Educational mismatch and productivity: evidence from LEED data on Italian firms

Laura Bisio, Matteo Lucchese

1. Introduction[1]

Over the past years, the role of the potential mismatch between the demand and supply of skills and qualifications has received considerable attention in Italy. However, the empirical evidence about the impact of this mismatch onto firms' productivity has not been fully documented so far. In the present paper, we investigate this issue empirically, exploiting the information available from the System of Statistical Registers built within the Italian National Institute of Statistics (Istat).

In particular, we focus on the "educational mismatch", defined as the difference between the educational attainment of workers (the highest level of education the worker has completed) and that "needed" for their job. In this way, over (under) education refers to situations where the individual's educational attainment is higher (lower) than the "required" level, thereby producing a surplus (deficit) of education. Indeed, this mismatch is the result of several overlapping factors, ranging from the adequacy of training to the (in)efficiency of the labour market or the ability of the economic system to absorb skilled labour. The latter issue increasingly depends on the speed at which technological change, and in particular the digitalization process, has changed the demand for skills in the last decades, especially for high-tech and knowledge-intensive industries.

The role of human capital as a key factor in improving firm's competitiveness has been already highlighted by Istat (Istat, 2018); investments in this area have been also recently found to be associated with an increase in firms resilience during the pandemic crisis (Istat, 2021). An analysis of the skill and qualification mismatch for the Italian economy is proposed by OECD (2016) and Monti e Pellizzari (2016), which aimed to provide statistical evidence of the roots of skill mismatch, based on the PIAAC survey results. More recently, the correlation between the ability to match the skills need and labour productivity has been pointed out by Fanti et al. (2021) for a representative sample of Italian firms based on the INAPP PEC survey.

In this paper we explore the effect of over/under education of employed workers on firms' productivity for the Italian economy on the basis of the work of Kampelmann and Rycx (2012), which provides evidence about the direct impact of educational mismatch on productivity using linked employer-employee data for a panel of Belgian firms covering the period 1999-2006.[2] By means of the Istat System of Statistical Registers, we are able to adapt the same analytical framework to Italian data, to contribute filling the gap in the literature about the link between human capital and firms' competitiveness in our economy. The results suggest that over/under education affects productivity growth in both manufacturing and services firms: in particular, over-education rises firm's productivity in medium and high-tech manufacturing firms as well as in less knowledge-intensive services, whereas under-education hampers productivity in manufacturing and services industries with a higher intensity of technology and knowledge.

This paper is organized as follows: section 2 presents the dataset and the empirical methodology; section 3 offers some preliminary descriptive statistics, section 4 shows the results and section 5 draws conclusions.

---

[1] An earlier version of this analysis appeared in the 2022 edition of "Istat Report on Competitiveness" (Istat, 2022).

[2] Mahy et al. (2015) extend the period of analysis to 2010 and highlight, among other results, that the effect of over-education on productivity is stronger in firms belonging to high-tech/knowledge-intensive industries – but with no distinction between manufacturing and services firms.

Laura Bisio, ISTAT, Italian National Institute of Statistics, Italy, bisio@istat.it, 0000-0003-0922-6359
Matteo Lucchese, ISTAT, Italian National Institute of Statistics, Italy, mlucchese@istat.it, 0000-0001-8331-7393

2. Data and empirical analysis

Our analysis is based on the integration of two different Statistical Registers (*Asia-Employment Register* and *Frame-SBS Register)*, covering almost the totality of Italian firms. The *Asia-Employment Register* (*Asia Occupazione*) is a LEED-type (Linked Employer Employee Database) one, which allows to obtain information related to firms, the workers employed therein and the main aspects of the work contracts; this dataset also provides information on the level of educational attainment achieved by each worker – via matching to the 2011 edition of the Population Census, updated through the "Information Base on education and qualifications" (*Base Informativa su istruzione e titoli di studio*, BIT). The "*Frame-SBS Register*", instead, provides data on firms' main economic and structural characteristics, including labour productivity.

