Mapping Skill Requirements in ICT Occupations using an IRT Modelling Approach

Alessandro Gallo, Francesca Adele Giambona, Lucia Buzzigoli and Laura Grassini[1]

Abstract

Online Job Advertisements (OJAs) represent a valuable resource for analysing labour market demand in contemporary economies. In this paper, we use proprietary Lightcast data to study ICT labour market demand in Italy. The dataset provides one of the most comprehensive views of online labour demand, covering millions of job postings collected daily from a wide range of sources, including specialised job portals and company websites. We propose a new approach to classify and map skill requirements for ICT occupations across selected Italian regions and provinces. The approach is based on a Multilevel Item Response Theory (IRT) model and delivers two main outputs. First, it constructs an index of "skill intensity" for each job advertisement, capturing the overall level of skill requirements and enabling territorial comparisons across regions and provinces. Second, it estimates parameters that identify the most critical skills for the ICT sector as a whole. These results can inform targeted economic and social policies and support early-career workers entering the ICT labour market. In addition, the skill-intensity index can be aggregated and incorporated into macroeconomic models to examine its relationship with territorial socio-economic indicators.

Key words: OJAs, Skills, Labour Market Demand, Item Response Theory, ICT.

Introduction

Online Job Advertisements (OJAs) provide real-time, detailed insights into the skills and requirements demanded by employers, offering a unique lens on labour market demand (Deming and Kahn, 2018; Deming and Noray, 2020; Atalay et al., 2025; Grasso and Tatsiramos, 2025). In a country marked by strong regional disparities and evolving economic challenges, analysing OJAs allows us to map skill requirements across sectors and territories, track emerging trends, and assess implications for employment, labour market participation, and economic inequality (Giambona et al., 2024). The use of OJAs is particularly relevant for analysing and tracking the evolution of Information and Communication Technology (ICT) occupations and the associated skills, as they constitute the main recruitment channel closely aligned with the preferences of prospective candidates (Kahlawi et al., 2024a; Kahlawi et al., 2024b). ICT occupations are defined using the ESCO[2] 2-digit groups of occupations 25 – *Information and communication technology professionals*[3]. To analyse OJAs we

[1] *Alessandro Gallo, University of Florence; email: alessandro.gallo@unifi.it*
Francesca Adele Giambona, University of Florence; email: francesca.giambona@unifi.it
Lucia Buzigoli, University of Florence; email: lucia.buzzigoli@unifi.it
Laura Grassini, University of Florence; email: laura.grassini@unifi.it

[2] ESCO stands for European Skills, Competences, and Occupations, more information are available at https://esco.ec.europa.eu/en

[3] We decided to focus on "professionals" since the role of skills for these types of occupations is more relevant than for lower level ones.

use Lightcast data. This data provides one of the most comprehensive views of online labour market demand, covering millions of job postings collected daily from a wide range of sources, including specialised job portals and company websites.
This paper contributes to the literature in several ways. First, it produces a measure of average "skill intensity" within each territory, enabling comparisons across occupations, geographical areas, and over time introduces. In addition, it proposes a novel approach to measuring and mapping the most important labour market skills across occupations and territorial areas.
Methodologically, the approach builds on the Item Response Theory (IRT) framework, a widely applied probabilistic model for measuring latent traits from categorical or dichotomous data (Samejima, 1969; De Ayala, 2009; Kamata & Vaughn, 2011). By adopting a Multilevel IRT model (Kamata and Vaugh, 2011; Chalmers, 2012, 2015), we derive an index of the *skill requirement level demanded in each job advertisement* named "skill intensity", supporting the potential longitudinal analyses of the most demanding occupations across the Italian regions. Moreover, this framework allows the estimation of parameters that identify the most critical ICT skills, also within each occupation.

Data and Descriptive Statistics

This work relies on OJA provided by Lightcast. Lightcast is a private labour market analytics firm specialising in the collection and harmonisation of job vacancy postings, which are used to analyse labour demand patterns, occupational structures, and job requirements. Beyond occupations, the dataset also includes information on skills requested, contract type, education required, companies, location, and other information. Most variables refer to official classifications (such as LAU and NUTS for geographical areas), and occupations and skills are traced back to the ESCO taxonomy. We started by considering all the OJAs for the year 2024[4], leading to a sample composed of 4,439,258 OJAs. We then focused on the ICT occupations, reducing the sample to 148,548 OJAs and roughly 1100 different skills. Some descriptive statistics from our sample are shown in Figure 1 and Table 1. Figure 1 illustrates the distribution of OJAs for ICT occupations across Italian regions. The majority of job ads are posted in Lombardia (29.8%), followed by Emilia-Romagna (16.3%) and Lazio (12.7%). Together with Veneto (9.0%) and Piemonte (7.3%), these regions account for approximately 75% of total ICT labour demand. We focus our analysis on these five regions for three main reasons. First, they represent the largest share of the ICT labour market, as shown in Figure 1. Second, they exhibit a more similar labour market structure than other Italian regions—particularly some Southern regions (e.g., Calabria, Sicilia, and Puglia), where the economy is more tourism-based—allowing for more reliable territorial comparisons. Finally, concentrating on these regions enables a more granular analysis, including the provincial level (NUTS-3) within our Multilevel IRT model. Table 1 reports the average number of skills requested per OJA, disaggregated by occupation (ESCO level 4), providing an indication of which occupations are the most demanding in terms of requirements.

