Accepted: 29 November 2024

DOI: 10.1111/ecca.12563

ORIGINAL ARTICLE

Economica LSE

Establishment size and the task content of jobs: evidence from 46 countries

Micole De Vera[1] | **Javier Garcia-Brazales**[2]

[1]Banco de España and IZA

[2]University of Exeter and LEAP (Bocconi)

Correspondence
Javier Garcia-Brazales, University of Exeter Business School, Department of Economics, Streatham Court, Rennes Drive, Exeter EX4 4PU.
Email: j.garcia-brazales2@exeter.ac.uk

Abstract

Using a mix of household- and employer-based survey data from 46 countries, we provide novel evidence that workers in larger establishments perform more non-routine analytical tasks, even within narrowly defined occupations. Moreover, workers in larger establishments rely more on the use of information and communication technologies to perform these tasks. We also document a 15% raw wage premium that workers in larger establishments enjoy relative to their counterparts in smaller establishments. A mediation analysis shows that our novel empirical facts on the task content of jobs are able to explain 5–20% of the establishment size wage premium, a similar fraction to what can be explained by selection of workers on education, gender and age.

KEYWORDS

cross-country evidence, establishment size, occupations, tasks, wage differential

JEL CLASSIFICATION

J24; J31; L25

1 | INTRODUCTION

Empirical studies show that the task composition of jobs can explain a large share of the dispersion of wages across occupations and time (Autor *et al.* 2003; Acemoglu and Autor 2011; Autor and Handel 2013; Goos *et al.* 2014; Acemoglu and Restrepo 2022). Most studies are, however, constrained by the fact that direct measures of the task content of jobs are not available in standard datasets. To overcome this limitation, authors typically resort to imputing the task content of occupations by means of alternative datasets such as the Occupation Information Network (O*NET) or the European Working Conditions Survey (EWCS). The use of these general task classifications relies on the implied assumption that the task composition of jobs under the same occupation is homogeneous. This strategy therefore rules out the possibility that heterogeneity in tasks drives within-occupation wage differences.

In this paper, we explore whether the assumption of homogeneity in the task composition of jobs within occupations holds empirically. To do this, we focus on a natural and relevant dimension along which the task content of a given occupation might vary: establishment size. Employer size is meaningful because it is a simple statistic that strongly correlates with total factor productivity and managerial quality both theoretically (Lucas 1978) and empirically (Bloom and Van Reenen 2007). Larger employers also tend to use more automation and offshoring (Alekseeva *et al.* 2021). Given this gradient in how production is organized, the task content of jobs in establishments of different sizes is likely to also differ (Ocampo 2022). These differences in the task content of jobs have concrete implications on the compensation of workers and may explain part of the so-called establishment size wage premium (ESWP).

Our main result is that the assumption of homogeneity in the task composition of jobs within occupations does not hold empirically. In particular, we find that, even *within* narrowly defined occupations, there are systematic differences in the task intensities performed by workers employed by establishments of different sizes. We document that the intensity of non-routine analytical tasks of workers in larger establishments is, on average, 5–15% of a standard deviation higher. The size of this gap is comparable in magnitude to the gap present between the countries at the two extremes of the development spectrum, as documented by Caunedo *et al.* (2023). We also report suggestive evidence in favour of an establishment size gradient in non-routine interpersonal and routine cognitive tasks. Moreover, we find that to undertake these tasks, workers in larger establishments rely significantly more on information and communication technologies (ICT). We interpret this result as an indication that workers in larger establishments may perform more non-routine analytical, non-routine interpersonal, and routine cognitive tasks through increased use of ICT.

These empirical patterns are robust to using both employer- and employee-based responses, and are widely present across the comprehensive set of 46 high-, middle- and low-income countries that we study after combining two large and representative datasets: the OECD Survey of Adult Skills and the World Bank Skills Measurement Surveys. Furthermore, replicating our analysis separately for each 1-digit ISCO occupation highlights that: (i) the patterns are still present when zooming into occupations, which increases the reliability of the within-occupation comparisons that we make between large and small establishments; and (ii) while the establishment size gradient in usage of ICT is present in all 1-digit occupations, the gradient for non-routine analytical tasks is concentrated in certain occupations, particularly professionals and elementary occupations. Finally, we provide evidence that our main finding is true not only for the average worker: the distributions of the intensity of performed tasks in larger establishments are also horizontally shifted relative to smaller establishments.

Though we acknowledge that we cannot rule out that the estimated gaps are partly confounded by selection of workers into firms of different sizes, it is reassuring that our main results hold after controlling for a rich set of observable worker characteristics, including education, cognitive and non-cognitive skills, and industry of employment. Moreover, following Oster (2019), we find that selection on unobservables would have to be *at least* as large as selection on observables for our estimated establishment size gradients in task content to be indistinguishable from zero.

This paper provides novel evidence of systematic differences in task intensity by comparable workers in the same occupation according to whether they are employed by a large or a small establishment, and that these differences help to explain wage differentials.[1] This result contributes to the recent literature documenting heterogeneity in task content within occupations (e.g. Deming and Kahn 2018; Stinebrickner *et al.* 2019), which has emphasized the role of dimensions other than employer size. Atalay *et al.* (2024) show that the variation in tasks is correlated with city size such that larger cities have higher intensity of analytical and interactive tasks, more technological requirements, and increased task specialization. Additionally, a number of papers report that occupational task content varies across countries (Dicarlo *et al.* 2016; Lewandowski *et al.* 2022; De La Rica *et al.* 2020; Caunedo *et al.* 2023).

Our results on the establishment size heterogeneity of tasks provide novel foundations to understand employer productivity and its dynamics. Our findings are consistent with the implications of static models that endogenize employers' decisions on how to allocate tasks for production (Ocampo 2022; Adenbaum 2023). For instance, Ocampo (2022) shows that automation may affect the task composition of occupations. Along with finding that workers in larger establishments use more ICT, we also document that they perform more non-routine analytical tasks compared to workers in smaller establishments. As employers grow larger, they likely accumulate more capital, automate, and offshore jobs (e.g. Jaimovich *et al.* 2023), which leads to changes in the task requirements of production. In particular, jobs may evolve to focus on non-routine analytical tasks and to use more ICT so as to complement the processes that aim to replace routine tasks. Future work exploring this mechanism may provide novel insights to understand the drivers of firm dynamics.

Moreover, the uncovered patterns on task heterogeneity connect naturally with the literature exploring the determinants of within-occupation wage dispersion, and provide a plausible novel driver of the ESWP—the empirical fact that larger establishments tend to pay their workers more for doing the same occupation. We explore the implications of the establishment size gradient in occupational task intensities on wage determination in two steps.

First, we document that, on average, workers in larger establishments earn about 15% more than their counterparts in smaller establishments, after controlling for 2-digit occupation codes. Our measured ESWP is consistent with the existence of large and economically significant employer-size wage premia found in other studies employing alternative datasets (Velenchik 1997; Gerlach and Hübler 1998; Schaffner 1998; Troske 1999; Winter-Ebmer 2001; Dobbelaere 2004; Söderbom *et al.* 2005; Lehmer and Möller 2010; Bloom *et al.* 2018; Colonnelli *et al.* 2018; Reed and Thu 2019; Lochner *et al.* 2020; Porcher *et al.* 2023). We extend previous analyses to show that this is not driven exclusively by a few workers in larger establishments who are paid disproportionately more. Rather, the distribution of wages in larger establishments is shifted to the right compared to the distribution of wages in smaller establishments.

Second, we conduct a mediation analysis to provide suggestive evidence on the sources of this ESWP, including our finding that task composition varies across establishments of differing sizes. A number of explanations for the existence of the ESWP have been proposed (Brown and Medoff 1989; Oi and Idson 1999): (i) large employers hire more skilled workers (worker selection); (ii) large employers have worse working conditions (compensating differentials); (iii) large employers have market power and share rents with workers (productivity); (iv) large employers have higher costs of monitoring and pay efficiency wages; and (v) large employers pay higher because of threat of unionization. In this paper, we explore the establishment size gradient in occupational task intensities as a complementary source of the ESWP. We find that this mechanism is able to explain over 10% of the raw ESWP. This proportion ranges from 5% to 20% across the countries in our sample. Our novel empirical pattern therefore accounts for an economically significant fraction of the ESWP that is comparable to the share explained by the sorting of higher-educated individuals into larger establishments.

These results open an exciting research avenue to explore the consequences of the employer size gradient in the task content of jobs on dynamic wage determination. Recent evidence indicates that early-career experience in large firms has dynamic rewards in future worker outcomes (Arellano-Bover 2024). Our findings suggest a compelling mechanism to rationalize this—the experience in performing non-routine analytical tasks and in using ICT accumulated by younger workers in larger establishments is rewarded with better future earnings prospects (Stinebrickner *et al.* 2019). We urge future research to probe along these lines.

1.1 | Outline of the paper

The remainder of this paper proceeds as follows. In Section 2, we describe the main datasets as well as our definition of occupations, and detail the measures of task content used in the analysis. In Section 3, we document novel facts on the heterogeneity of occupational task contents across establishments of differing sizes, and discuss potential drivers. In Section 4, we measure the ESWP and study a number of explanations for its existence, including the establishment size gradient in task intensity. Finally, in Section 5, we conclude with a summary of the findings and a discussion of future directions for work. The Appendix contains additional results.

2 | DATA AND MEASUREMENT

2.1 | Data sources

We take advantage of the availability of cross-country harmonized surveys reporting the tasks performed by individuals in their work to construct a rich dataset covering working (not self-employed) individuals aged 16–65 across 46 countries at various stages of economic development. We combine two main datasets.

2.1.1 | The OECD Survey of Adult Skills

The Survey of Adult Skills is a cross-sectional, cross-country survey conducted under the OECD Programme for International Assessment of Adult Competencies (PIAAC). This survey aims to measure cognitive skills (literacy, numeracy and problem solving in technology-rich environments), as well as skills used both at work and in other contexts. It is representative of a country's adult population aged 16–65, with around 5000 individuals participating in each country.[2] There have been three rounds of data collection (2008–13, 2012–16 and 2016–19). We focus on the surveys collected from the following 30 countries: Belgium, Chile, Cyprus, Czech Republic, Denmark, Ecuador, France, Greece, Hungary, Ireland, Israel, Italy, Japan, Kazakhstan, South Korea, Lithuania, Mexico, the Netherlands, New Zealand, Norway, Peru, Poland, Russian Federation, Singapore, Slovak Republic, Slovenia, Spain, Sweden, the UK and the USA.[3]

Full earnings information is not available in the public-use files of New Zealand, Peru, Singapore and the USA. Instead, earnings are reported only in deciles.[4] These countries are still employed in the quantification of the gradient in task intensity by firm size.

2.1.2 | World Bank Skills Measurement surveys

Our second main data source is the World Bank's STEP Skills Measurement Programme surveys. They also are cross-sectional surveys that aim to measure the demand and supply of skills in urban areas of low- and middle-income countries, which allows us to complement the set of high- and middle-income countries available in the OECD's Survey of Adult Skills. There are two types of surveys in the programme: household-based and employer-based.

The household-based surveys interview a randomly selected household member (aged 15–64) about their personal education and training history, work status and history, skills used in their job, earnings, individual competencies, and non-cognitive traits and abilities (e.g. personality, behaviour, risk preferences). Sample sizes vary from 3000 to 4000 individuals. We focus on the surveys that contain consistent questions regarding tasks and skills, corresponding to the following 11 countries: Armenia, Bolivia, China (Yunnan Province), Colombia, Georgia, Kenya, Laos, Macedonia, Sri Lanka, Ukraine and Vietnam.[5]

Additionally, in some countries, firms were also surveyed using an employer-based questionnaire. In this module, an informed respondent from around 300–500 firms per country reported the worker composition of the firm, the skills required from workers in different occupations, and the amount of in-firm training provided. We use the employer-based surveys of the following nine countries: Albania, Armenia, Azerbaijan, Bosnia-Herzegovina, Georgia, Kenya, Kosovo, Serbia and Vietnam. Note that four of these countries have also conducted the household-based survey, which allows us to document consistent evidence of within-occupation task heterogeneity from both the employer's and the employee's perspective.

2.1.3 | Strengths and limitations of data used

The main virtue of these datasets is the availability of information about the tasks performed by individuals in their own work that are comparable across a wide range of countries. The main limitation is that they are cross-sectional. In particular, note that although both STEP and PIAAC surveys were conducted over multiple rounds across years, only one country (the USA) was surveyed twice with different sets of respondents. In the absence of a panel, we are limited in the mechanisms that we explore. For instance, we cannot control for additional individual heterogeneity outside the characteristics that we observe, nor can we speak to the dynamics of task requirements and wages.

