Training, Automation, and Wages: International Worker-Level Evidence

Oliver Falck[a] Yuchen Guo[a] Christina Langer[b] Valentin Lindlacher[c]
Simon Wiederhold[d]

[a]*ifo Institute Munich & LMU Munich* [b]*Stanford Digital Economy Lab* [c]*TU Dresden*
[d]*Halle Institute for Economic Research (IWH) & University of Halle*

Abstract

Job training is widely regarded as crucial for protecting workers from automation, yet there is a lack of empirical evidence to support this belief. Using internationally harmonized data from over 90,000 workers across 37 industrialized countries, we construct an individual-level measure of automation risk based on tasks performed at work. Our analysis reveals substantial within-occupation variation in automation risk, overlooked by existing occupation-level measures. To assess whether job training mitigates automation risk, we exploit within-occupation and within-industry variation. Additionally, we employ entropy balancing to re-weight workers without job training based on a rich set of background characteristics, including tested numeracy skills as a proxy for unobserved ability. We find that job training reduces workers' automation risk by 4.7 percentage points, equivalent to 10 percent of the average automation risk. The training-induced reduction in automation risk accounts for one-fifth of the wage returns to job training. Job training is effective in reducing automation risk and increasing wages across nearly all countries, underscoring the external validity of our findings. Women tend to benefit more from training than men, with the advantage becoming particularly pronounced at older ages.

Keywords: Job Training, Human Capital, Automation, Technological Change, Entropy Balancing

JEL: J24, J31, J61, O33

[⋆]This research has greatly benefited from discussions with David Deming, Florian Englmaier, Carl Benedikt Frey, Thomas Lemieux, Raffaella Sadun, Ludger Woessmann, and with conference participants at Verein für Socialpolitik, EALE, SOLE, ifo Venice Summer Institute, and CESifo Junior Workshop or the Economics of Education. Margarita Gatsou provided excellent research assistance. We gratefully acknowledge funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 101004703. Guo gratefully acknowledges financial support from Deutsche Telekom AG and the generous hospitality at the Harvard Growth Lab. Langer gratefully acknowledges financial support from the Dieter Schwarz Foundation. The views expressed in this paper are those of the authors and do not necessarily represent the views of the funding organizations.

1. Introduction

The fear of technology displacing workers has been an ongoing worry since the dawn of the industrial revolution. Historical accounts document the protests and destruction of machinery in England and Germany as workers resisted the mechanization of manufacturing (Goldin and Katz, 1998; Mokyr et al., 2015). In recent decades, this anxiety has resurfaced due to significant advances in information and communication technology (ICT) and industrial robotics becoming a prominent issue in public discourse and labor disputes. Unlike in the past, economists today have the data and analytical tools to rigorously assess the extent to which technology substitutes human labor. While several studies have documented the negative impacts of technological advancements on employment and wages (e.g., Acemoglu and Autor, 2011; Goos et al., 2014; Arntz et al., 2016; Cortes, 2016), it is rare for entire occupations to disappear (e.g., Atalay et al., 2020; Bachmann et al., 2022). One reason for this is that technology typically automates only specific tasks within an occupation, leaving room for workers to adapt their roles. This raises a critical question: Can job training enable workers to update their task portfolio, thereby reducing their susceptibility to being replaced by technology?

To address this question, we use detailed individual-level data on job tasks and job training from the Programme for the International Assessment of Adult Competencies (PIAAC). We construct a novel measure of automation risk that varies at the individual worker level and we leverage rich data on job training and workers' background characteristics to address selection into training. Applying entropy balancing and using variation within occupations, within industries, and within countries, we find that job training substantially reduces the average risk of automation.

PIAAC provides comprehensive and internationally comparable information on job tasks across 37 countries for more than 90,000 workers. It surveys a broad spectrum of tasks performed at work, including manual, cognitive, digital, and social domains. Crucially, PIAAC allows us to identify tasks that are particularly difficult to automate, such as those requiring (a) social intelligence for navigating complex social interactions, (b) cognitive intelligence for complex reasoning, and (c) perception and manipulation for executing physical tasks in unstructured environments. By leveraging this rich task-level data, we construct an individual-level measure of automation risk. Intuitively, this measure is a weighted share of job tasks, each with different degrees of susceptibility to automation coming from Frey and Osborne (2017). We document substantial variation in automation risk within occupations.[1] In fact, our analysis reveals a notable overlap in individual-level automation risk even between occupations at opposite ends of the occupation-level automation risk spectrum.[2] Moreover, we demonstrate that *every* occupation exhibits substantial within-occupation variation in automation risk, emphasizing the importance of going beyond existing occupation-level automation measures (e.g., Pajarinen and Rouvinen, 2014; Brzeski and Burk, 2015; Frey and Osborne, 2017; Nedelkoska and Quintini, 2018).[3]