The empirical analysis covers a large set of Italian firms with at least 20 workers over the period 2014–2019. Both labour productivity and mismatch variables are evaluated with respect to employees – i.e. self-employed are excluded from the analysis –, while employment is measured in terms of annual average job positions, based on the worker's weekly work attendance.[3]

As already mentioned, the empirical analysis is an application to the Italian case of the ORU (Over, Required and Under Education) model performed by Kampelmann and Rycx (2012), based on a longitudinal LEED data structure. The ORU model consists in a two-step procedure. The first step is aimed at computing the aggregate measures of over/under-education of workers at the firm level. The latter are calculated on the basis of the years of education "required" for a given type of "occupation", that – in our case – is identified by the combination of three elements: the economic sector in which the firm operates (2-digit economic sector according to Nace Rev.2), the workers' qualification (blue-collar, white-collar, apprentice, middle manager, manager/supervisor, other type) and their age class (15-29, 30-49, 50 and more). The "required" years of education correspond to the modal years of education of the workers employed within each type of occupation[4]. A worker is defined as over (under) educated if his/her years of education are higher (lower) than those required by the type of occupation in which he/she is employed. Once the years of over/under-education are calculated at the worker-level, three distinct measures are derived at the firm level by averaging the number of years of, respectively, "required" (REQ), over- (OVER) and under-education (UNDER) of the workers within each firm. As in Kampelmann and Rycx (2012), the following equations describe the firm-level "mismatch" variables:

- $REQ_j = \frac{1}{n_j} \sum_{i=1}^{n_j} REQ_{i,j}$ where $REQ_{i,j}$ is the "required" years of education for the type of occupation of the worker i employed in firm j, while n_j it is the number of employees in firm j;

- $OVER_j = \frac{1}{n_j} \sum_{i=1}^{n_j} OVER_{i,j}$ where $OVER_{i,j}$ is the difference between the number of years of education attained by worker i employed in firm j and the "required" number of years for his/her occupation, when the worker is over-educated; it is 0 otherwise;

---

[3] We are only able to investigate a specific type of mismatch occurring in the labour market, i.e. the poor matching in terms of education required/attained at the firm/worker. We cannot study e.g. the lack of matching between workers' skills or professional status and those needed by the firms. In addition, though the imbalances – of either qualification or skills – that can occur at the aggregate level are found to be related to mismatches at the individual level (Montt, 2015), they fall out of the scope of our analysis.

[4] The Asia-Employment Register reports the following 7 levels of educational attainment (i.e. 7 degrees) to which specific amounts of educational years are associated (in parenthesis): no education or primary education (5); lower secondary education (8); technical and professional upper secondary education (11); upper secondary education (13); tertiary education, 1st level degree (16); tertiary education, 2nd level degree (18); Ph.D. (21). It should be noticed that the educational attainment level does not have full coverage in the Asia-Employment Register (see below).

- $UNDER_j = \frac{1}{n_j}\sum_{i=1}^{n_j} UNDER_{i,j}$ where $UNDER_{i,j}$ is the difference between the number of years of education attained by the worker i employed in firm j and the "required" number of years for his/her occupation, when the worker is under-educated; it is 0 otherwise.

Thus, the sum of the three measures (REQ_j, $OVER_j$ and $UNDER_j$) is equal to the average years of education of the employees employed in firm j.

The second step is the estimate of a labour productivity function at the firm level, where the dependent variable is defined as value added per worker and the measures of educational mismatch are the key explanatory variables:

$$\ln PROD_{j,t} = \beta_0 + \beta_1\, lnPROD_{j,t-1} + \beta_2\, REQ_{j,t-1} + \beta_3\, OVER_{j,t-1} + \beta_4\, UNDER_{j,t-1} + \beta_5\, \bar{F}_{j,t} + \beta_6\, \bar{W}_{j,t} + \delta_t + \varepsilon_{j,t}$$

The regression also includes two vectors of control variables, $\bar{F}_{j,t}$ and $\bar{W}_{j,t}$, respectively related to firm's (2-digit economic sector according to Nace Rev.2, firm age, firm size, unit labour costs) and labour force characteristics (firm's average age of workers, the share of workers under 29 and over 50 years old, the share of female workers, the share of workers by professional status, the share of temporary and part-time workers). In addition, the lagged dependent variable controls for the potential persistency of labour productivity, while business cycle-related effects are taken into account by year dummies (δ_t).