2.2 | Measuring establishment size, occupational task content and wages

2.2.1 | Establishment size and the presence of large establishment gaps

Both datasets provide a measure of workplace or establishment size based on the number of employees, reported in bins. We define establishments that have at least 50 employees as large.[6] The establishment size gaps that we measure in the main analyses compare workers in establishments that have at least 50 employees to workers in establishments with fewer than 50 employees. The qualitative results are robust to the use of alternative cut-offs for the definition of 'large'.

2.2.2 | Occupations

Occupations typically refer to a 'set of jobs whose main tasks and duties are characterized by a high degree of similarity'.[7] In statistical analyses within economics, occupation classifications also serve as a convenient way to summarize heterogeneity across jobs. Though they are useful, we show that these occupational classifications may mask heterogeneity in within-occupation task intensity that, in turn, could help to explain within-occupation wage inequality.

We focus on within-occupation variation as defined by the widely used 2008 version of the International Standard Classification of Occupations (ISCO-08) by the International Labour Organization. The ISCO-08 is a four-level hierarchical, nested classification that is coded using four digits. The first digit refers to one of the 10 'major groups' that typically reflect skill complexity. Each 'major group' is divided into several 'sub-major groups' (second digit), which in turn are divided into one or more 'minor groups' (third digit), which in turn are divided into one or more 'unit groups' (fourth digit). There are a total of 43 sub-major groups, 130 minor groups, and 436 unit groups. The most detailed occupational code available in the PIAAC and STEP surveys has three digits.

For parsimony, our preferred specifications use 2-digit ISCO-08 codes, which we believe balances the trade-off between the specificity of the occupations and sample size concerns. Though other datasets often include occupation codes at a more granular level, a number of papers that study job polarization and tasks in the labour market discuss occupations at a similar aggregation to ours (e.g. Spitz-Oener 2006; Goos *et al.* 2014; Fonseca *et al.* 2018). We show that the results are quantitatively similar using 3-digit ISCO-08 codes.

2.2.3 | Task content of occupations

We follow the approach in Caunedo *et al.* (2023) to construct task measures that are comparable to well-established definitions in the literature (e.g. Autor *et al.* 2003; Acemoglu and Autor 2011). We distinguish five task components of occupations: non-routine analytical (NRA), non-routine interpersonal (NRI), routine cognitive (RC), routine manual (RM), and non-routine manual (NRM). Non-routine analytical tasks involve reading and thinking creatively. Non-routine interpersonal tasks require interacting with others (e.g. through advising, negotiating, teaching). Routine cognitive tasks require structured repetition of activity planning and time management. Routine manual tasks involve physically demanding activities. Finally, non-routine manual tasks involve manual dexterity. The exact variables employed are listed in Appendix Table A1.

We create an individual index measuring the intensity of a particular task category in two steps. First, we standardize the responses to each task variable to have within-country mean zero and standard deviation unity. Second, to obtain the index for a task category, we average the standardized responses to the task variables, and re-standardize the result to again have within-country mean zero and standard deviation unity. By construction, these measures are interpreted as intensities in units of standard deviations relative to the country mean. In Online Appendix OA-C, we consider an alternative construction of task intensity indices using multiple correspondence analysis.

2.2.4 | Usage of ICT

A particular focus of our paper is on documenting the intensity with which workers use technologies such as computers and specific software as part of their work. Though we report results relating to the usage of ICT alongside the task dimensions discussed above, we do not consider it mutually exclusive to those tasks. Rather, *we interpret the use of ICT as a means through which the tasks are performed*. In Appendix Table A1, we show the questions in the surveys that are relevant to measure the use of ICT. We create an index in a similar manner as for the above measures of task content.

2.2.5 | Wages

STEP's worker-based survey reports log hourly earnings in USD. The PIAAC survey elicits hourly wages (in levels of the domestic currency), but for some countries wages are reported only in bins.[8] To quantify the establishment size wage gap, we focus on log hourly wages in non-self-employment work. We deflate the values to 2018 local currency, and use 2018 exchange rates to US dollars to convert earnings to real 2018 USD.

2.2.6 | Demographics and additional controls

To increase the comparability of demographic variables across surveys, we first consider the following standard controls: gender, age block (10-year groups starting from age 16 and ending at 65), and three education categories based on ISCED 2008—(i) primary education or less (ISCED 1), (ii) up to a professional tertiary education degree (ISCED 5), and (iii) bachelor's degree and above (ISCED 5A and beyond).

We further aim to better account for the potential sorting of workers with higher ability or higher non-cognitive skills into larger establishments. In terms of cognition, for STEP countries, we standardize, at the country level, the proportion of correct responses over the total number of questions in three different linguistic tests (vocabulary, sentence and passage). For PIAAC countries, we use the first imputation in both the numeracy and literacy competencies, and we verify that the results hold when employing item response theory over the ten imputations available in the survey (Khorramdel *et al.* 2020). In terms of non-cognitive abilities, STEP provides pre-constructed measures for the following traits: openness, stability, agreeableness and grit. We employ the (standardized) first principal component from these four dimensions. Though PIAAC is known for being less well-equipped for measuring individual non-cognitive traits, we still use a number of measures that have previously been shown to predict earnings. In particular, we follow Anghel and Balart (2017) in using measures of cultural engagement, social trust and political efficacy, and we follow Cabrales *et al.* (2014) in employing a measure of motivation for learning. We combine these four measures by taking their first principal component to proxy for the respondent's non-cognitive skills.

We consider the following sectoral classification for STEP countries: (i) agriculture, fishing and mining; (ii) manufacturing and construction; (iii) commerce; and (iv) other services. For PIAAC, we use more detailed information encompassing 21 different industries.

Finally, in additional specifications reported in the Appendix, we account for regional variation in industrial and demographic composition by employing regional fixed effects, which are always available in STEP but are missing for a subset of PIAAC countries (Italy, Norway and the USA). Importantly, the nature of these regions changes across countries, even within the same survey. For instance, among STEP countries, regions refer to metropolitan areas in Colombia, while in China (Yunnan), they pertain to census enumeration areas. In PIAAC countries, the geographical information corresponds to OECD TL-2 territorial levels (representing the first administrative tier of sub-national government), which are defined politically. Given that this hinders the interpretation of these fixed effects and that they are available for only a subset of countries, we consider the inclusion of regional controls as a robustness check rather than as part of our main specification.

2.3 | Summary statistics

We report summary statistics for the 36 countries for which we have a continuous measure of wages in Online Appendix Table OA-A1. To ensure that our results are not driven by extrapolation, we also impose that each establishment size and 2-digit occupation code cell has at least five observations. The number of observations after focusing on working-age individuals who are not self-employed varies from 857 (Greece) to 3984 (UK) among PIAAC countries, and from 360 (Laos) to 1289 (Vietnam) among STEP countries. In general, small establishments are more prevalent, but there is significant cross-country variation. For instance, in Belgium and the Netherlands 50% of the establishments are large, while the share is around 20% in Ecuador and Greece.

3 | ESTABLISHMENT SIZE GRADIENT IN THE TASK CONTENT OF JOBS

In this section, we document from various perspectives our novel stylized fact that even within narrowly defined occupation groups, there are significant differences in the task composition of jobs across workers of establishments of different sizes. To quantify such a gradient, we estimate versions of the regression

$$T_i = \beta \times \mathrm{LE}_{j(i)} + X_i'\gamma + \delta^o_{o(i)} + \delta^c_{c(i)} + \varepsilon_i, \tag{1}$$

where T_i is the measure of task content of the job of worker i, $\mathrm{LE}_{j(i)}$ is an indicator of whether the establishment $j(i)$ of individual i has at least 50 employees, X is a vector of individual- and establishment-level characteristics, and δ^o and δ^c are occupation-code and country fixed effects, respectively. We focus on our five main task categories (NRA, NRI, RC, RM, NRM) as well as the use of ICT as outcomes. Our coefficient of interest is β, which captures the average difference in intensity in doing task T between two observably equivalent workers in the same country and occupation who differ in that one is employed by a large establishment and the other by a small one.[9] We report standard errors clustered at the country level.

In panels A and B of Table 1, we report estimates of β in equation (1) using the pooled samples of PIAAC and STEP countries, respectively. In column (1), we present the establishment size gradient in task intensity controlling only for occupation and country fixed effects. Qualitatively, we find that workers in larger establishments perform more non-routine analytical tasks and make more intensive use of ICT. We also find suggestive evidence that workers in larger establishments perform more routine cognitive tasks, a pattern that is more evident in the STEP pooled sample than in the PIAAC one.[10]

We do not find a difference in the intensity with which manual tasks, either routine or non-routine, are performed between workers in larger and smaller establishments. The establishment size gradient in non-routine interpersonal tasks is concentrated among PIAAC countries (we return to this later when we document that the gradients for these tasks appear to be more country-specific). Note that by controlling for occupation fixed effects at the 2-digit level based on the ISCO-08 classification, we account for the possibility that the occupational structure of large and small establishments differs in a way that could explain these patterns.

In column (2) of Table 1, we build on our baseline results to account for potential confounders of the establishment size gradient. We include industry fixed effects to rule out the possibility that the gaps are driven by larger establishments disproportionately concentrating in industries that use certain tasks more intensively, or have establishments that are more productive. Moreover, we introduce individual controls for education, gender and age to account for the most salient sources of worker selection into establishments that are typically recorded in standard datasets. Finally, we take advantage of the availability of measures of cognitive and non-cognitive skills in the surveys that we use to show that the task gap barely changes after their inclusion, which reinforces the idea that worker selection cannot fully explain the gradient in task requirements. We find qualitatively similar results, which suggests that our results do not simply arise from differential selection of workers into small and large establishments.[11]

In terms of economic magnitude, focusing on column (2) of Table 1, we find that the average worker in a large establishment performs around 11.7% and 6.6% of a standard deviation more non-routine analytical tasks than a worker in a small establishment in the PIAAC and STEP countries, respectively. We interpret this standard deviation as relative to the country-specific distribution of performed tasks. The average worker in the large establishment also uses 13.2% and 12.5% of a standard deviation more ICT compared to the average worker in smaller establishments in STEP and PIAAC countries, respectively. In terms of routine cognitive tasks, the average large-establishment worker in PIAAC countries performs 3.5% of a standard deviation more than their counterparts in smaller establishments; the difference is larger, at 19.1% of a standard deviation, in the STEP sample. The size of these gaps is comparable to that present between the countries at the two extremes of the development spectrum, as measured by GDP per capita (Caunedo *et al.* 2023).

TABLE 1 Pooled estimates of establishment size gradient in the task content of jobs.

Outcome variable	(1)	(2)	(3) $\hat{\delta}$
Panel A: PIAAC			
Non-routine analytical	0.155***	0.117***	1.488
	(0.019)	(0.017)	
Non-routine interpersonal	0.074***	0.055**	1.427
	(0.022)	(0.022)	
Routine cognitive	0.008	0.035**	−0.632
	(0.015)	(0.014)	
Routine manual	−0.006	0.011	−0.149
	(0.015)	(0.010)	
Non-routine manual	−0.007	0.001	−0.030
	(0.017)	(0.014)	
Use of ICT	0.156***	0.132***	1.525
	(0.010)	(0.010)	
Sample size	65,151	65,151	
Panel B: STEP			
Non-routine analytical	0.125***	0.066**	0.395
	(0.032)	(0.023)	
Non-routine interpersonal	−0.015	−0.035	−0.601
	(0.030)	(0.034)	
Routine cognitive	0.169**	0.191***	−4.020
	(0.067)	(0.060)	
Routine manual	−0.020	0.007	−0.068
	(0.034)	(0.033)	
Non-routine manual	−0.048	−0.060	−4.489
	(0.038)	(0.035)	
Use of ICT	0.194***	0.125**	0.696
	(0.057)	(0.043)	
Sample size	8339	8339	
Controls			
Country fixed effects	Yes	Yes	
2-digit occupation fixed effects	Yes	Yes	
Industry fixed effects		Yes	
Individual demographics		Yes	
Individual cognition/non-cognition		Yes	

Notes: Regressions of a given category of task requirement intensity on an indicator of large establishment (at least 50 employees). Each row refers to a specific task category. Additional controls are indicated in the lower part of the table. Individual demographics include education, gender and age. Regressions are conducted separately for the pooled sample of PIAAC and STEP countries in panels A and B, respectively. Standard errors are reported in parentheses and clustered at the country level. Column (3) reports the estimated $\hat{\delta}$ from Oster (2019).
***, **, * indicate $p < 0.01$, $p < 0.05$, $p < 0.1$, respectively.

