrative assistants	0.526	0.476	0.409	620,565	0.84%
4110	Assistant agents and administrative assistants	5211	Operators in stores and markets	0.526	0.476	0.409	584,142	0.79%
5211	Retail operators in stores and markets	4211	Cashiers and ticket sellers (except bank tellers)	0.474	0.571	0.570	538,597	0.73%
4211	Cashiers and ticket sellers (except bank tellers)	5211	Operators in stores and markets	0.474	0.571	0.570	532,885	0.72%
5142	Workers in waste collection services for cleaning and maintaining public areas	5143	Workers in building maintenance services	0.412	0.313	0.397	376,313	0.51%
5134	Workers in customer service in food, beverage and hotel establishments	5211	Operators in stores and markets	0.579	0.683	0.587	356,459	0.48%
5143	Workers in building maintenance services	5142	Workers in waste collection, cleaning and conservation services in public areas	0.412	0.313	0.397	329,779	0.45%
4221	Receptionists	4110	Assistant agents and administrative assistants	0.684	0.717	0.677	309,277	0.42%
5211	Retail operators in stores and markets	5134	Workers in customer service in food, beverage and hotel establishments	0.579	0.683	0.587	300,968	0.41%
2312	Higher education teachers in elementary school (first to fourth grade)	3312	High school teachers in elementary school	1.000	0.985	0.956	285,799	0.39%
7823	Drivers of small and medium-sized vehicles	7825	Drivers of general cargo vehicles	0.647	0.700	0.673	285,567	0.39%
5211	Retail operators in stores and markets	3541	Specialists in product promotion and sales	0.737	0.533	0.516	284,780	0.38%
3312	Secondary education teachers in elementary school	2312	Higher education teachers in elementary school (first to fourth grades)	1.000	0.985	0.956	284,345	0.38%
3541	Specialists in product promotion and sales	5211	Operators in stores and markets	0.737	0.533	0.516	278,922	0.38%
4110	Assistant agents and administrative assistants	4221	Receptionists	0.684	0.717	0.677	264,640	0.36%

Table 15: JSI 1, 3 and 5 across the dyads with the 15 highest share of transitions out of occupation A (period 1)

Origin CBO		Destination CBO		JSI 1	JSI 3	JSI 5	Total num. of transitions	Share on the transitions out of CBO A
2252	Doctors in surgical specialties	2251	Clinical doctors	0.900	0.725	0.719	18,411	75.4%
2253	Doctors in diagnostic and therapeutic medicine	2251	Clinical doctors	0.950	0.783	0.758	1,614	66.7%
3411	Commercial aviation pilots, flight mechanics and similar	2153	Aircraft pilots	0.944	0.788	0.793	1,030	58.6%
3213	Aquaculture technicians	3171	System and application development technicians	0.471	0.450	0.514	1,383	57.0%
9153	Biomedical equipment maintenance and repair technicians	3222	Nursing technicians and assistants	0.800	0.625	0.541	4,601	56.0%
7256	Assemblers of aircraft systems and structures	3143	Vehicle mechanics technicians	0.650	0.667	0.667	1,787	51.4%
7824	Bus drivers, metropolitan and highway drivers	7825	General cargo vehicle drivers	0.700	0.699	0.728	58,143	50.7%
2153	Aircraft pilots	3411	Commercial aviation pilots, flight mechanics and related professionals	0.944	0.788	0.793	866	48.2%
7630	Multipurpose clothing manufacturing professionals	7632	Clothing machine operators	0.600	0.340	0.404	28,246	46.4%
8484	Grain tasting and classification workers and similar	5211	Shop and market salespeople	0.214	0.200	0.226	4,027	46.1%
3764	Models	5211	Shop and market salespeople	0.000	0.074	0.118	383	45.2%
2711	Chefs and similar	5132	Cooks	0.950	0.759	0.631	578	42.9%
3514	Court officials and similar	4110	Assistant agents and administrative assistants	0.632	0.516	0.496	18,831	42.3%
3763	Event presenters, programs and shows	5211	Shop and market salespeople	0.650	0.513	0.470	3,331	40.3%
5241	Home salespeople	5211	Shop and market salespeople	0.700	0.550	0.463	23,602	40.0%

Table 16: JSI 1, 3 and 5 across the dyads with the 15 highest share of transitions out of occupation A (period 2)