Since workers in PIAAC report their training activities in the year preceding the survey, we can exploit within-occupation variation to investigate the role of job training in reducing workers' vulnerability to automation. Despite the data being cross-sectional, the rich background information provided by PIAAC—including tested numeracy skills as a proxy for general ability—helps mitigate biases due to selection into training. We also control for other factors typically associated with participation in training, such as educational attainment, socio-demographic characteristics, full-time employment status, industry, and employer size. In our preferred specification, which incorporates country, industry, and occupation fixed effects and applies entropy balancing, we find that participation in job training reduces a worker's automation risk by 4.7 percentage points (pp). This is equivalent to one-fourth of a standard deviation in automation risk or ten percent of the mean risk. The magnitude of the training effect is comparable to the difference in automation risk between ICT professionals (0.40) and ICT technicians (0.44) or between business administration professionals (0.39) and business administration associates (0.43), with professional occupations in the same field typically involving more complex tasks and therefore lower automation risk.

Notably, the estimate on training effectiveness changes only little across specifications once we include occupation fixed effects. This robustness suggests that, once we solely rely on within-occupation variation, selection biases into training are unlikely to be a major identification issue. This conclusion is further supported by a bounding analysis

[1] We also observe significant variation in automation risk between occupations, consistent with the existing literature (e.g., Autor et al., 2003; Goos et al., 2014).

[2] Arntz et al. (2017) use U.S.-specific data from PIAAC to illustrate that estimates of automation risk are strongly inflated when task variation within occupations is ignored.

[3] Our main analysis focuses on occupations at the two-digit level, as these are consistently available across all countries in PIAAC. However, in the more restricted sample of countries that report occupations at the four-digit level, we observe similar patterns of within-occupation variation in automation risk. In particular, the average standard deviation in automation risk across countries and occupations at the two-digit occupation level is 0.167, while it is 0.169 at the four-digit occupation level.

(Oster, 2019). Overall, our results indicate that training enables workers to expand their task portfolios and increasingly engage in tasks with a lower risk of automation.

We also investigate whether the reduction in automation risk through job training translates into actual wage gains. Our analysis shows that participation in job training leads to a marked increase in hourly wages, with our preferred specification indicating an 8.2 percent wage increase. Notably, similar to our findings for automation risk, the greatest reduction in estimated training effectiveness occurs when accounting for occupation fixed effects. The wage increase from job training is comparable to the wage gradient associated with an additional year of schooling in industrialized countries (Hanushek et al., 2015). Further analysis reveals that approximately one-fifth of the wage gains from training can be attributed to its role in reducing workers' automation risk, underscoring how the shift toward less automatable, technology-complementary tasks is rewarded in the labor market.

Crucially, our individual-level automation risk measures allow us to advance the existing literature by exploiting variation in automation risk within occupations. Ideally, however, we would like to use individual-level panel data to track changes in automation risk over time and control for selection into training by accounting for pre-training automation risk in our estimations. While this approach is not feasible with the cross-sectional international PIAAC data, we can approximate or even observe workers' pre-training automation risk in country-specific extensions. These analyses yield results that are consistent with our baseline estimations.

First, we leverage a repeated cross-section of the PIAAC survey available for the United States in 2012 and 2017. Using the no-training group from 2012, we impute individual-level past automation risk for workers in the 2017 sample. While past automation risk strongly predicts contemporaneous automation risk, including it as a control has minimal impact on the training estimate. Moreover, drawing on the approach outlined by Kleven et al. (2019) and Kleven et al. (2024), we construct a pseudo panel by matching observations in the 2017 survey to observations in the 2012 survey based on an extensive set of observable characteristics. This matched pseudo panel allows us to add fixed effects to our preferred specification to account for unobserved characteristics of matched pairs.