The aim of the analysis is to verify how over/under-education can affect productivity (value added per worker) at the firm level, conditional to the average years of education required in each firm. The productivity equation can be consistently estimated by pooled ordinary least square (POLS), but the existence of firm-specific time-invariant factors influencing both labour productivity and the explanatory educational variables can make the estimated coefficients by POLS biased. The so-called "heterogeneity bias" can be properly tackled by a fixed-effects (FE) estimator. However, a second source of bias may also arise due to time-varying unobserved factors making educational mismatch being determined by the dynamics of firms' productivity (and *viceversa*)[5]. Such endogeneity issue undermines the unbiasness of the FE estimator. Thus, to take into account of both the heterogeneity and the simultaneity issues –as properly proposed by Kampelmann and Rycx (2012) – we adopt the dynamic "System-GMM" (Generalized Method of Moments) estimator by Arellano and Bover (1995) and Blundell and Bond (1998).

Finally, we apply this analysis to a balanced panel of over 36,500 manufacturing and services firms with at least 20 workers, operating during the whole period 2014-2019[6]. For the sake of robustness and adapting the work of Kampelmann and Rycx (2012) to our dataset, the original microdata underwent a few cleaning steps. In particular, we exclude firms with a share of missing values concerning workers' educational degree above 20%[7], type of "occupations" with less than 30 observations (workers) and firms for which labour productivity value lies below/above the $1^{st}/99^{th}$ percentile.

3. Descriptive statistics

Figure 1 shows the evolution of the number of required years of education, over-education and

---

[5] The interested reader may refer to Kampelmann and Rycx (2012) for a more thorough review of studies addressing this issue in the educational mismatch literature.

[6] We consider the following sections: C, G, H, I, J, L, M and N according to the Nace Rev.2 classification.

[7] The remaining missing values have been replaced with the required years of education in the relative type of occupation (we recode about 4% of total workers each year). It is worth noting that the share of missing values is rather constant across years, thus the cleaning procedure – either in the form of replaced or deleted observations – has been applied uniformly across time.

under-education at the firm level between 2014 and 2019, for the whole set of manufacturing and services firms in each year, according to the quartiles of their annual distribution and the mean values. The average number of required years grew from 10.65 to 11.07, with a slow but steady upward shift of the distribution, stronger in 2018 and 2019. In addition, the inter-quartile range increased from 2.59 in 2014 to 2.77 in 2019, revealing a widening of the dispersion of required years of education.

In the same period, over-education remained almost steady (around 1.2 years), while under-education increased from -0.70 years in 2014 to -0.75 in 2019 – indeed, a shrinking of years of under-education in absolute terms corresponds to an increase of the phenomenon. Both over and under-education exhibit standard deviation and interquartile range increasing over time, pointing to a growing divergence among firms in terms of their educational mismatch. At the sectoral level – not shown in Figure 1 –, the required years of education in the manufacturing sector slightly grew from 10.12 in 2014 to 10.36 in 2019, while the increase has been stronger in the service sectors (from 11.15 to 11.67). In 2019, over-education is more pronounced in the manufacturing sector (1.29 years and 1.10 years respectively), while under-education is higher in services (-0.61 and -0.87 years).