3.1 | Robustness to selection on unobservables

The results above show that the gaps in skill use persist after the inclusion of a rich set of observable covariates. Still, the cross-sectional nature of the data limits the amount of unobserved individual heterogeneity that we can control for. To show that the facts that we uncover are unlikely to be driven by selection on unobservables, we follow Oster (2019), who suggests estimating the value of a parameter, denoted as δ, indicating how much stronger/weaker selection on unobservables would have to be, relative to selection on observables, to render the coefficient of interest indistinguishable from zero.[12] A value $\delta = 1$ indicates that selection on unobservables would have to be as large as selection on observables to make the coefficient of interest equal zero. Since one would typically believe that the included controls are capable of explaining a large fraction of the variation in the outcome, Oster (2019) suggests that $\delta \geq 1$ is a rule of thumb to be confident that selection on unobservables is not a large issue. The estimated $\hat{\delta}$ values for the different establishment size gradients in tasks are reported in column (3) of Table 1. We find that the estimated δ values for the tasks for which we identified quantitatively significant gaps comfortably satisfy the proposed rule of thumb.[13]

3.2 | Alternative treatment of occupations

Our main interest is in documenting within-occupation heterogeneity in task composition. While we have demonstrated the robustness of our results to the use of various degrees of specificity in the occupation codes (2-digit and 3-digit), in Table 2 we take an even more flexible approach. In particular, we estimate our preferred specification separately for subsamples defined by 1-digit occupation codes, while still controlling for dummies of 2-digit occupation codes. This approach has the conceptually attractive feature of allowing for more precise within-occupation comparisons of task intensity between establishments of different sizes by considering subsamples that contain more similar occupations.[14] We find that workers in larger establishments across virtually all 1-digit occupation codes perform more non-routine analytical tasks and use more ICT in their work, within both PIAAC and STEP countries, as is consistent with the strong effects that we found in the pooled results. Not surprisingly, there are differences in the magnitude of the establishment size gradient depending on which 1-digit occupation code we focus on. For instance, among PIAAC countries, the largest gaps in the performance of non-routine analytical tasks and in the use of ICT are seen among services and sales workers. Looking at the undertaking of routine cognitive tasks, we find that that the establishment size gradient in tasks is driven by elementary occupations and clerical support workers in the PIAAC sample. Finally, for most 1-digit occupation categories, in both the PIAAC and STEP samples, we find small establishment size differences in the performance of routine and non-routine manual tasks—as we found in our main specification.

3.3 | Cross-country comparisons

We complement the pooled analysis from Table 1 by exploring the presence of the establishment size gradient separately for each country. We report the results in Figure 1. What we find aligns well with our previous results: (i) the larger reliance on non-routine analytical tasks and the use of ICT is present in virtually all countries in our sample; and (ii) the higher intensity on non-routine interpersonal and routine cognitive tasks, while prevalent throughout many countries, also features a subset of countries for which the effects are not distinguishable from zero. Appendix Figure A1 shows that the differences in the performance of routine and non-routine manual tasks are mostly indistinguishable from zero and, if any, are negative.

TABLE 2 Establishment size gradient in the task content of jobs, by 1-digit occupation codes.

1-digit ISCO-08 category	Task category NRA	NRI	RC	RM	NRM	ICT	No. of observations
Panel A: PIAAC							
Managers	0.087	−0.003	−0.010	−0.020	0.021	0.036	4157
	(0.052)	(0.042)	(0.024)	(0.047)	(0.060)	(0.033)	
Professionals	0.088***	0.051**	−0.045**	−0.016	−0.082	0.085***	13,472
	(0.026)	(0.024)	(0.020)	(0.040)	(0.063)	(0.014)	
Technicians & associate professionals	0.132***	−0.036	0.021	−0.035	0.059	0.086***	9601
	(0.027)	(0.025)	(0.025)	(0.038)	(0.038)	(0.027)	
Clerical support workers	0.050	0.040	0.112***	−0.111**	0.011	0.096***	7503
	(0.044)	(0.035)	(0.040)	(0.050)	(0.040)	(0.025)	
Services & sales workers	0.167***	0.160**	−0.022	0.103**	0.069*	0.223***	12,175
	(0.031)	(0.061)	(0.059)	(0.042)	(0.035)	(0.030)	
Craft & related trade workers	0.076	−0.022	0.087	−0.032	−0.084**	0.190***	6344
	(0.093)	(0.072)	(0.109)	(0.039)	(0.040)	(0.064)	
Plant & machine operators, & assemblers	0.163*	0.060	0.051	−0.022	0.021	0.179***	5244
	(0.080)	(0.066)	(0.108)	(0.024)	(0.041)	(0.028)	
Elementary occupations	0.113***	0.131***	0.124***	0.120***	0.027	0.200***	6132
	(0.023)	(0.043)	(0.039)	(0.042)	(0.056)	(0.038)	
Panel B: STEP							
Managers	0.150	−0.015	0.369**	−0.170	−0.128	0.260	451
	(0.131)	(0.152)	(0.142)	(0.150)	(0.160)	(0.202)	
Professionals	0.017	0.068	0.074	−0.046	−0.095**	0.177*	2142
	(0.032)	(0.061)	(0.048)	(0.055)	(0.038)	(0.085)	
Technicians & associate professionals	0.149*	0.271*	0.058	0.150**	−0.027	−0.052	788
	(0.079)	(0.126)	(0.123)	(0.060)	(0.096)	(0.067)	
Clerical support workers	0.035	−0.126	0.060	0.069	0.088	0.198***	895
	(0.067)	(0.098)	(0.101)	(0.038)	(0.131)	(0.038)	
Services & sales workers	0.202***	0.047	0.223***	−0.117	0.009	0.218*	1847
	(0.062)	(0.080)	(0.065)	(0.075)	(0.084)	(0.106)	
Craft & related trade workers	0.107	−0.214*	0.275**	0.135	0.121	0.093	687
	(0.122)	(0.105)	(0.092)	(0.104)	(0.095)	(0.119)	
Plant & machine operators, & assemblers	0.286**	−0.122	0.194	−0.024	−0.295**	0.139*	588
	(0.108)	(0.068)	(0.111)	(0.075)	(0.120)	(0.075)	
Elementary occupations	0.087	0.204*	0.150	0.042	−0.064***	0.100*	941
	(0.064)	(0.102)	(0.150)	(0.107)	(0.018)	(0.049)	

Notes: Coefficient of an indicator for large establishment in a regression of task content intensity on indicator of large establishment (at least 50 employees) by 1-digit ISCO-08 occupation codes and the set of controls in column (2) of Table 1. We do not report the 1-digit categories corresponding to armed forced occupations and skilled agricultural, forestry and fishery workers due to small sample size. Standard errors are reported in parentheses and clustered at the country level.
***, **, * indicate $p < 0.01$, $p < 0.05$, $p < 0.1$, respectively.

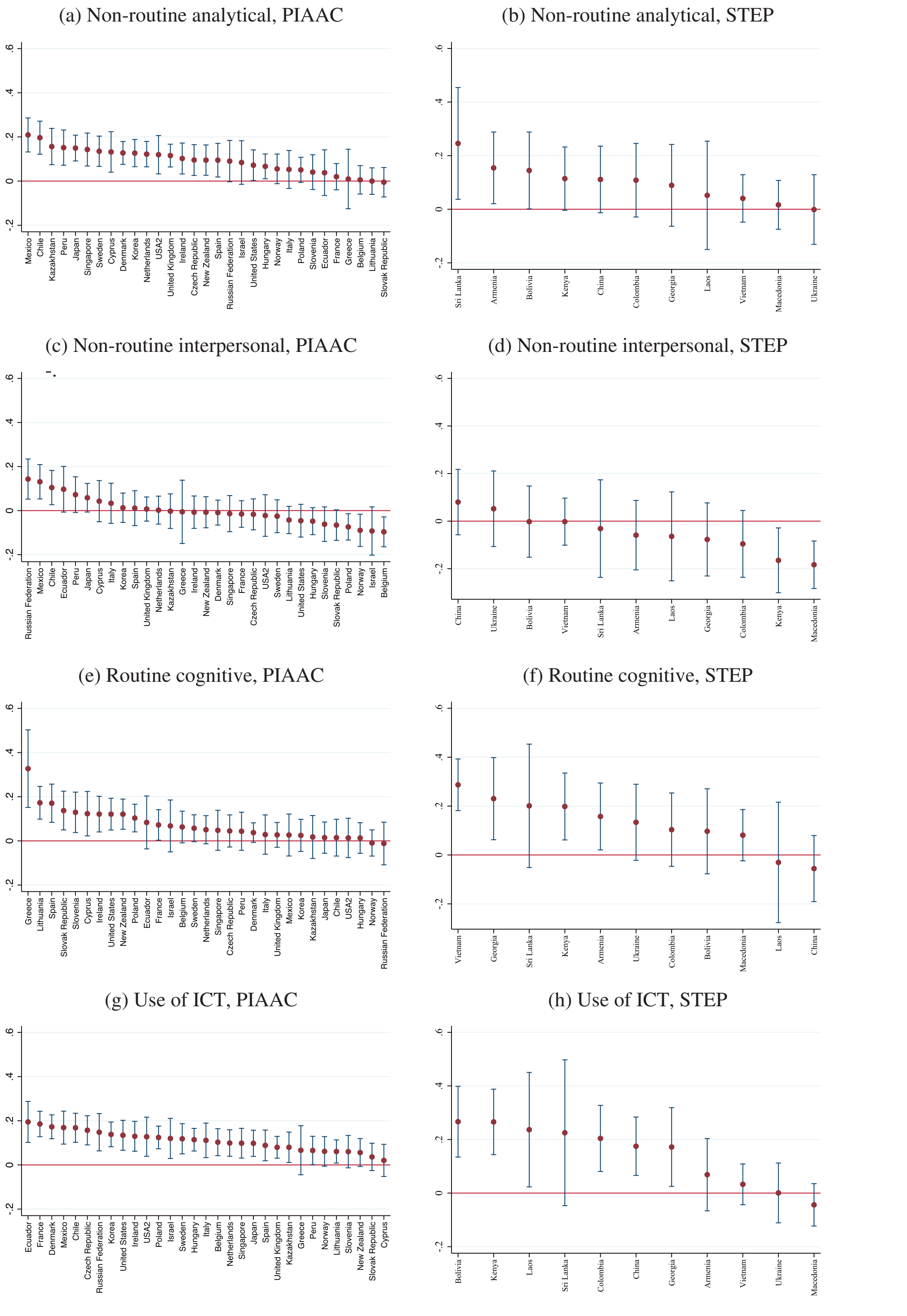

FIGURE 1 Establishment size gradient in the task content of jobs by country, within 2-digit occupations. *Notes*: Coefficient of an indicator for large establishment in a regression of task content intensity on indicator of large establishment (at least 50 employees) and the set of controls in column (2) of Table 1. Regressions are done separately for each country. Countries are ordered by decreasing point estimates. 'United States' refers to the survey conducted in 2012, and 'USA2' to the survey conducted in 2017. Reported confidence intervals at 95% confidence level computed using heteroscedasticity-robust standard errors.

3.4 | Country-level correlates of the establishment size task gradient

Although the qualitative patterns uncovered above are fairly similar across countries, some quantitative differences arise. There are several reasons why this might be the case, including: (i) differences in labour market institutions; and (ii) differences in the relevance of firms with at least 50 employees. We explore whether these differences can be explained by the level of development across countries, focusing on two indicators: log GDP per capita, and the proportion of the population that has completed at least tertiary-level education. In Appendix Figures A2 and A3, we plot the establishment size gradients in the task content of jobs against log GDP per capita and the fraction of population with at least tertiary education, respectively. We highlight two empirical patterns. First, the establishment size differences in the use of non-routine analytical tasks are uncorrelated with log GDP per capita and only slightly positively correlated with the proportion of the population that is at least tertiary educated. This suggests that this pattern is not driven by economic development. In contrast, the establishment size gradient in the performance of routine cognitive and the use of ICT is negatively correlated with economic development—in richer countries and countries with a more educated population, the establishment size differences in the task content of jobs are smaller.

3.5 | Differences in the *distribution* of task intensity by establishment size

So far, our results document average differences in task composition of occupations between establishments of different sizes. In Online Appendix OA-D, we extend our analysis by also exploring the differences in the distribution of task intensity. For this, we employ distributional regressions in the spirit of Chernozhukov *et al.* (2013), and show that the large establishment differences in non-routine analytical tasks and in the use of ICT are present at multiple thresholds throughout the support of the distribution. This demonstrates that the mean differences that we find are not driven solely by discrepancies in the upper tail of the distribution of tasks performed. Instead, the distributions of the intensity of performed tasks in larger establishments are all shifted to the right compared to those in smaller establishments. For the case of routine and non-routine manual tasks, the insignificant differences are seen throughout the intensity distribution, except for the case of routine manual tasks for large establishments, where there is suggestive evidence of a widening of the intensity distribution in larger firms.