Origin CBO		Destination CBO		JSI 1	JSI 3	JSI 5	Total num. of transitions	Share on the transitions out of CBO A
2252	Doctors in surgical specialties	2251	Clinical doctors	0.900	0.725	0.719	49,674	73.0%
201	Senior officers of the military police	212	Corporals and soldiers of the military police	0.947	0.846	0.834	61,384	69.1%
6122	Agricultural producers in the cultivation of fibrous plants	6221	Agricultural workers in grass cultivation	0.529	0.369	0.351	633	59.8%
2253	Doctors in diagnostic and therapeutic medicine	2251	Clinical doctors	0.950	0.783	0.758	4,522	57.0%
2711	Chefs and similar	5132	Cooks	0.950	0.759	0.631	11,086	50.3%
2261	Osteopaths and chiropractors	2251	Clinical doctors	0.850	0.681	0.690	2	50.0%
6310	Multipurpose fishermen	6312	Coastal and high seas fishermen	1.000	1.000	1.000	728	49.5%
3281	Necropsy technicians and taxidermists	3222	Nursing technicians and assistants	1.000	0.689	0.657	1,073	49.4%
7824	Urban and highway bus drivers	7825	Drivers of general cargo vehicles	0.700	0.699	0.728	108,148	45.9%
7630	Multipurpose clothing manufacturing professionals	7632	Operators of sewing machines for clothing	0.600	0.340	0.404	44,927	45.7%
202	Military police captains	201	Senior officers of the military police	1.000	1.000	1.000	4,006	41.7%
212	Military police corporals and soldiers	201	Senior officers of the military police	0.947	0.846	0.834	65,683	40.9%
3514	Court officials and similar	4110	Assistant agents and administrative assistants	0.632	0.516	0.496	34,087	40.9%
2429	Intelligence professionals	1114	Civil service managers	0.700	0.526	0.550	96	40.7%
3411	Commercial aviation pilots, flight mechanics and similar	2153	Aircraft pilots	0.944	0.788	0.793	1,006	40.5%

The UNU-MERIT WORKING Paper Series

2025-01 *Development strategies for the green hydrogen economy in emerging economies* by Fabianna Bacil, Anthony Black, Marina Domingues, Jun Jin, Rasmus Lema, Glen Robbins and Sören Scholvin

2025-02 *Do global value chains and local capabilities matter for economic complexity in EU regions?* by R. Boschma, E. Hernández-Rodríguez, A. Morrison and C. Pietrobelli

2025-03 *Innovation and productivity: the recent empirical literature and the state of the art* by Jacques Mairesse, Pierre Mohnen and Ad Notten

2025-04 *Digital Technologies, Hiring, Training, and Firm Outcomes: Evidence from AI and ICT Adoption in Indian firmss* by Sneha Marydas, Nanditha Mathew, Giordano De Marzo and Carlo Pietrobelli

2025-05 *New technologies and employment: the state of the art* by Marco Vivarelli and Guillermo Arenas Díaz

2025-06 *Formal-Informal Supply Chain Linkages and Firm Productivity in Sub-Saharan Africa: The Role of Human Capital* by Gbenoukpo Robert Djidonou, Neil Foster-McGregor and Nanditha Mathew

2025-07 *Internet Access and Youth's Mental Health and Well-being: Evidence from Ethiopia* by Trang Thi Pham and Pui-Hang Wong

2025-08 *A Twin Transition or a policy flagship? Emergent constellations and dominant blocks in green and digital technologies* by Linnea Nelli, Maria Enrica Virgillito and Marco Vivarelli

2025-09 *Good for business, not so much for the environment? Entry into importing and the energy intensity of Indian plants* by Michele Delera, Nanditha Mathew and Tania Treibich

2025-10 *The role of business visits in fostering R&D investment* by Marco Vivarelli, Mariacristina Piva and Massimiliano Tani

2025-11 *Not as innocent as it seems? The effects of "neutral" messaging on refugee attitudes* by Tobias Hillenbrand, Bruno Martorano and Melissa Siegel

2025-12 *Europeanizing Universities: A Catalyst for Social Cohesion and Sustainable Economic Development* by Jo Ritzen

2025-13 *Robots & AI Exposure and Wage Inequality: A Within Occupation Approach* by Florencia Jaccoud

2025-14 *Social Protection and Climate Disasters: Brazil's Responses to the 2024 Rio Grande do Sul Floods* by Raquel Tebaldi

2025-15 *A New Perspective on European Labour Migration: A Case Study of the Opportunity Cost for Greece and Poland* by Katrin Marchand, Pavlos Liagkas, Dani Smith and Aleksandra Wojnar

2025-16 *Does Training in AI Affect PhD Students' Careers? Evidence from France* by Pierre Boutros, Eliana Diodati, Michele Pezzoni and Fabiana Visentin

2025-17 *Endogenous R&D elasticities of productivity functions and BERD bias: Explorations of functional coefficient regressions* by Thomas Ziesemer

2025-18 *A data set for domestic and foreign private and public R&D stocks and labour-augmenting technical change for 44 rich or emerging economies with explorations on panel unit roots, cointegration, growth rate slowdown, and con- or divergence* by Thomas Ziesemer

2025-19 *Avenues to Maximizing Value Added from Critical Minerals* by Carlo Pietrobelli and Jorge Valverde

2025-20 *Corruption as a Gendered Barrier: Informal Institutions and the Limits of Democratization in North Macedonia* by Milka Muratovska, Ina Kubbe and Ortrun Merkle

2025-21 *Generative AI: A Double-Edged Sword for Western Democracy* by Jo Ritzen

2025-22 *Artificial intelligence as a complement to other innovation activities and as a method of invention* by Guillermo Arenas Díaz, Mariacristina Piva and Marco Vivarelli

2025-23 *Social Media and Son Preference: Evidence from India* by Praachi Kumar and Bruno Martorano

2025-24 *Human Capital and Occupational Mobility: The Case of Brazil* by Renata Mayer Gukovas