Second, we take this analysis one step further for Germany, which, as a unique feature among participating countries, created an individual-level panel by re-surveying participants from the original PIAAC sample 3.5 years later.[4] Exploiting this panel dimension,

[4] The United States and Germany are the only PIAAC countries with multiple waves of data collection.

we directly control for past automation risk in a value-added specification. Corroborating the U.S. pseudo-panel evidence, past automation risk is a strong predictor of contemporaneous automation risk, whereas adding it as a control yields only a slight reduction in the estimated training coefficient. The same applies when we also control for past training participation. This suggests that our baseline model already effectively accounts for selection into training based on past automation risk and general training propensity. Observing changes in automation risk and training over the 3.5-year period also enables us to estimate individual fixed effects regressions, which account for time-invariant unobserved heterogeneity. Although the variation available for the panel analysis is limited due to the relatively short time span, we still find a consistent negative relationship between job training and automation risk.

To understand how job training reduces automation risk, we investigate the specific tasks performed by workers after receiving training and the types of skills they develop. Our results indicate that job training increases the use of tasks that are difficult to automate (i.e., negotiation, complex problem-solving, teaching, advising others, influence others, and plan others' work) to a very similar extent. Interestingly, training also increases the use of tasks more susceptible to automation (e.g., manual dexterity or solving simple problems), but to a much smaller degree. Additionally, we find that job training increases workers' digital skills, which are also assessed in PIAAC. This result indicates that job training often involves learning to work with new technology.[5] These shifts in task composition and improvements in digital skills suggest that training helps workers adapt to technological change by equipping them with skills that complement, as opposed to compete with, automation technologies.

Finally, our individual-level analysis allows us to explore for whom training generates the largest impact on workers' task portfolios and wages. Examining heterogeneity in training effectiveness is crucial because it provides insights into how different demographic groups and countries may benefit differently from job training. Understanding these differences can guide the development of more targeted and effective training policies. We find that training tends to be more effective at reducing automation risk and increasing wages for women compared to men, with gender differences being particularly pronounced at older ages. This greater effectiveness for women may be attributed to differences in the tasks they perform (Black and Spitz-Oener, 2010), or their higher representation in

[5] The content of job training is not explicitly reported in PIAAC. However, prior work confirms that learning to work with new technology is often a main focus of job training measures, particularly in environments undergoing rapid technological change (Ma et al., 2024).

industries where training yields higher returns. Our large country sample also enables us to investigate cross-country heterogeneity. We find that training is significantly related to outcomes in almost all countries, whereas the magnitude of the training estimate varies considerably across countries. These patterns of heterogeneity are remarkably consistent across both automation risk and wage outcomes.

We also investigate the effectiveness of job training by training characteristics. We find that longer training activities are more effective in reducing automation risk and increasing wages, suggesting that the benefits of job training arise from genuine skill acquisition rather than mere signaling effects. Additionally, training is most effective when employers fully or partially finance it, which is the case for nearly 70 percent of workers in our sample.[6] This suggests that training-induced productivity gains are more likely to materialize when employers have a vested interest in the success of the training.

Our paper contributes to three key strands of literature. First, we contribute to the literature on the labor market effects of job training, which has primarily focused on wages, productivity, and employment prospects (LaLonde, 1986; Blundell et al., 1999; Lechner, 1999; Goux and Maurin, 2000; Pischke, 2001; Dearden et al., 2006; Goerlitz, 2011; Hidalgo et al., 2014; Goerlitz and Tamm, 2016; Adhvaryu et al., 2023).[7] These studies typically use observational data with varying identification strategies, with a few studies even exploiting random assignment into training (Schwerdt et al., 2012; De Grip and Sauermann, 2012; Adhvaryu et al., 2023). This literature consistently finds that job training improves labor market outcomes.[8] Our study is the first to show that a considerable portion of the wage gains from job training can be attributed to its role in reducing workers' risk of automation.

Additionally, our work is related to studies on the effectiveness of active labor market policies and training programs for unemployed workers (Hujer et al., 2006; Card et al., 2010; Kluve, 2010; McCall et al., 2016).[9] Schmidpeter and Winter-Ebmer (2021) find that training programs for unemployed workers are especially effective in improving the chances of finding a job for individuals previously employed in routine-intensive occupa-

[6] Firm-level studies demonstrate that productivity gains from training often exceed the associated wage increases, which could be a strong incentive for firms to pay for training (Konings and Vanormelingen, 2015; Ma et al., 2024).