3.6 | When does the gradient arise?

We provide two pieces of evidence to suggest that the establishment size gradient in the task content of jobs does not arise from larger employers exclusively assigning more non-routine analytical tasks and usage of ICT to workers with longer tenure in the establishment. Rather, this establishment size gradient is already reflected in the labour demand of the employers and thus appears at the beginning of the job tenure and early in the workers' careers.

First, in Appendix Table A4, we re-estimate our preferred specification conditioning first on workers having been in their current job for a short period of time (up to 2 years) and then additionally on being young (less than 25 years old). This excludes workers who may have adopted more task-intensive work as they progressed in their careers. Among these young workers with short tenure, we find establishment size gradients in non-routine analytical tasks and the use of ICT that are of the same sign and comparable in magnitude to the full sample.

Second, we leverage the availability of employer surveys from the World Bank STEP skills surveys programme to examine whether the task requirements of firms already differ at the time of hiring.[15] In these surveys, employers answer a limited set of questions on the skill requirements of occupations in the workplace. Based on the questions asked in the survey, we are able to identify only tasks that pertain to the following categories: (i) non-routine analytical tasks and (ii) use of ICT. To limit the burden on the survey respondent, STEP elicits only two occupations (randomly selected out of nine categories). Appendix Table A5 reports estimates of average differences in task requirements between large and small employers, within occupation categories, for the pooled sample of nine countries where the STEP employer survey is available. We again find that large employers require more non-routine analytical tasks and use more ICT as early as in the hiring stage. This further suggests that the origins of the gradients arise from systematic differences in how production or technology is organized by employer size rather than from how workers are able to accumulate more specialized tasks over their tenure or career.

3.7 | Discussion: sources of the establishment size gradient in task content

The more fundamental question remains: *why* do workers in the same occupation engage in different task intensities depending on their establishment's size? To the best of our knowledge, there is no unified framework that may be used to exhaustively explore the potential drivers of the establishment size gradient, so in this subsection we focus on providing several well-grounded plausible explanations, relate them to our empirical findings, and highlight that further theoretical and empirical analyses will be needed to shed more light on this question.

To make the exposition easier to follow, let us take a concrete example. Our results suggest that an accounting professional in a larger establishment performs non-routine analytical tasks more intensely and uses more ICT than another accounting professional in a smaller establishment, even within the same industry and controlling for the innate skills of the accountants. This is consistent with Adenbaum (2023), who argues that larger, more productive establishments are able to organize production such that workers are more specialized. The accountant in the small establishment may need to perform additional administrative tasks, such as sorting mail or answering phones, whereas in larger establishments, specific workers are hired to perform these administrative tasks, leaving the accountant to focus on the tasks in which he specializes, which tend to be more non-routine analytical and use more ICT. We find some support towards this story in that, for certain occupation categories, we document that workers in larger establishments display higher intensity in some task categories at the expense of engaging less in other task categories. However, our pooled results from Table 1 show that on average, workers in larger establishments perform more non-routine analytical tasks and use more ICT without significantly decreasing their performance of other task categories.

A second plausible rationale is rooted in the idea that establishment size does not only relate to organizational capacity. Larger establishments are naturally more complex. The increased complexity in larger establishments may lead workers in the same occupation to perform tasks with different intensities to obtain the same goal. The accountant in the small firm will come across the same financial transactions in their day-to-day work as the accountant in the large establishment, but the latter is more likely to encounter complex financial transactions in their day-to-day work that involve more non-routine analytical work.

To deal with such complexity, larger establishments tend to adopt technologies that complement the scale of their operations (e.g. Alekseeva *et al.* 2021; Lashkari *et al.* 2024). While the accountant in the smaller establishment may just need a simple spreadsheet to prepare the employer's financial statements, the accountant in the larger establishment would be required to use more specialized accounting software (which helps the larger establishment to process more complex transactions, facilitates communication with other relevant personnel, and might improve replicability). One can make similar arguments for other occupations. For instance, the most complex equipment used by a family bakery might be an industrial oven, while a baker in a large bread manufacturing plant has to work with complex equipment throughout the whole bread-making process.

Regardless of where the firm size gradient in task content comes from, we have shown that there is heterogeneity in the task content of jobs, even within the same occupation. This heterogeneity may be fundamental in helping us to better understand labour markets. For instance, in the next section, we show that the differences in task content between large and small establishments can explain an economically significant portion of the wage gaps observed between workers in larger and smaller establishments.

4 | THE ESWP AND THE ROLE OF INDIVIDUAL SELECTION, SECTORS AND TASKS

In this section, we first document in Subsection 4.1 the presence of an ESWP— both on average and throughout the wage distribution—using the pooled PIAAC and STEP samples separately. We then explore in Subsection 4.2 how much of this raw gap can be explained linearly by various mechanisms, including selection of individuals into occupations, and differences in the task composition of occupations.

4.1 | The ESWP: cross-country evidence

Similar to how we documented the establishment size gap in the task content of jobs in Section 3, in this subsection we explore the establishment size gap in wages. We estimate the regression

$$\ln w_i = \beta \times \mathrm{LE}_{j(i)} + X_i'\gamma + \delta^o_{o(i)} + \delta^c_{c(i)} + \varepsilon_i, \tag{2}$$

where $\ln w_i$ is log real hourly wages in 2018 USD of individual i. The rest of the specification is similar to equation (1). We interpret β as a measure of the ESWP, which is how much more workers with similar observables in establishments with at least 50 employees are paid (in log points) relative to those in smaller establishments within the same occupation and country.

The estimated ESWP controlling only for country and 2-digit ISCO-08 occupation code fixed effects (column (1) of Table 3) is measured to be about 16.5% and 25.7% for PIAAC and STEP countries, respectively.[16] These results show that the establishment size wage gaps are not perfectly explained by differences in the occupational structure of establishments. In column (2), we saturate the regression with additional controls, including our proposed mediator, the establishment size gradient in tasks.[17] Though the estimated ESWP falls after the inclusion of these possible mediators, we find that there is still a substantial ESWP left unexplained. In the next subsection, we explore how much of the explained ESWP can be attributed to the different mediators considered.

Expanding on the existing literature, we explore the ESWP beyond the comparison of *average* wages between large and small establishments. In columns (3)–(5) of Table 3, we report the β coefficients in a quantile regression version of equation (2) at the 10th, 50th and 90th quantiles, controlling for tasks, industry fixed effects and individual controls. We find that the worker at the 10th percentile in larger establishments is paid 0.117 and 0.216 log points more than the worker at the 10th percentile in smaller establishments in PIAAC and STEP, respectively. For the median worker, the difference is about 0.106 and 0.143 log points, and it is 0.085 and 0.056 log points at the 90th percentile. This suggests that the entire wage distribution of large establishments is shifted to the right compared to smaller establishments, even within occupations.

TABLE 3 Pooled estimates of the ESWP.

	Mean regressions		Quantile regressions		
	(1)	(2)	(3) p10	(4) p50	(5) p90
PIAAC					
ESWP	0.153***	0.111***	0.117***	0.106***	0.085***
	(0.014)	(0.012)	(0.010)	(0.010)	(0.012)
Sample size	54,782	54,782	54,782	54,782	54,782
STEP					
ESWP	0.229***	0.185***	0.216***	0.143***	0.056***
	(0.028)	(0.025)	(0.034)	(0.022)	(0.030)
Sample size	8339	8339	8339	8339	8339
Controls					
Country fixed effects	Yes	Yes	Yes	Yes	Yes
2-digit occupation fixed effects	Yes	Yes	Yes	Yes	Yes
Tasks		Yes	Yes	Yes	Yes
Industry fixed effects		Yes	Yes	Yes	Yes
Individual demographics		Yes	Yes	Yes	Yes
Individual cognition/non-cognition		Yes	Yes	Yes	Yes

Notes: Columns (1) and (2) show regressions of log hourly wages (in 2018 USD) on an indicator of large establishment (at least 50 employees) and the set of controls in columns (1) and (2) of Table 1, respectively. The PIAAC countries for which continuous wage data are not available are excluded. Appendix Table A6 additionally reports the point estimates and standard errors for the various tasks and computer use, which are not reported in the present table for brevity. Columns (3)–(5) show results from quantile regressions under the specification in column (2). Standard errors are reported in parentheses and clustered at the country level.
***, **, * indicate $p < 0.01$, $p < 0.05$, $p < 0.1$, respectively.

In Online Appendix OA-E, we provide a richer analysis of the ESWP. We find that the existence of an economically significant ESWP is present when we look at more detailed occupational categories or at individual countries.

4.2 | Sources of the ESWP

There are a number of plausible reasons for the existence of the ESWP. In this subsection, we explore the role in wage determination of sorting by (i) individual characteristics, (ii) industry characteristics, and (iii) differential task content of jobs. To quantify their relative importance, we conduct a simple mediation analysis adopting the two-step conditional decomposition developed in Gelbach (2016). A desirable feature of his approach is that the results from the decomposition are independent of the order in which the mediators are introduced in the regression. A limitation, however, is that we require measurement of the key mediators to avoid omitted variable biases. The decomposition begins with a raw estimate of the ESWP, β^{raw}, from the regression

$$\ln w_i = \beta^{\text{raw}} \times \text{LE}_{j(i)} + \delta^{o,\text{raw}}_{o(i)} + \delta^{c,\text{raw}}_{c(i)} + \varepsilon^{\text{raw}}_i, \tag{3}$$

where $\ln w_i$ is log real hourly wages, $\text{LE}_{j(i)}$ is the indicator for worker i being in a large firm, δ^o are occupation fixed effects (2-digit ISCO code), and δ^c are country fixed effects. This raw ESWP estimate coincides with the estimate in column (1) of Table 3. The second step of

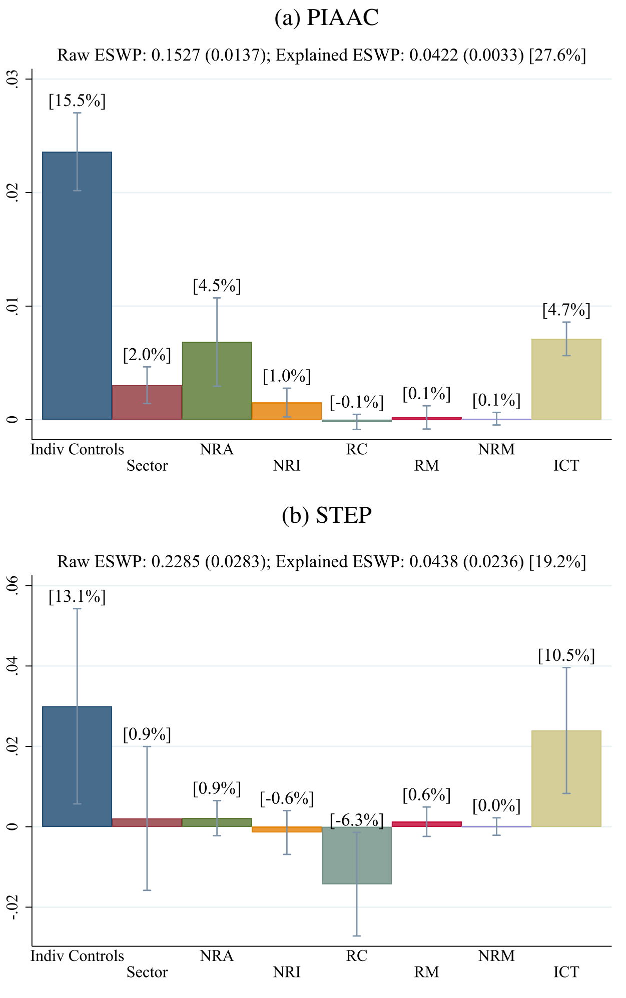

FIGURE 2 Gelbach decomposition of ESWP, pooled. *Notes*: Pooled PIAAC and STEP samples. Raw ESWP refers to the estimate in column (1) of Table 3. Explained ESWP is the difference between the estimates in columns (1) and (2) in that table. The vertical axis is the amount of the ESWP explained by the corresponding component. Numbers in brackets indicate percentages of the raw ESWP. Reported confidence intervals at 95% confidence level. Standard errors are clustered at the country level.

the decomposition consists of re-estimating the above equation after the inclusion of a set of individual controls X_i that are believed to mediate the ESWP:

$$\ln w_i = \beta^{\text{full}} \times \text{LE}_{j(i)} + X_i'\gamma + \delta^{o,\text{full}}_{o(i)} + \delta^{c,\text{full}}_{c(i)} + \varepsilon^{\text{full}}_i. \tag{4}$$

In our case, the variables incorporated in X are (i) individual characteristics, including sex, age and education, (ii) sector dummies, and (iii) the task content of jobs and usage of ICT reported by workers. In other words, this regression replicates the specification in column (2) of Table 3. The difference $\beta^{\text{raw}} - \beta^{\text{full}}$ is interpreted as the part of the ESWP that we are able to explain by controlling for X. Gelbach (2016) uses the formula for the omitted variable bias to apportion the explained part of the ESWP to each of the component variables of X.[18]

The decomposition results are summarized graphically in Figure 2. We find that the mediators that we consider are able to explain 27.6% of the raw ESWP in PIAAC, and 19.2% in STEP. Individual characteristics (age, sex and education) explain a significant portion of the

ESWP, around 15.5% and 13.1% in PIAAC and STEP, respectively. This suggests that large establishments pay more on average because they hire workers who are older, more educated, and better skilled. This sorting pattern of workers to larger employers has been documented recently by Arellano-Bover (2021). Importantly, though human capital sorting (by occupation or education) explains a large portion, it is unable to fully explain the existence of the ESWP.