Table 1: Average number of skills requested in ICT ads by the occupations (ESCO level 4) in 2024.

[4] Although our project aims to consider also other available years (from 2019 to 2025).

Occupation (ESCO level 4)	
Software developers	26,2
Web and multimedia developers	24,7
Systems analysts	20,7
Database and network professionals	19,7
Applications programmers	19,7
Computer network professionals	19,2
Software and app. dev. & analysts	17,3
Systems administrators	17,1
Database designers and administrators	15,8

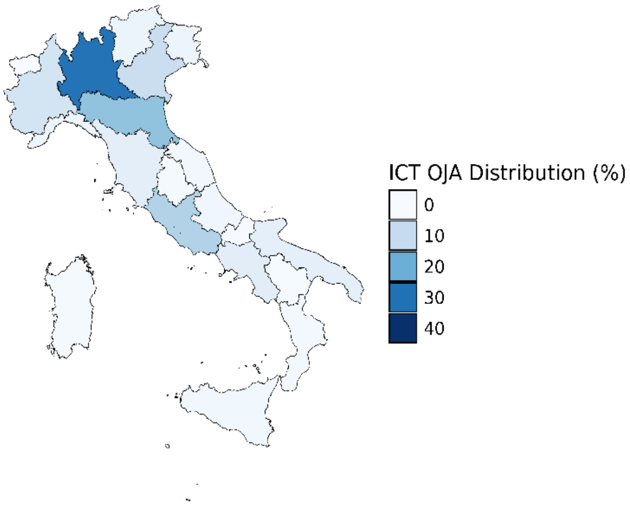

Figure 1: Choropleth map describing the distribution (%) of ICT OJA in Italian regions in 2024.

Methodology

IRT is a probabilistic framework widely used to measure latent traits through categorical items, including both dichotomous and ordered polytomous responses (Samejima, 1969; De Ayala, 2009). This methodology was originally developed to analyse students' performance in multi-item tests by modelling the probability of selecting a specific response category using a logistic function. These probabilities depend on parameters related to both individual characteristics and item properties. In that context, performance is measured using responses to a set of test items, typically multiple-choice questions. Formally, consider a set of students $k = 1, \dots, N$ and a set of items $X_i = X_1, \dots, X_K$, where $K, N \in \mathbb{N}$. Let X_{ki} denote the response of student k to item i, with $X_{ki} = 1$ if the student answers item i correctly and $X_{ki} = 0$ if the answer is incorrect or missing. To evaluate students' performance, one of the most used specifications is the two-parameter logistic (2PL) IRT model:

$$P(X_{pi} = 1 | \theta_p, a_i, b_i) = \frac{\exp\left[a_i(\theta_p - b_i)\right]}{1 + \exp\left[a_i(\theta_p - b_i)\right]},$$

where θ_k is the performance of the student k and a_i, b_i are two parameters referring to the items. a_i is the discrimination parameter and captures how strongly item i distinguishes between low- and high-performing students (given their ability θ). A higher a_i implies that item i provides greater discrimination. b_i is the difficulty parameter and reflects how difficult item i is. Formally, b_i is the value of ability θ at which the probability of a correct response equals 0.50. Therefore, higher b_i values indicate that a higher ability level is required to have a 50% chance of answering item i correctly (i.e., the item is more difficult).

In the context of the Italian labour market, and focusing on the demand side captured by OJAs, we define $k = 1, \dots, N$ as the set of OJAs, while the items X_i are interpreted as the skills requested in the postings. Accordingly, θ_k can be interpreted as the skill intensity required by OJA k. Each skill X_i can then be characterised by the parameters a_i and b_i: a_i measures how strongly the skill discriminates between low- and high-skill-intensity OJAs, while b_i indicates the level of θ_k at which the probability of requiring that skill equals 0.50. Therefore, X_{ki} is a dichotomic variable categorized as, $X_{ki} = 1$ denotes if the OJA k requests the skill i and $X_{ki} = 0$ otherwise.

Formally, since the OJA data have a complex multilevel structure, we estimate the following IRT multilevel model (Fox & Glas, 2001; Kamata and Vaugh, 2011; Chalmers, 2012, 2015), where we consider three levels: regions (NUTS-2, 20 regions) provinces (NUTS-3, 41 provinces) and occupations (ESCO level 4, 9 occupations).

$$\Pr(X_{ki} = 1 \mid \theta_k, a_i, b_i) = \frac{\exp[a_i(\theta_k - b_i)]}{1 + \exp[a_i(\theta_k - b_i)]}.$$

In this model, θ_k is given by:

$$\theta_p = \mu + v_{r(k)} + \gamma_{p(k)} + u_{g(k)} + \varepsilon_k,$$

with random effects:

$$\begin{aligned}
v_r &\sim \mathcal{N}\left(0, \sigma^2_{\text{reg}}\right), && r = 1, \dots, R,\\
\gamma_p &\sim \mathcal{N}\left(0, \sigma^2_{\text{prov}}\right), && r = 1, \dots, P,\\
u_g &\sim \mathcal{N}(0, \sigma^2_{\text{occ}}), && g = 1, \dots, G,\\
\varepsilon_k &\sim \mathcal{N}(0,1).
\end{aligned}$$

v_r is the region-specific shift in skill intensity, u_g is the occupation-specific shift in skill intensity, γ_p is the province-specific shift in skill intensity and ε_k captures residual heterogeneity within occupation–region cells[5].