Figure 1. The required years of education, over-education and under-education - 2014-2019 (annual average by firm for the whole set of firms each year)

302
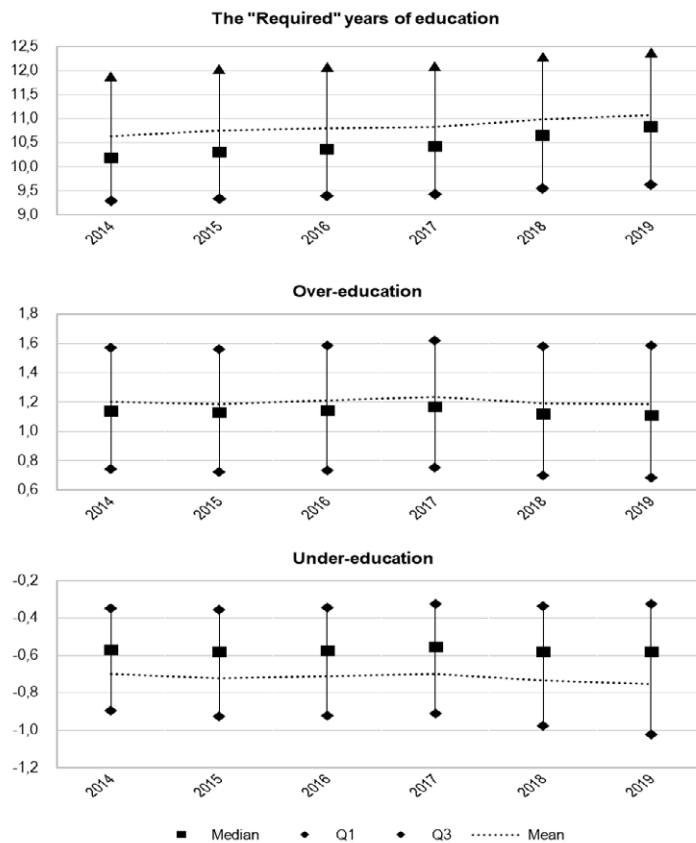
4. Results

The results of our estimates by GMM-SYS are reported in Table 1.[8] They show the effects of the educational mismatch on firm labour productivity, according to the different technological and knowledge intensity of sectors.[9] In each specification, the absence of second-order autocorrelation of the residuals to the differences has been verified using the Arellano-Bond test (Arellano and Bond, 1991), while the set of instruments is valid according to the Hansen test (Hansen, 1982). Results from POLS and FE estimators are not shown for the sake of brevity, but are available upon requests from the authors.

A one-unit (year) increase in the mean required years of education leads to an increase in firm productivity in both manufacturing and services sectors, but with greater intensity in high-tech industry; in addition, over-educated workers appear to be more productive and to bring a productivity premium to the firms in which they work, while under-educated workers hamper the productivity of the firms where they are employed[10]. Among manufacturing firms, the influence of over-education raises with the technological intensity, while it acts as a competitive factor especially for less knowledge-intensive services. Interestingly, our estimates highlight a (negative) impact of under-education for firms in high and medium-high technological industries and in knowledge-intensive services, where the relatively higher degree of complexity of production processes probably entails higher costs of using less educated human capital.

Table 1. The impact of the educational mismatch on firm productivity in Italian firms

	Dep: Labour productivity			
	GMM-SYS			
	High and Medium-High Tech Manufacturing	Low and Medium-Low Tech Manufacturing	Knowledge Intensive Services	Less Knowledge Intensive Services
Lagged Labour Productivity	0.067 **	0.375 ***	0.466 ***	0.065 ***
	(0.026)	(0.065)	(0.069)	(0.018)
Required Education	0.115 ***	0.046 *	0.038 *	0.083 ***
	(0.036)	(0.027)	(0.022)	(0.021)
Over-Education	0.099 ***	0.031 **	0.056 *	0.119 ***
	(0.037)	(0.015)	(0.029)	(0.020)
Under-Education	0.081 **	0.020	0.050 *	0.053
	(0.036)	(0.026)	(0.027)	(0.039)
Workers and Firm variables	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes
Arellano-Bond test (AR2)[a]	0.135	0.252	0.271	0.298
Hansen-J test [b]	0.228	0.645	0.215	0.176
Observations	32,064	65,429	16,087	66,583
Firms	6,616	13,386	3,348	13,637

*Standard errors in parentheses. Significance levels: * p < 0.1 , **p < 0.05, *** p < 0.01.*
a) P-value associated to the Arellano-Bond statistics testing null of absence of serial correlation of differentiated errors at the second lag.
b) P-value associated to the Hansen-J statistics testing the null of exogeneity of instruments.