The third to eighth bars in both Figures 2(a) and 2(b) report the fractions of the raw ESWP that are explained by the differences in the tasks performed and in ICT use by the workers documented in Section 3. To help us to better interpret the results, in Appendix Table A6, we report the coefficients of the tasks on log wages in the regressions of Table 3. In particular, we document that non-routine analytical tasks, non-routine interpersonal tasks and the use of ICT have positive returns on wages, whereas routine cognitive and routine manual tasks have negative returns, consistent with existing evidence (Autor and Handel 2013; Stinebrickner *et al.* 2019). Something important to notice is that the Gelbach (2016) decomposition estimates the contribution of each mediator *keeping* the other mediators constant. Hence while the different task components may be predictive of wages, the variation that explains the ESWP is largely mediated by the variation in the usage of ICT rather than variation in the task content.

Returning to Figure 2, we find that the establishment size gradient in the performance of non-routine analytical tasks explains about 4.5% of the raw ESWP in PIAAC. Moreover, differences in the use of ICT explain an additional 4.7% of the raw ESWP in PIAAC, and 10.5% in STEP. In STEP countries, apart from the large role of ICT, we find that routine cognitive tasks, which are disproportionately undertaken by workers in large establishments but have sizeable negative returns on wages, explain −6.3% of the gap.

Overall, we take these results to reflect not only the disproportionately higher intensity with which workers perform various tasks and use ICT in larger establishments, but also the growing importance of computer skills (Alekseeva *et al.* 2021) in the labour market.[19] Combined, these task components and the use of ICT explain more than 10% of the raw ESWP, a magnitude comparable to that explained jointly by education, age and sex.[20]

The main concern in the performed decomposition analysis is the potential presence of omitted variables. While we do have a rich set of worker controls (including cognitive and non-cognitive skills that should reduce the potential for the presence of unobserved determinants of worker selection into establishments), not observing establishment characteristics, in particular performance, may be a concern. First, the ESWP may be partially driven by differences in establishment productivity—in many models of the labour market, including rent-sharing models or search and matching models, more productive employers pay higher wages to their workers. Unfortunately, we do not observe measures of employer productivity. To partially address this issue, we control for the sector in which the worker works with the aim of accounting for aggregate productivity differences across sectors. We find that the sectoral membership of the worker only partly explains the existence of the ESWP—about 2% of the raw ESWP in PIAAC countries.

A second concern is that the ESWP may be driven by spatial differences in wages. In an attempt to capture within-country spatial differences in wages, we repeat the decomposition including regional fixed effects, and report the results in Appendix Figure A4.[21] We find that the regional dummies are able to explain a non-negligible fraction of the ESWP (around 9% and 20% in PIAAC and STEP countries, respectively). Importantly, we show that this does not come at the expense of shifting the importance of the tasks performed, as their importance in the decompositions remains about the same size.

4.2.1 | Cross-country comparisons

The results of the decomposition exercise by country are summarized graphically in Online Appendix Figure OA-B2, which focuses on countries for which both the ESWP and the explained portion of the ESWP are statistically significant. We find that the proportion of the raw ESWP explained by the controls that we consider varies between 20% and 40%. In terms of broad patterns, basic individual characteristics such as age, sex and education consistently explain a significant portion of the raw ESWP (between 10% and 30%). Sectoral membership is intermittently statistically and economically significant in a handful of countries. In countries where this component accounts for a statistically significant portion, sectors explain around 5–20% of the raw ESWP.

The establishment size gradients in the performance of non-routine analytical tasks and the use of ICT explain, in general, a total of about 5–20% of the raw ESWP. The establishment size gradient in the performance of non-routine analytical tasks explains between 5% and 8% of the raw ESWP whenever it contributes to a statistically significant share of the ESWP. Among the countries for which the ICT component is statistically significant, the estimates lie mostly between 3% and 9% of the raw ESWP, with a couple of countries where the use of ICT contributes more substantially to the ESWP.[22]

5 | CONCLUSION

In this paper, we document novel stylized facts about the heterogeneity in occupational task intensity across establishments. We find that individuals working in larger establishments report that they perform non-routine analytical tasks more frequently and use ICT more intensively, even within narrowly defined occupations. We complement these empirical facts with demand-side information confirming that larger employers indeed require workers to perform more non-routine analytical and ICT-intensive tasks.

Moreover, we document the existence of an economically significant establishment size wage premium (ESWP) of about 15%. We provide suggestive evidence on the role of task heterogeneity in explaining this ESWP. By controlling for individual characteristics (age, gender, education, cognition and non-cognition) of the workers, sector and task content of jobs, we explain about 28% of the raw ESWP in PIAAC countries, and 19% in STEP countries. Differences in the task contents of jobs are able to explain over 10% of the raw ESWP. Therefore, accounting for within-occupation heterogeneity in the task content of jobs enriches our understanding of wage gaps in the labour market.

We consider that our work opens two natural avenues for future research.

First, an unresolved question is how these task differences arise in a dynamic economy. In the Introduction, we suggested that as employers grow larger, they invest in automation and/or offshore more work, which transforms the organization of production. These larger establishments engage workers in complementary tasks such as non-routine analytical and routine cognitive tasks. Moreover, these tasks are performed with more ICT. While our results are consistent with this micro-founded mechanism of firm dynamics, it is difficult to establish its consistency with reality in the absence of panel data of employers and tasks.

Second, we leave for future study other implications for the labour market of the establishment size task gradient. We have suggestive evidence of its role in static wage determination, but lack exogenous identifying conditions to argue for its causal nature. The implications of our results on dynamic wage determination remain unexplored. More specifically, our results may serve as a nexus between two seemingly parallel strands of the literature. (i) A number of studies show that having experience in certain tasks has different returns in the market: analytical tasks and the use of ICT have been found to have high market returns, especially in recent years (Stinebrickner *et al.* 2019; Alekseeva *et al.* 2021). (ii) There is evidence that experience in large firms also has higher returns in the market (Arellano-Bover 2024). Our results suggest a plausible mechanism for the bigger dynamic returns to working in larger employers—workers in larger firms gain more experience in performing non-routine analytical tasks and the use of ICT, which are highly valued in the labour market.

ACKNOWLEDGMENTS

We thank Manuel Arellano, Jaime Arellano-Bover and Samuel Bentolila for valuable comments and suggestions. We also thank Andrés Carmona for his research assistance. The data used in this paper can be obtained from the websites of the OECD's Survey of Adult Skills (https://www.oecd.org/skills/piaac) and the World Bank's STEP Skills Measurement Program (https://microdata.worldbank.org/index.php/collections/step).

The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Banco de España or the Eurosystem.

ENDNOTES

1 Contemporaneous work by Adenbaum (2023) shows that larger firms hire a wider variety of occupations and make jobs more specialized, which means that workers in the same occupation perform different tasks depending on their firm's size. We provide complementary evidence to show that (i) this pattern is widespread across occupations and countries, and (ii) differences in task intensity within occupations partly explain wage inequality.

2 The USA conducted a second round of data collection to get more reliable estimates for certain subgroups. In the graphs presented in this paper, we label the results based on the first round (conducted in 2012) as 'United States', and those based on the second round (conducted in 2017) as 'USA2'.

3 We exclude Turkey because we cannot construct our measure of non-cognitive skills of the worker. We confirm that the results from specifications that do not control for this variable are similar when including Turkey.

4 Peru lacks this information as well.

5 Though Ghana has a household-based survey that contains the relevant variables, we exclude it because of the small sample size that remains after sample selection. The Philippines was also a survey country, but a different questionnaire was used.

6 The OECD classifies micro enterprises as those with fewer than 10 employees, small enterprises as those with 10–49 employees, medium-sized enterprises as those with 50–249 employees, and large enterprises as those with 250 or more workers. See https://data.oecd.org/entrepreneur/enterprises-by-business-size.htm (accessed 18 November 2024).

7 See https://isco-ilo.netlify.app/en/isco-08 (accessed 18 November 2024).

8 A complete overview on the availability of the task and wage information for each country can be found in Appendix Subsection A.1.

9 Throughout the paper, in the pooled regressions, we use probability weights, adjusted based on the populations of the various countries in 2018 (with the exception of China, for which we use the population of Yunnan—the only province of the country that was surveyed). Intuitively, this weighting approach places more weight on observations from large-population countries.

10 Apart from the gap in economic development between PIAAC and STEP countries, another potential explanation behind this difference is the fact that the subcomponents of the routine cognitive category may capture different skills in PIAAC and STEP. As seen in Appendix Table A1, PIAAC focuses on planning-related activities, while STEP expands the interest to the actual execution of tasks that require routine cognitive skills (e.g. short, repetitive tasks).

11 In Appendix Table A2, we show the estimates when we introduce the confounders sequentially. Moreover, we include two additional columns where we (i) saturate the regression with the interactions of all our controls with country fixed effects to allow the returns to such controls to vary across countries, and (ii) include regional fixed effects to account for spatial differences in the presence of large and small establishments, and in tasks and occupations. We find quantitatively similar results, so we keep column (2) in Table 1 as our preferred specification for parsimony. In Online Appendix Table OA-B1, we find the same qualitative patterns employing 3-digit occupation fixed effects instead. In Online Appendix Table OA-B2, we report the results from the PIAAC sample when we do not discard countries with missing regional information. We do not identify any substantial differences. Appendix Table A3 shows that these findings remain under alternative measures of establishment size.

12 A more detailed explanation of the methodology is given in Online Appendix OA-F.

13 In column (7) of Appendix Table A2, we estimate the corresponding Oster (2019) δ values, where we include industry fixed effects in the baseline regression. The long regression, therefore, adds only individual controls. Relative to selection based on observable individual characteristics only, the estimated δ values are generally smaller in magnitude but still indicate that our main results are robust to selection on unobservables.

14 The main drawback is that the sample size for some 1-digit occupations is limited, particularly among STEP countries, which may affect the precision of our estimates and the extent to which we rely on extrapolation.

15 We acknowledge that the results in this subsection are based on a small number of low- and middle-income countries, so external validity is limited.

16 Note that $\exp(0.153) - 1 \approx 0.165$.

[17] In Appendix Table A6, we report how the estimated ESWP changes as the mediators are introduced sequentially.
[18] A more detailed explanation of the methodology is in Online Appendix OA-F.
[19] A potential concern is that ICT, which explains a large fraction of the ESWP in both samples, is itself a mediator of the role of tasks; that is, after tasks are assigned to workers, ICT use is decided as a function of the tasks. In Online Appendix Figure OA-B1, we replicate the same analysis excluding ICT as a potential mediator. We find that the fraction of the ESWP that tasks can explain is mostly unaffected. This suggests that ICT use is an independent mechanism in itself.
[20] In STEP countries, routine cognitive tasks explain 6% of the closing of the gap in wages between workers in larger and smaller establishments. This is on top of the differences in the use of ICT explaining 10% of the widening gap in wages between workers in larger and smaller establishments.
[21] We note that in STEP, only urban areas are surveyed, which partially alleviates the urban–rural differences that we might expect in wages.
[22] We also report the version of the decomposition where we additionally include regional fixed effects as mediators by country (see Online Appendix Figure OA-B3). The qualitative results remain, and region emerges as a contributor of its own to the ESWP for PIAAC countries.
[23] Variable names are based on the Albanian survey, where 'x' stands for either a or b.

REFERENCES

Acemoglu, D. and Autor, D. (2011). Skills, tasks and technologies: implications for employment and earnings. In D. Card and O. Ashenfelter (eds), *Handbook of Labor Economics*, Vol. **4**. Amsterdam: Elsevier, pp. 1043–171.

——— and Restrepo, P. (2022). Tasks, automation, and the rise in U.S. wage inequality. *Econometrica*, **90**(5), 1973–2016.