[7] For comprehensive overviews of the literature, see Leuven (2005), Bassanini et al. (2007), and De Grip and Sauermann (2013).

[8] A substantial body of literature also explores whether adult education enhances non-pecuniary outcomes, including well-being, health, and civic, political, and cultural engagement. For an overview of this literature and supporting empirical evidence, see Ruhose et al. (2019).

[9] See Card et al. (2018) for an overview.

tions. However, Nedelkoska and Quintini (2018) and Heß et al. (2023) demonstrate that workers most exposed to automation are, in fact, the least likely to participate in training.

Second, our study contributes to the literature by investigating the role of job training in reducing workers' automation risk within an international context. We reveal a crucial mechanism through which training can influence wage and employment outcomes—especially in light of rapid technological advancement. By utilizing harmonized data from 37 countries, we demonstrate that the effectiveness of training in mitigating automation risk is present across diverse technological landscapes and levels of automation risk across countries.[10] This international perspective strongly enhances the external validity of our findings, addressing a gap in the literature that has typically relied on data from single countries or specific experiments.[11] At the same time, we uphold the rigorous standards for addressing bias from selection into job training. We do so by exploiting within-occupation variation and ensuring that workers with and without training are similar in tested numeracy skills (capturing unobserved ability) and in a large set of additional control variables, using entropy balancing techniques.

Third, our study complements the growing body of research on the labor market effects of technology, which has evolved from a skill-based to a task-based approach (**?**Krusell et al., 2000; Autor et al., 2003; Acemoglu and Autor, 2011; Akerman et al., 2015). This research documents that technological advances have favored higher-skilled workers and those in so-called non-routine jobs; largely due to their higher adaptability to new technologies and their lower susceptibility to technological change (e.g., Deming and Noray, 2020). Studies highlight impacts like job loss (Braxton and Taska, 2023), or the role of specific technologies such as mechanized telephone operations (Feigenbaum and Gross, 2024) or the introduction of robots across industries (Hirvonen et al., 2022; Adachi et al., 2024). Our study extends this research by examining how job training helps workers update their task portfolios, reducing their vulnerability to automation and enhancing their

[10] The average automation risk in our data ranges from 38 percent in Norway to 58 percent in the Slovak Republic, with a cross-country mean of 46 percent.

[11] Salas-Velasco (2009) and Fialho et al. (2019) utilize international datasets—Careers after Higher Education: a European Research Survey (CHEERS) and PIAAC, respectively—to explore the relationship between job training and labor market outcomes. While the former focuses on the impact of on-the-job training for European university graduates on wages, the latter takes a broader, policy-oriented perspective, providing descriptive analyses of training frequency, trends across countries over time, and associated wage effects. Unlike our study, neither addresses the role of automation in the context of training. Brunello et al. (2023), using data from the European Investment Bank Investment Survey (EIBIS), examine the impact of automation on employer-provided training at the firm level, finding that firms appear less inclined to offer training to employees after adopting advanced digital technologies. A detailed overview of country-specific studies on the effects of training is available in Ma et al. (2024).

ability to adapt to technological change. Additionally, we show that job training improves digital skills, enabling workers to engage with complex technologies more effectively.

The rest of the paper is organized as follows: Section 2 presents our international worker data, while Section 3 describes how we use these data to construct individual-level measures of automation risk. Section 4 sets out our empirical strategy. Section 5 presents our results on the effects of training on workers' automation risk and wages, discusses the robustness of our estimates, and provides evidence for mechanisms. Section 6 provides evidence for heterogeneities in training effectiveness across countries, socio-demographic groups, and the type of training. Section 7 concludes.

2. PIAAC Data

2.1. International Data

Our empirical analysis is based on data from the Programme for International Assessment of Adult Competencies (PIAAC), a large-scale survey administered by the Organisation for Economic Co-operation and Development (OECD, 2013, 2016). The survey includes representative samples of working-age individuals (16 to 65 years) from 39 countries, collected in three rounds between 2011 and 2017.[12] The survey items are harmonized across diverse economic and cultural contexts, enabling both international analyses, when pooling all countries, and reliable cross-country comparisons. The international harmonization is especially valuable for measuring job training, workplace tasks, and adult skills, as no other dataset provides such measures in a comparable fashion across a wide range of countries.