Discussion and Expected Results

This approach is suitable for two main reasons. First, the latent trait can be interpreted as an index of "skill intensity" that is the *level of skills required* in a job advertisement, under the assumption that *a higher number of requested skills reflects a higher level of requirements demanded*[6]. This provides a measure of the requirement level asked in each job ad. These unit level results can be aggregated to produce regional and provincial estimates of the average skill requirement demanded, as well as to track trends over time by exploiting the longitudinal nature of the data.

Second, this method allows us to identify the most relevant skills through the parameters a_i and b_i. The difficulty parameter b_i captures how *difficult (rare)* a skill is in OJAs: it indicates the level of the latent trait θ_{kg} at which an OJA has a 0.50 probability of including that skill. The discrimination parameter a_i measures how strongly a skill differentiates between job ads requiring low versus high levels of θ_{kg}. In other words, skills with high discrimination are especially informative because they are disproportionately associated with the most skill-intensive positions. Therefore, possessing highly discriminating skills is likely to be particularly important for accessing the most demanding OJAs.

Identifying the skills that are most predictive of high skill intensity is relevant from both an economic and a policy perspective. Labour market research consistently shows that an increase in the number

[5] To estimate the model, we use the R package mirt and the function mixedmirt(). Estimation is performed via the Metropolis–Hastings Robbins–Monro (MH-RM) algorithm, which is the standard approach for this class of models in the literature (Chalmers, 2012; Chalmers, 2015).

[6] This may be considered a strong assumption considering a broad ICT sector, however the Multilevel IRT approach mitigates considerably this issue, levelling the skills across job ads considering occupation, province and region of the job ads.

(and complexity) of skills required is associated with higher wages and, more generally, with more rewarding occupations (Deming and Kahn, 2018; Deming and Noray, 2020; Atalay et al., 2025; Grasso and Tatsiramos, 2025). For this reason, understanding which skills are most strongly linked to access to more skilled positions can be a useful tool for early-career workers entering the ICT labour market.

This approach can provide new insights into the analysis of ICT labour market skills demand. On the one hand, it offers a clearer picture of the most in-demand skills across ICT occupations, from both a territorial and potentially from an intertemporal perspective. On the other hand, the estimated skill intensity index can be incorporated into macro-econometric models to examine its impact on other relevant economic outcomes.

Data Disclaimer: This article reflects solely the results and opinions of the authors and does not necessarily represent the position of Lightcast, which shall not be held liable for its content.

References

Atalay, E., Sotelo, S., & Tannenbaum, D. (2024). The geography of job tasks. *Journal of Labor Economics*, *42*(4), 979-1008.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of statistical Software*, *48*, 1-29.

Chalmers, R. P. (2015). Extended mixed-effects item response models with the MH-RM algorithm. *Journal of Educational Measurement*, *52*(2), 200-222.

De Ayala, R. J. (2009). *The theory and practice of item response theory*. Guilford Publications.

Deming, D., & Kahn, L. B. (2018). Skill requirements across firms and labour markets: Evidence from job postings for professionals. *Journal of Labour Economics*, *36*(S1), S337-S369.

Deming, D. J., & Noray, K. (2020). Earnings dynamics, changing job skills, and STEM careers. *The Quarterly Journal of Economics*, 135(4), 1965-2005.

Fox, J. P., & Glas, C. A. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, *66*(2), 271-288.

Giambona, F., Kahlawi, A., Buzzigoli, L., Grassini, L., & Martelli, C. (2024). Skills in online job ads: An analysis of Italian regions. *Socio-Economic Planning Sciences*, *94*, 101916.

Grasso, G., & Tatsiramos, K. (2025). Demand for Skills and Wage Inequality. Preprint available at: https://giuseppe-grasso.github.io/uploads/GT_INPS_oct_2025.pdf

Kahlawi, A., Buzzigoli, L., Giambona, F., Grassini, L., & Martelli, C. (2024a). Online job ads in Italy: a regional analysis of ICT professionals. Statistical Methods & Applications, 33(2), 609-633.

Kahlawi, A., Grassini L., & Buzzigoli L. (2024b). The use of online job ads to analyse skill changes in ICT occupations. *Statistica Applicata,* vol. 36, pp. 249-271.

Kamata, A., & Vaughn, B. K. (2011). Multilevel IRT modeling. In *Handbook of advanced multilevel analysis* (pp. 41-57) edited by Joop Hox and J. Kyle Roberts. Routledge.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, 34(S1), 1-97.