Adenbaum, J. (2023). Endogenous firm structure and worker specialization. Working Paper; available online at https://jacobadenbaum.github.io/files/Adenbaum2023.pdf (accessed 25 November 2024).

Alekseeva, L., Azar, J., Giné, M., Samila, S. and Taska, B. (2021). The demand for AI skills in the labor market. *Labour Economics*, **71**, 102002.

Anghel, B. and Balart, P. (2017). Non-cognitive skills and individual earnings: new evidence from PIAAC. *SERIEs*, **8**(4), 417–73.

Arellano-Bover, J. (2021). Who gets their first job at a large firm? The distinct roles of education and skills. *AEA Papers and Proceedings*, **111**, 465–9.

——— (2024). Career consequences of firm heterogeneity for young workers: first job and firm size. *Journal of Labor Economics*, **42**(2), 549–89.

Atalay, E., Sotelo, S. and Tannenbaum, D. (2024). The geography of job tasks. *Journal of Labor Economics*, **42**(4); available online at https://www.journals.uchicago.edu/doi/abs/10.1086/725360 (accessed 25 November 2024).

Autor, D. H. and Handel, M. J. (2013). Putting tasks to the test: human capital, job tasks, and wages. *Journal of Labor Economics*, **31**(S1), S59–S96.

———, Levy, F. and Murnane, R. J. (2003). The skill content of recent technological change: an empirical exploration. *Quarterly Journal of Economics*, **118**(4), 1279–333.

Barro, R. J. and Lee, J. W. (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics*, **104**, 184–98.

Bloom, N., Guvenen, F., Smith, B. S., Song, J. and von Wachter, T. (2018). The disappearing large-firm wage premium. *AEA Papers and Proceedings*, **108**, 317–22.

——— and Van Reenen, J. (2007). Measuring and explaining management practices across firms and countries. *Quarterly Journal of Economics*, **122**(4), 1351–408.

Brown, C. and Medoff, J. (1989). The employer size–wage effect. *Journal of Political Economy*, **97**(5), 1027–59.

Cabrales, A., Dolado, J. J. and Mora, R. (2014). Dual labour markets and (lack of) on-the-job training: PIAAC evidence from Spain and other EU countries. CEPR Discussion Paper no. 10246.

Caunedo, J., Keller, E. and Shin, Y. (2023). Technology and the task content of jobs across the development spectrum. *World Bank Economic Review*, **37**(3), 479–93.

Chernozhukov, V., Fernández-Val, I. and Melly, B. (2013). Inference on counterfactual distributions. *Econometrica*, **81**(6), 2205–68.

Colonnelli, E., Tåg, J., Webb, M. and Wolter, S. (2018). A cross-country comparison of dynamics in the large firm wage premium. *AEA Papers and Proceedings*, **108**, 323–7.

De La Rica, S., Gortazar, L. and Lewandowski, P. (2020). Job tasks and wages in developed countries: evidence from PIAAC. *Labour Economics*, **65**, 101845.

Deming, D. and Kahn, L. B. (2018). Skill requirements across firms and labor markets: evidence from job postings for professionals. *Journal of Labor Economics*, **36**(S1), S337–S369.

Dicarlo, E., Lo Bello, S., Monroy-Taborda, S., Oviedo, A. M., Sanchez Puerta, M. L. and Santos, I. V. (2016). The skill content of occupations across low and middle income countries: evidence from harmonized data. IZA Discussion Paper no. 10224.

Dobbelaere, S. (2004). Ownership, firm size and rent sharing in Bulgaria. *Labour Economics*, **11**(2), 165–89.

Fonseca, T., Lima, F. and Pereira, S. C. (2018). Job polarization, technological change and routinization: evidence for Portugal. *Labour Economics*, **51**, 317–39.
Gelbach, J. B. (2016). When do covariates matter? And which ones, and how much? *Journal of Labor Economics*, **34**(2), 509–43.
Gerlach, K. and Hübler, O. (1998). Firm size and wages in Germany—trends and impacts of mobility. *Empirica*, **25**(3), 245–61.
Goos, M., Manning, A. and Salomons, A. (2014). Explaining job polarization: routine-biased technological change and offshoring. *American Economic Review*, **104**(8), 2509–26.
Jaimovich, N., Zhang, M. B. and Vincent, N. (2023). Under the hood of the routine share decline. *Economics Letters*, **234**, 111437.
Khorramdel, L., von Davier, M., Gonzalez, E. and Yamamoto, K. (2020). Plausible values: principles of item response theory and multiple imputations. In D. B. Maehler and B. Rammstedt (eds), *Large-scale Cognitive Assessment*. Cham: Springer International Publishing, pp. 27–47.
Lashkari, D., Bauer, A. and Boussard, J. (2024). Information technology and returns to scale. *American Economic Review*, **114**(6), 1769–815.
Lehmer, F. and Möller, J. (2010). Interrelations between the urban wage premium and firm-size wage differentials: a microdata cohort analysis for Germany. *Annals of Regional Science*, **45**(1), 31–53.
Lewandowski, P., Park, A., Hardy, W. and Du, Y. (2022). Technology, skills, and globalization: explaining international differences in routine and nonroutine work using survey data. *World Bank Economic Review*, **36**(3), 687–708.
Lochner, B., Seth, S. and Wolter, S. (2020). Decomposing the large firm wage premium in Germany. *Economics Letters*, **194**, 109368.
Lucas, R. E. (1978). On the size distribution of business firms. *Bell Journal of Economics*, **9**(2), 508–23.
Ocampo, S. (2022). A task-based theory of occupations with multidimensional heterogeneity. Centre for Human Capital and Productivity Working Paper no. 2022-2.
Oi, W. Y. and Idson, T. L. (1999). Firm size and wages. In O. C. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, Vol. **3**. Amsterdam: Elsevier, pp. 2165–214.
Oster, E. (2019). Unobservable selection and coefficient stability: theory and evidence. *Journal of Business & Economic Statistics*, **37**(2), 187–204.
Porcher, C., Rubinton, H. and Santamaría, C. (2023). JUE insight: The role of establishment size in the city-size earnings premium. *Journal of Urban Economics*, **136**, 103556.
Reed, T. and Thu, T. (2019). The large-firm wage premium in developing economies. World Bank Policy Research Working Paper no. 8997.
Schaffner, J. A. (1998). Premiums to employment in larger establishments: evidence from Peru. *Journal of Development Economics*, **55**(1), 81–113.
Spitz-Oener, A. (2006). Technical change, job tasks, and rising educational demands: looking outside the wage structure. *Journal of Labor Economics*, **24**(2), 235–70.
Stinebrickner, R., Stinebrickner, T. and Sullivan, P. (2019). Job tasks, time allocation, and wages. *Journal of Labor Economics*, **37**(2), 399–433.
Söderbom, M., Teal, F. and Wambugu, A. (2005). Unobserved heterogeneity and the relation between earnings and firm size: evidence from two developing countries. *Economics Letters*, **87**(2), 153–9.
Troske, K. R. (1999). Evidence on the employer size-wage premium from worker–establishment matched data. *Review of Economics and Statistics*, **81**(1), 15–26.
Velenchik, A. D. (1997). Government intervention, efficiency wages, and the employer size wage effect in Zimbabwe. *Journal of Development Economics*, **53**(2), 305–38.
Winter-Ebmer, R. (2001). Firm size, earnings, and displacement risk. *Economic Inquiry*, **39**(3), 474–86.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: De Vera, M. and Garcia-Brazales, J. (2024). Establishment size and the task content of jobs: evidence from 46 countries. *Economica*, 1–32. https://doi.org/10.1111/ecca.12563

APPENDIX

A.1 Data availability and measurement

A.1.1 Data availability by country

For the reader's convenience, we report data availability for each of the 46 countries employed in our study:

For PIAAC:

- (worker-based) tasks + continuous wages: Belgium, Chile, Cyprus, Czech Republic, Denmark, Ecuador, France, Greece, Ireland, Israel, Italy, Japan, Kazakhstan, South Korea, Lithuania, Mexico, Netherlands, Norway, Poland, Russian Federation, Slovak Republic, Slovenia, Spain, Sweden, UK
- (worker-based) tasks + bin-based wages (hence excluded from the ESWP analyses): Hungary, New Zealand, Peru, Singapore, USA (both rounds)
- (worker-based) tasks and no wages: Peru.

For STEP:

- task intensity based on worker-based survey + wages: Bolivia, China (Yunnan province), Colombia, Georgia, Laos, Macedonia, Sri Lanka, Ukraine
- task intensity based both on worker- and employer-based surveys + wages: Armenia, Kenya, Vietnam
- task intensity only based on employer-based surveys (and no wages): Albania, Azerbaijan, Bosnia-Herzegovina, Kosovo, Serbia.

A.1.2 Construction of task content measures using worker-based surveys

Table A1 summarizes the mapping, following Caunedo *et al.* (2023), of the questions in PIAAC and STEP to the different task dimensions in which we are interested: non-routine analytical, non-routine interpersonal, routine cognitive, routine manual, non-routine manual, and the usage of ICT.

Our preferred choice of variable construction standardizes each subcomponent of a task category to have mean 0 and standard deviation 1 across all the respondents in a given country. Then all the subcomponents of the category are used to obtained the simple average (i.e. equal weights assigned to each subcomponent). The resulting mean for each task category is once again standardized to have mean 0 and standard deviation 1 across the respondents in the country.

A.1.3 Construction of task content measures using employer-based surveys

In the employers' questionnaire, a knowledgeable person was asked about the task requirements for two randomly selected occupations. One of them would be chosen among manager/professional/technician, while the second would be selected from clerk/services/sales/crafting/operator/elementary occupation. We are able to obtain a measure of employer-specific demand that closely matches the construction of our non-routine analytical and ICT use measures from the workers' survey, but not for the other task categories. We follow the same procedure as for the worker-based skill intensity measures, and construct employer requirements as the standardized value (mean 0 and standard deviation 1) of the simple mean of the standardized scores in each of the following questions.

- Non-routine analytical. (a) Does the job involve reading? (m_30x_1) (b) Does the job involve writing using correct spelling and grammar? (m_30x_2) (c) Does the job involve maths? (m_30x_3) (d) Does the job involve solving problems that take 30 minutes or more to solve? (m_30x_4) (e) Does the job involve speaking other languages? (m_30x_5).[23] Possible answers were yes, no.

- ICT: what is the highest level of computer use involved in this job? (m_3_08) Possible responses were none, straightforward, moderate, complex, specialized.

A.2 Additional tables and figures

TABLE A1 Mapping of survey questionnaires to task categories.

	STEP surveys		PIAAC surveys	
Task category	Item description	Item nos	Item description	Item nos
Non-routine analytical	Type of documents read and frequency	A-4, A-5-(1–6)	Type of documents read and frequency	G_Q01(a–h)
	Think creatively	B-10	Think creatively	F_Q05b
Non-routine interpersonal	Personal relationship	B-5, B-6	Personal relationship	F_Q02a, F_Q02d, FQ_04a, FQ_04b
	Guiding/coaching	B-13	Guiding/coaching	F_Q02b, F_Q02e, F_Q03b
Routine cognitive	Freedom how to decide work	B-14	Planning activities	FQ_03a
	Presence of short, repetitive tasks	B-16	Organizing own time	FQ_03c
	Learning new things	B-17		
Routine manual	Physical demand	B-3	Long physical work	FQ_06b
Non-routine manual	Driving car, truck, three-wheeler	B-7	Use/accuracy hand/fingers	FQ_06c
	Repair/maintain electronic equipment	B-8		
Use of ICT	Used a computer	B-18	Used a computer	G_Q04

Notes: For STEP countries, we diverge from Caunedo *et al.* (2023) in constructing our measure of routine manual tasks by not including the category of operating heavy machinery, which does not have a clear counterpart in PIAAC. If available, we also employ the question 'As a regular part of this work, do you have to read the following … Other?' (A-5-7) as an additional subcomponent in the construction of the non-routine analytical measure for STEP countries.

TABLE A2 Pooled estimates of establishment size gradient in the task content of jobs.