We utilize detailed information on participation in job training, our treatment variable, from PIAAC's background questionnaire, where workers report on their training activities in the twelve months prior to the survey.[13] We define a training measure as *job training* if it involved either organized on-the-job training sessions (e.g., training by supervisors or

[12] We exclude data for the Russian Federation in our analysis. According to the OECD (2013), data for the Russian Federation are preliminary and are not representative of the entire Russian population because they do not include the population of the Moscow municipal area. Moreover, we treat England and Northern Ireland as one country: the United Kingdom. Thus, our sample has a total of 37 industrialized countries.

[13] Respondents in PIAAC dedicated significant time to the assessment. They spent approximately 25 to 40 minutes completing a detailed background questionnaire, followed by an additional 50 to 60 minutes on the cognitive skill assessment. Data collection was conducted in respondents' homes under the supervision of trained interviewers, utilizing a computer-based application (with a paper version available as an alternative). Each individual assessment began with the background questionnaire and then proceeded to the skill test (OECD, 2019).

co-workers) or, if not classified as on-the-job training, was reported to be job-related.[14] On average, 54.6 percent of workers participated in job training in the year prior to the survey having been conducted (see Table B.1).[15] Training participation decreases with age but increases considerably with education and firm size. Women participate slightly more than men, but men are more likely to have employer-financed training. The median duration of job training is four days, decreasing with age and increasing with education and firm size. More than two-thirds of training activities (68.9 percent) are fully or partly employer-financed, with larger firms providing more financial support.[16]

Job training is distinct from broader categories of adult learning and education, which may include more general education programs, lifelong learning initiatives, or informal skill acquisition. We specifically refer to interventions directly linked to improving performance in one's current job or firm. However, PIAAC also provides information on participation in broader educational measures that are not directly applicable to the workplace (i.e., open/distance education, seminars/workshops, or private lessons). We include non-job-related training as a control variable to capture participants' general motivation to engage in training activities.

Our primary outcome of interest is task use at the workplace, which is captured in PIAAC's background questionnaire. Respondents are asked about the extent to which they engage in various job tasks, such as using accuracy with their hands or fingers (perception manipulation), solving complex problems (creative intelligence), or negotiating with others (social intelligence).[17] Based on these detailed items on task use, we construct a measure of automation risk at the individual level (see Section 3.1 for details).

[14] The exact questions asked were: (i) "During the last twelve months, have you attended any organized sessions for on-the-job training or training by supervisors or co-workers?" (ii) "Was this activity [respondents could choose between different types of training] mainly job-related?"

[15] Training participation varies considerably across countries (between 20.7 percent in Kazakhstan and 27.3 percent in Greece as countries with low training frequencies and 71.8 percent in Finland; cross-country mean: 52.2 percent).

[16] The international PIAAC data do not provide any information on the specific content of training programs. However, in a follow-up survey of the German PIAAC study (see Section 2.2), respondents were asked about the types of training they received. Notably, "Computer or software use, information technology (IT)" emerged as the most significant area of focus, with twelve percent of respondents reporting this as their main field of training. Other notable fields of training include "Security, e.g., first aid, occupational safety" (ten percent), "Health care" (eight percent), as well as "Personal development, communication skills", "Business knowledge", and "Project management, leadership skills" (six percent each). Thus, while the content of job training is very diverse, IT-related training seems to be particularly important in the current labor market.

[17] Responses are provided on a five-point Likert scale: 1 - Never, 2 - Less than once a month, 3 - Less than once a week but at least once a month, 4 - At least once a week but not every day, 5 - Every day.

In addition to examining the effect of training on automation risk, we also consider its impact on hourly wages as a direct measure of labor market success.[18]

PIAAC also assesses respondents' key workplace skills in literacy, numeracy, and problem-solving in technology-rich environments (referred to as digital skills).[19] PIAAC defines these skill measures as follows (OECD, 2013):

> *Numeracy*: ability to access, use, interpret, and communicate mathematical information and ideas in order to engage in and manage the mathematical demands arising from a range of situations in adult life;
>
> *Literacy*: ability to understand, evaluate, use, and engage with written texts to participate in society, to achieve one's goals, and to develop one's knowledge and potential;
>
> *Digital*: ability to use digital technology, communication tools and networks to acquire and evaluate information, communicate with others and perform practical tasks.