---

[8] Table 1 only shows the estimated coefficients related to the mismatch variables, while those related to the control variables are not reported. Anyway, the results are in line with our expectations and available for the interested reader.

[9] We use Eurostat "High-tech aggregation by NACE Rev. 2" (3-digit for manufacturing, 2-digit for services), available at: https://ec.europa.eu/eurostat/cache/metadata/Annexes/htec_esms_an3.pdf.

[10] Because mean years of under-education take negative values by construction, a positive regression coefficient indicates a negative correlation between under-education and productivity – i.e. productivity rises when mean years of over-education increase or under-education decreases.

5. Conclusions

Providing strong empirical evidence of the relationship between human capital and firm productivity at the different levels of the technology ladder, our results offer some relevant implications that may steer the policy action towards an increase of the education levels achieved by the working population and a reduction of the mismatch between the demand and supply of skills and qualifications. The availability of longitudinal microdata at the firm level is indeed the main strength of this analysis, which applies and adapts to the Italian case the ORU framework proposed by Kampelmann and Rycx (2012) for a panel of Belgian firms.

There are, of course, several enhancements of our empirical analysis – e.g. improving the identification of specific types of occupations, controlling for potential "birth cohort" effects, exploring the potential mismatch among types of occupations and workers' relative fields of study – that have to be tackled by future work. And it would be also important to try disentangling the channels through which the productivity premium is achieved – e.g. those linked to the complementarities with digital technologies (see OECD, 2022). However, as the empirical evidence on this phenomenon is relatively scarce, we think that this analysis offers a useful, though preliminary, contribution to the ongoing debate on this crucial issue for the development of the Italian economy.

References

Arellano, M., and Bond, O. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, **58**(2), pp. 277–297.

Arellano, M., and Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, **68**(1), pp. 29-51.

Blundell, R., and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, **87**(1), pp. 115-143.

Fanti, L., Guarascio, D., Tubiana, M. (2021). Skill mismatch and the dynamics of Italian companies' productivity. *Applied Economics*, **53**(59), pp. 6790-6803.

Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, **50**(4), pp. 1029–1054.

Istat (2018). *Rapporto sulla Competitività dei settori produttivi*. Istat, Roma.

Istat (2022). *Rapporto sulla Competitività dei settori produttivi*. Istat, Roma.

Istat (2021). *Rapporto Annuale sulla situazione del Paese*. Istat, Roma.

Kampelmann S., and Rycx F. (2012). The Impact of Educational Mismatch on Labour Productivity. Evidence from Linked Panel Data. *Economics of Education Review*, **31**(6), pp. 918-931.

Mahy, B., Rycx, F., Vermeylen, G. (2015). Educational Mismatch and Firm Productivity: Do Skills, Technology and Uncertainty Matter?. *De Economist,* **163**, pp. 233-262.

Monti, P., and Pellizzari M. (2016). Skill Mismatch and Labour Shortages in the Italian Labour Market. *Innocenzo Gasparini Institute for Economic Research Policy Brief*, **2**.

Montt, G. (2015). The causes and consequences of field-of-study mismatch: An analysis using PIAAC. *OECD Social, Employment and Migration Working Papers*, **167**, OECD Publishing, Paris.

OECD (2016). *Skills Matter. Further Results from the Survey of Adult Skills*. OECD Publishing, Paris.

OECD (2022). *Closing the Italian digital gap: The role of skills, intangibles and policies*. OECD Science, Technology and Industry Policy Papers, **126**, OECD Publishing, Paris.