Outcome variable	(1)	(2)	(3)	(4)	(5)	(6)	(7) $\hat{\delta}$
Panel A: PIAAC							
Non-routine analytical	0.155***	0.159***	0.127***	0.117***	0.114***	0.111***	1.387
	(0.019)	(0.019)	(0.016)	(0.017)	(0.017)	(0.017)	
Non-routine interpersonal	0.074***	0.086***	0.064***	0.055**	0.053**	0.048**	0.796
	(0.022)	(0.023)	(0.022)	(0.022)	(0.022)	(0.022)	
Routine cognitive	0.008	0.006	0.024	0.035**	0.031*	0.039**	−0.581
	(0.015)	(0.015)	(0.015)	(0.014)	(0.016)	(0.015)	
Routine manual	−0.006	−0.000	0.006	0.011	0.019	0.008	−0.172
	(0.015)	(0.014)	(0.011)	(0.010)	(0.012)	(0.011)	
Non-routine manual	−0.007	−0.008	−0.001	0.001	0.006	−0.005	−0.025
	(0.017)	(0.017)	(0.015)	(0.014)	(0.015)	(0.013)	
Use of ICT	0.156***	0.164***	0.140***	0.132***	0.126***	0.121***	1.177
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	
Sample size	65,151	65,151	65,151	65,151	65,151	65,151	
Panel B: STEP							
Non-routine analytical	0.125***	0.129***	0.069**	0.066**	0.067**	0.081**	0.381
	(0.032)	(0.030)	(0.024)	(0.023)	(0.027)	(0.029)	
Non-routine interpersonal	−0.015	−0.003	−0.033	−0.035	−0.062*	−0.039	−0.292
	(0.030)	(0.033)	(0.037)	(0.034)	(0.033)	(0.028)	
Routine cognitive	0.169**	0.162**	0.188**	0.191***	0.226***	0.194***	−2.359
	(0.067)	(0.066)	(0.064)	(0.060)	(0.052)	(0.053)	
Routine manual	−0.020	−0.019	0.007	0.007	0.011	0.022	−0.066
	(0.034)	(0.035)	(0.032)	(0.033)	(0.022)	(0.028)	
Non-routine manual	−0.048	−0.042	−0.060	−0.060	−0.093**	−0.086**	−2.683
	(0.038)	(0.038)	(0.037)	(0.035)	(0.031)	(0.031)	
Use of ICT	0.194***	0.201***	0.126**	0.125**	0.106**	0.102**	0.644
	(0.057)	(0.056)	(0.043)	(0.043)	(0.042)	(0.043)	
Sample size	8339	8339	8339	8339	8339	8339	
Controls							
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	
2-digit occupation fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	
Industry fixed effects		Yes	Yes	Yes	Yes	Yes	
Individual demographics			Yes	Yes	Yes	Yes	
Individual cognition/non-cognition				Yes	Yes	Yes	
Region fixed effects					Yes	Yes	
Country interactions						Yes	

Notes: Extension of Table 1 employing different sets of controls. Regressions of a given category of task requirement intensity on an indicator of large establishment (at least 50 employees). Each row refers to a specific skill category. Additional controls are sequentially included across columns and are indicated in the lower part of the table. Individual demographics include education, gender and age. Regressions are conducted separately for the pooled sample of PIAAC and STEP countries in panels A and B, respectively. Column (7) reports estimates of δ from Oster (2019), with column (2) as the short regression and column (4) as the long regression. Standard errors are reported in parentheses and clustered at the country level.
***, **, * indicate $p < 0.01$, $p < 0.05$, $p < 0.1$, respectively.

TABLE A3 Establishment size gradient in the task content of jobs with more detailed treatment of establishment size.

	NRA (1)	NRI (2)	RC (3)	RM (4)	NRM (5)	COMP (6)
Panel A: PIAAC						
Firm size: 11–50	0.086***	0.061***	0.027	0.065***	−0.013	0.095***
	(0.012)	(0.019)	(0.028)	(0.018)	(0.016)	(0.016)
Firm size: 51–250	0.135***	0.076***	0.045	0.072***	−0.009	0.154***
	(0.016)	(0.020)	(0.032)	(0.020)	(0.014)	(0.013)
Firm size: 250–1000	0.162***	0.060**	0.093***	0.025	−0.007	0.216***
	(0.028)	(0.027)	(0.026)	(0.019)	(0.015)	(0.019)
Firm size: >1000	0.272***	0.178***	0.003	0.007	0.003	0.246***
	(0.034)	(0.051)	(0.032)	(0.032)	(0.046)	(0.029)
Observations	65,151	65,151	65,151	65,151	65,151	65,151
R-squared	0.353	0.288	0.160	0.265	0.093	0.399
Panel B: STEP						
Firm size: 2–5	0.055	0.401***	−0.078	0.145**	0.110	0.036
	(0.072)	(0.047)	(0.107)	(0.050)	(0.078)	(0.065)
Firm size: 6–15	0.087	0.256***	0.115	0.092	−0.012	0.179***
	(0.069)	(0.031)	(0.105)	(0.057)	(0.076)	(0.048)
Firm size: 16–25	0.134*	0.096**	0.160**	0.064	0.027	0.194**
	(0.068)	(0.033)	(0.071)	(0.046)	(0.070)	(0.062)
Firm size: 26–50	0.166**	0.197***	0.299***	0.127**	−0.054	0.273***
	(0.060)	(0.047)	(0.079)	(0.042)	(0.081)	(0.053)
Firm size: 51–200	0.146*	0.198***	0.322***	0.094	−0.013	0.271***
	(0.066)	(0.055)	(0.065)	(0.065)	(0.067)	(0.073)
Firm size: >200	0.220***	0.190***	0.280**	0.119*	−0.070	0.346***
	(0.066)	(0.051)	(0.090)	(0.053)	(0.097)	(0.106)
Observations	7818	7818	7818	7818	7818	7818
R-squared	0.408	0.297	0.175	0.218	0.221	0.492

Notes: Replication of column (2) of Table 1 where the large establishment indicator has been replaced by multiple indicators for whether the establishment's number of employees falls into specific ranges. These ranges are not consistent between PIAAC and STEP, and the omitted category is having up to 10 employees in PIAAC, and being the sole employee in STEP. The sample size for STEP is decreased relative to Table 1 because we have excluded Ukraine from the sample (the elicited firm size categories could not be homogenized with those of the other STEP countries). Standard errors are reported in parentheses and clustered at the country level.
***, **, * indicate $p < 0.01$, $p < 0.05$, $p < 0.1$, respectively.

TABLE A4 Establishment size gradient in the task content of jobs, young workers with short tenure, pooled estimates.

	PIAAC		STEP	
Task category	Short tenure	+ Age < 25	Short tenure	+ Age < 25
NRA	0.115***	0.096***	0.070	0.135*
	(0.015)	(0.020)	(0.054)	(0.064)
NRI	0.041	0.084	−0.109	−0.166
	(0.031)	(0.058)	(0.045)	(0.100)
RC	0.029	−0.030	0.168**	0.221***
	(0.031)	(0.086)	(0.059)	(0.046)
RM	−0.010	0.033	0.017	0.219***
	(0.018)	(0.032)	(0.037)	(0.025)
NRM	−0.003	0.128***	−0.025*	0.120
	(0.017)	(0.045)	(0.011)	(0.069)
Use of ICT	0.121***	0.230***	0.183**	0.145
	(0.013)	(0.030)	(0.076)	(0.088)
Observations	28,220	5716	2656	810

Notes: Pooled PIAAC and STEP samples. Coefficient of an indicator for large establishment in a regression of task content intensity on indicator of large establishment (at least 50 employees) controlling for the set of controls in column (2) of Table 1. Columns (1) and (3) restrict the sample to workers with short tenure. In PIAAC, there is no direct question about tenure, so we proxy short tenure by an individual having worked for multiple employers in the last five years. STEP does provide information on the months that the individual has worked for the firm. We are therefore able to define short tenure in a more demanding manner, as having worked for the current employer for up to 24 months. Columns (2) and (4) additionally require the worker to be up to 25 years old. Standard errors are reported in parentheses and clustered at the country-level.
***, **, * indicate $p < 0.01$, $p < 0.05$, $p < 0.1$, respectively.

TABLE A5 Evidence from the demand side: task requirements.

Task category	LE estimate	Observations
Non-routine analytical	0.181 (0.035)	8338
Use of ICT	0.184 (0.023)	8212

Notes: The table reports the coefficient in a regression of a task measure (non-routine analytical tasks and use of ICT, separately) on an indicator of large employer (LE) and fixed effects for sector, country and occupation asked at random by the surveyors. In parentheses, we report the p-values of the test that the effects are null using wild-bootstrapped standard errors clustered at the country level. The estimating sample is obtained by pooling the information obtained from all the countries participating in the STEP employer surveys. Further details on the construction of the outcome variables are provided in Subsection A.1.

TABLE A6 Pooled estimates of the ESWP, explicitly documenting the returns on tasks.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: PIAAC							
ESWP	0.153***	0.132***	0.129***	0.114***	0.111***	0.097***	0.095***
	(0.014)	(0.014)	(0.014)	(0.013)	(0.012)	(0.017)	(0.014)
Non-routine analytical		0.068***	0.066***	0.048***	0.044***		
		(0.010)	(0.009)	(0.008)	(0.007)		
Non-routine interpersonal		0.022**	0.024**	0.021**	0.020**		
		(0.010)	(0.010)	(0.008)	(0.008)		
Routine cognitive		−0.029***	−0.029***	−0.026***	−0.022***		
		(0.006)	(0.006)	(0.008)	(0.007)		
Routine manual		−0.039***	−0.037***	−0.036***	−0.033***		
		(0.007)	(0.007)	(0.006)	(0.006)		
Non-routine manual		−0.025***	−0.025***	−0.017**	−0.016**		
		(0.007)	(0.007)	(0.007)	(0.007)		
Use of ICT		0.052***	0.052***	0.049***	0.045***		
		(0.006)	(0.006)	(0.004)	(0.004)		
Sample size	54,782	54,782	54,782	54,782	54,782	54,782	54,782
Panel B: STEP							
ESWP	0.229***	0.213***	0.210***	0.185***	0.185***	0.158***	0.139***
	(0.028)	(0.023)	(0.023)	(0.024)	(0.025)	(0.030)	(0.029)
Non-routine analytical		0.039	0.038***	0.016	0.017		
		(0.022)	(0.021)	(0.019)	(0.016)		
Non-routine interpersonal		0.105**	0.107**	0.094**	0.095**		
		(0.040)	(0.040)	(0.034)	(0.035)		
Routine cognitive		−0.090***	−0.089***	−0.084***	−0.085***		
		(0.017)	(0.017)	(0.016)	(0.016)		
Routine manual		−0.057***	−0.057***	−0.062***	−0.062***		
		(0.015)	(0.015)	(0.017)	(0.016)		
Non-routine manual		0.019	0.018	−0.001	−0.001		
		(0.020)	(0.020)	(0.023)	(0.023)		
Use of ICT		0.141***	0.140***	0.123***	0.124***		
		(0.023)	(0.022)	(0.025)	(0.026)		
Sample size	8339	8339	8339	8339	8339	8339	8339
Controls							
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2-digit occupation fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Tasks		Yes	Yes	Yes	Yes	Yes	Yes
Industry fixed effects			Yes	Yes	Yes	Yes	Yes
Individual demographics				Yes	Yes	Yes	Yes
Individual cognition/non-cognition					Yes	Yes	Yes
Region fixed effects						Yes	Yes
Country interactions							Yes

Notes: Extension of Table 3 where we additionally estimate specifications with alternative sets of controls, and also report the results on the returns of tasks on wages. Column (1) does not report estimates for tasks since tasks are not part of that specification. Columns (6) and (7) do not report them because tasks are interacted with country fixed effects, so the level effect of the task lacks a meaningful interpretation. Standard errors are reported in parentheses and clustered at the country level.
***, **, * indicate $p < 0.01$, $p < 0.05$, $p < 0.1$, respectively.