As a potential mechanism, we investigate the effect of job training on digital skills, given that job training often focuses on equipping workers with the necessary skills to operate new hardware or software (e.g., OECD, 2017).[20] In contrast, we treat numeracy skills as a control variable, which accounts for selection into training based on ability.[21] In addition to numeracy skills, we utilize the extensive set of observable worker characteristics provided by PIAAC for entropy balancing, which helps mitigate estimation bias arising from selection into job training (see Section 4).

[18]PIAAC's Public Use File reports hourly wages for Austria, Canada, Germany, Hungary, Sweden, Turkey, and the United States only as a worker's decile rank in the country-specific wage distribution. For Germany, we obtained the Scientific Use File, which provides continuous wage information. For the remaining countries, we follow Hanushek et al. (2015) by assigning the decile median of hourly wages to each survey participant belonging to the respective decile of the country-specific wage distribution. Using wages in coarse categories in some countries is unlikely to affect our results, as Hanushek et al. (2015) demonstrate that using decile medians instead of continuous wages has no substantive impact on their returns-to-skills estimates. Additionally, we trim the bottom and top one percent of the wage distribution in each country to limit the influence of outliers.

[19]These skill measures have been shown to yield considerable earnings returns in all participating countries (e.g., Hanushek et al., 2015, 2017; Falck et al., 2021).

[20]The assessment of digital skills was an optional component of the survey, and Cyprus, France, Italy, and Spain did not participate in this assessment. Additionally, digital skills were only tested for participants who successfully completed an initial computer core test, which measured basic digital competencies such as using a keyboard and mouse or scrolling through text on the screen. For further details, see (Falck et al., 2021).

[21]PIAAC measures each of these skill domains on a 500-point scale. For exposition, we standardize scores in the regression analyses to have a mean of zero and a standard deviation of one in the estimation sample. Additionally, following Hanushek et al. (2015), we use the first plausible value of the PIAAC scores in each domain throughout.

To ensure that our results are comparable across outcomes, we limit the sample to workers who provide information on both workplace tasks and wages, allowing our automation and wage regressions to be based on the same sample of workers. The final sample encompasses 91,470 workers.[22]

2.2. Country-Specific Extensions

In addition to the international PIAAC dataset, we leverage two country-specific extensions—those from the United States and Germany—to control even more rigorously for selection into job training than in the broader cross-country analysis.[23] Both countries provide data over time, which allows us to investigate changes in the automation risk and exploit (quasi) individual-level variation within individuals.

United States. The United States is the only country that participated in two waves of the PIAAC survey, conducted in 2011/2012 and 2017.[24] Both survey rounds were administered under identical conditions, ensuring consistent data collection. Moreover, the background questionnaire included the same questions in both years, allowing for direct comparisons over time. The U.S. PIAAC data are repeated cross sections: Samples were drawn to be representative of the U.S. population aged 16 to 65 in the respective survey year, with different individuals being sampled in each wave.

Germany. In Germany, PIAAC offers a unique longitudinal dimension by re-surveying the sample of German adults who participated in the initial PIAAC assessment in 2011/2012 (referred to as PIAAC-L). This follow-up survey took place approximately 3.5 years later, in 2015, enabling us to track changes in training participation and job tasks over time at the individual level.[25]

From the original 5,379 participants in the 2012 PIAAC wave, a re-taker sample of 3,263 individuals (60.7 percent) was re-tested in 2015 (Rammstedt et al., 2017; Zabal et al., 2017). Survey re-takers are slightly positively selected in terms of achievement; a pattern also observed in other longitudinal assessment surveys (Martin et al., 2021).

[22] We also provide results for an unrestricted sample that includes observations without wages, yielding very similar findings.

[23] For consistency with the other PIAAC countries, our international results use only the 2012 wave for the United States and Germany.

[24] For exposition, we consistently refer to the earlier PIAAC wave as the 2012 wave.

[25] As part of PIAAC-L, Germany also conducted additional follow-up surveys of the original PIAAC participants in 2014 and 2016. However, these surveys did not include detailed questions on job tasks, making it impossible to construct individual-level measures of automation risk from these data.

However, PIAAC provides sampling weights to adjust the re-taker sample to the general population, ensuring that results are representative of the entire German population.