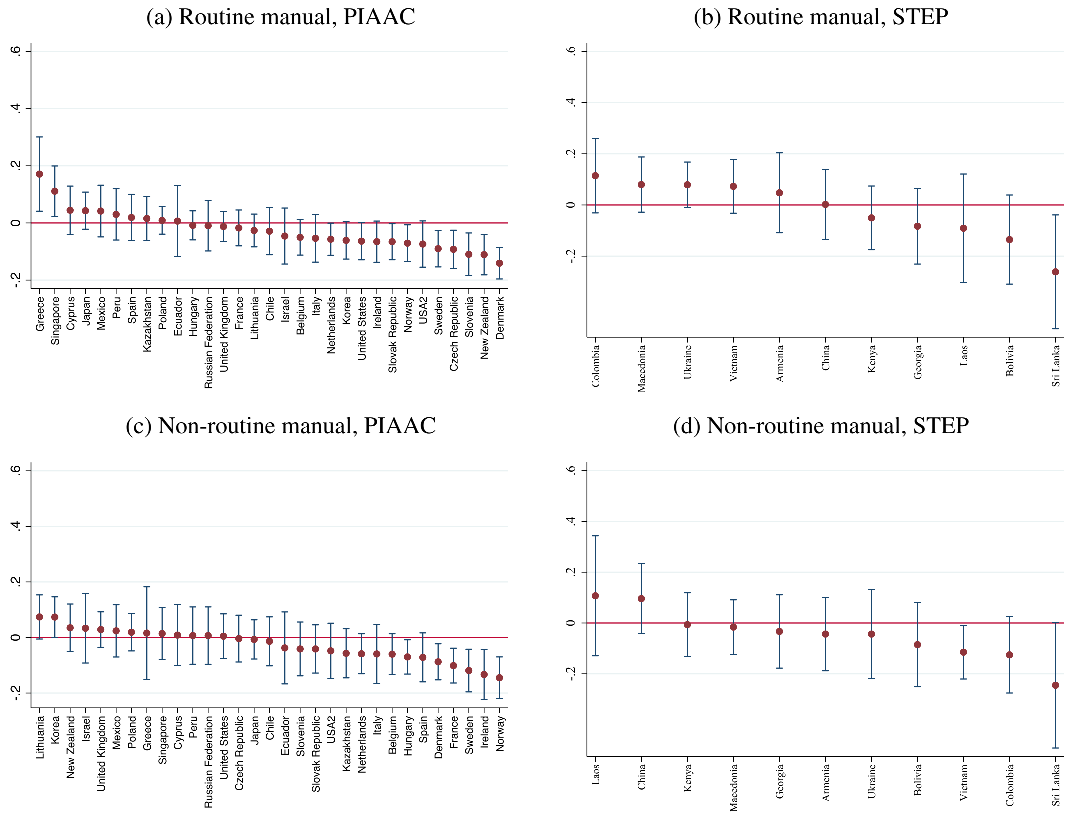

FIGURE A1 Establishment size gradient in the task content of jobs by country, within 2-digit occupations. *Notes*: Continuation of Figure 1. Coefficient of an indicator for large establishment in a regression of task content intensity on indicator of large establishment (at least 50 employees) and the set of controls in column (2) of Table 1. Regressions are done separately for each country. Countries are ordered by decreasing point estimates. 'United States' refers to the survey conducted in 2012, and 'USA2' to the survey conducted in 2017. Reported confidence intervals at 95% confidence level computed using heteroscedasticity-robust standard errors.

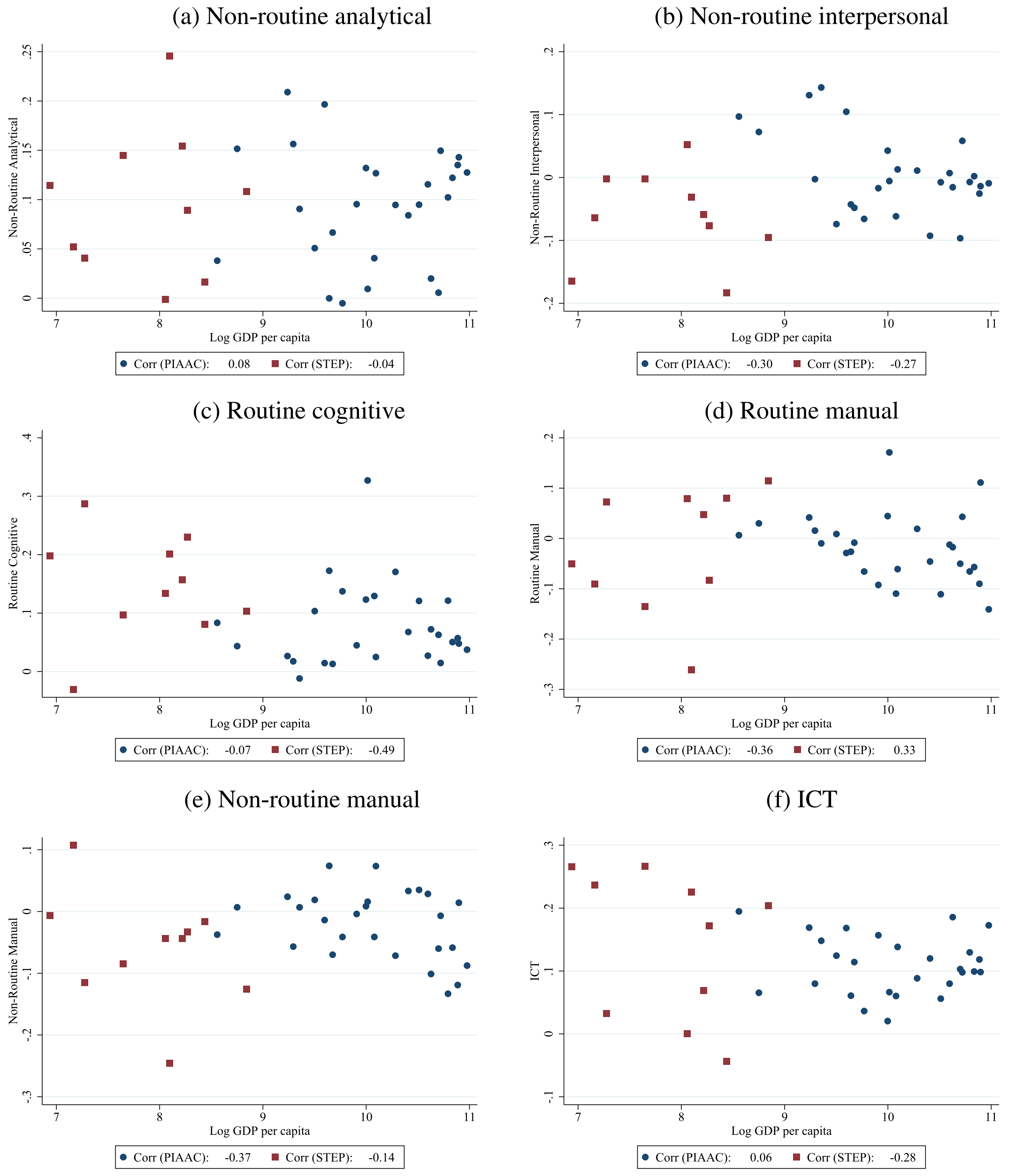

FIGURE A2 Correlations of the establishment size gradient in task content with log GDP per capita. *Notes*: Correlations of the establishment size gradient in task content with country-level log GDP per capita. Correlations weighted by estimated precision of estimated establishment size gradients.

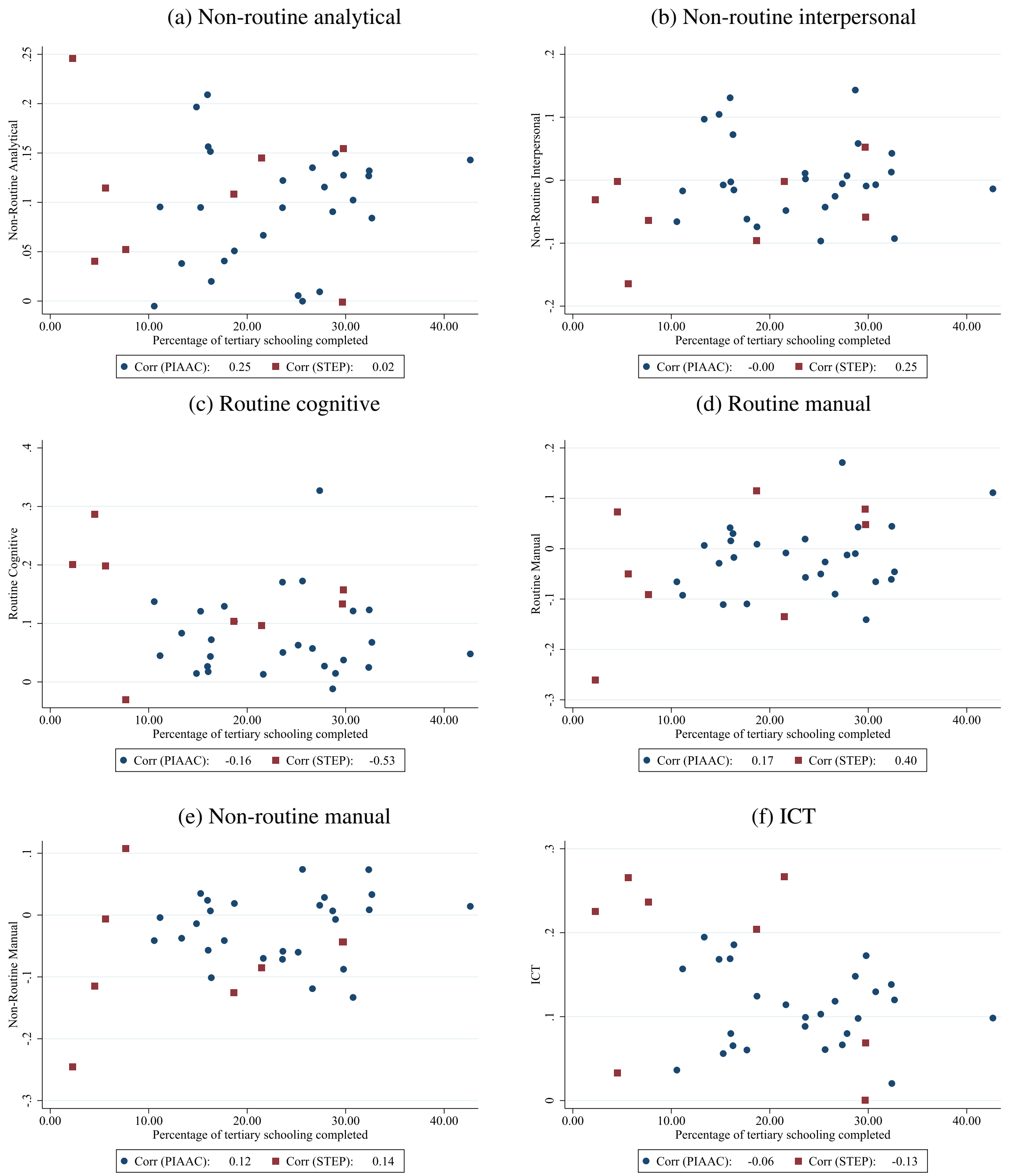

FIGURE A3 Correlations of the establishment size gradient in task content with fraction of population with at least tertiary education. *Notes*: Correlations of the establishment size gradient in task content with country-level proportion of population with at least tertiary-level education (as measured using the Barro and Lee (2013) methodology on 2015 data). Correlations weighted by estimated precision of estimated establishment size gradients.

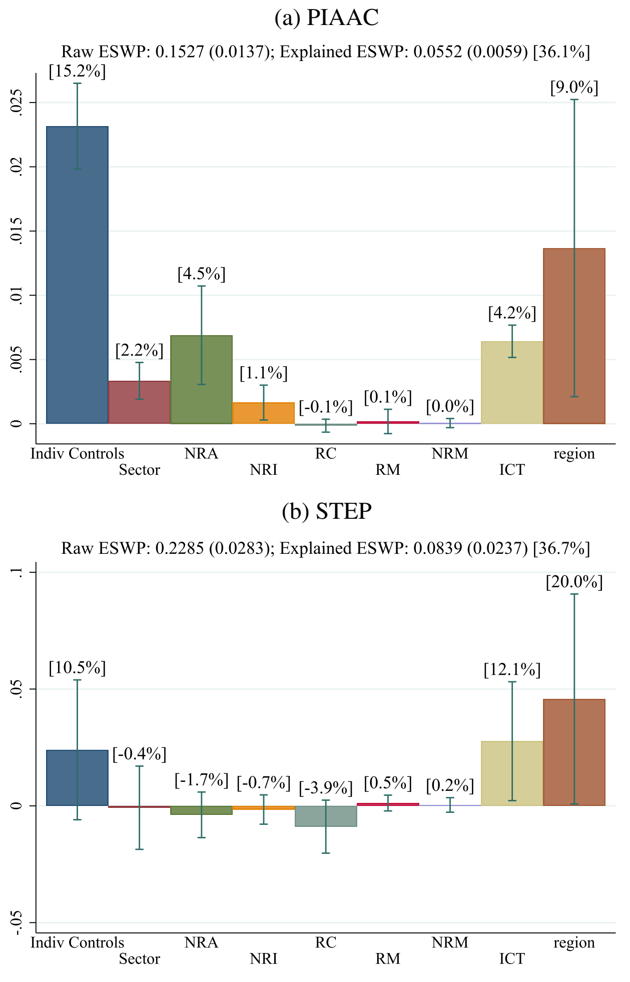

FIGURE A4 Gelbach decomposition of ESWP with region fixed effects as mediators, pooled. *Notes*: Pooled PIAAC and STEP samples. We include region as a potential independent mediator. Numbers in brackets indicate percentages of the raw ESWP. Reported confidence intervals at 95% confidence level. Standard errors are clustered at the country level.