3. Automation Risk

In this section, we describe our approach to constructing an individual-level measure of automation risk, building on, and extending, the methodology developed by Nedelkoska and Quintini (2018). While their research pioneered the use of PIAAC data to estimate automation risk at the occupational level across different countries, we refine this approach to capture automation risk at the individual level. This advancement allows us to study how job training impacts a worker's specific risk of automation, rather than relying solely on aggregated occupational categories. We show that our novel measure of automation risk at the individual level provides new insights regarding workers' susceptibility to automation and their need to adapt to evolving technological change.

3.1. Measuring Automation Risk at the Individual Level

Our methodology generally follows the framework established by Nedelkoska and Quintini (2018), who estimated the probability of tasks becoming fully automated within a given occupation. However, instead of aggregating this information to generate occupation-level risk measures, we focus on the individual tasks that each worker performs. This enables us to derive a personalized automation risk score based on the specific tasks reported by each individual.

The construction of our individual-level automation risk measure involves two key steps. First, Nedelkoska and Quintini (2018) utilized expert assessments from Frey and Osborne (2017) to identify occupations at risk of automation, considering factors such as the difficulty of automating tasks that require social intelligence, complex problem-solving, and manual dexterity.[26] They then used a logistic regression model to predict automation risk, with the probability of full automation as the dependent variable and task usage data from PIAAC as the independent variables.[27] This relationship between

[26] The occupational automation risk in Frey and Osborne (2017) is based on the identification of primary engineering bottlenecks encountered by mobile robotics and machine learning developers. To achieve this, Frey and Osborne (2017) conducted interviews with engineering scientists during a 2013 workshop at Oxford University's Engineering Sciences Department. The scientists were asked to assess whether tasks associated with 70 different occupations could be automated using advanced computer-controlled equipment. Occupations where all tasks were deemed automatable received a value of one, whereas occupations where only some or no tasks could be automated received a value of zero.

[27] See Table B.2 for the exact wording of the task questions.

tasks and automation risk is estimated using Canadian PIAAC data. The Canadian data provide more detailed occupational information and a greater sample size than other PIAAC countries; this enables direct mapping to the occupational classifications used by Frey and Osborne (2017). The coefficients obtained from this regression are then applied to calculate the automation risk of jobs at a more aggregated occupational classification level, both within Canada and for other countries. The coefficients from this, shown in Table B.3, reflect each task's contribution to the overall automation risk. For instance, tasks involving influencing others, solving complex problems, or negotiating, tend to reduce an individual's automation risk. Conversely, tasks such as solving simple problems or using manual dexterity increase the risk.

To construct our individual-level automation risk measure, we apply the coefficients from Table B.3 as weights to the task usage values reported by workers in PIAAC. We then sum these weighted values across all relevant job tasks and input the sum into a logistic function to predict each worker's individual automation risk. This results in a score ranging from zero (low probability of full automation) to one (high probability of full automation).[28]

Automation Risk in 2015 for Germany. We utilize the longitudinal dimension of the German PIAAC survey (see Section 2.2) to develop an automation risk measure that reflects the period of approximately 3.5 years following the initial assessment and investigate changes in the automation risk at the individual level. Specifically, we began by identifying job tasks related to routine or automatable work in the 2015 survey data. The relevant tasks incorporate measures such as solving difficult problems, carrying out short, repetitive tasks, and organizing one's own work (see Table B.4 for a full list of variables). Since the questions used to measure automation risk in 2015 differ from those in the 2012 survey—upon which our primary automation risk measure is based—we ensured comparability between the 2012 and 2015 measures by applying the following method. First, we estimated a logit regression of the 2012 automation risk, aggregated at the four-digit ISCO level, on the individual-level job task measures from the 2015 survey. Data are linked based on the occupation of the 2015 respondents. The coefficients from this regression provide reasonable correlations with the 2012 occupation-level automation risk (Table B.5). For instance, solving difficult problems, dealing with unexpected situations, and checking the work of others has a strong negative correlation with automation

[28] The logistic function ensures that our predicted individual-level automation risk lies between zero and one.

risk, while obtaining detailed specifications for tasks, carrying out short, repetitive tasks, and having similar work days are positively correlated with automation risk. We then apply these coefficients to all individuals in the 2015 survey to generate our automation risk variable, similar to the approach outlined above. The resulting automation risk measure for 2015 exhibits a strong correlation with the 2012 measure ($\rho = 0.48$), indicating consistency and comparability between the two periods.

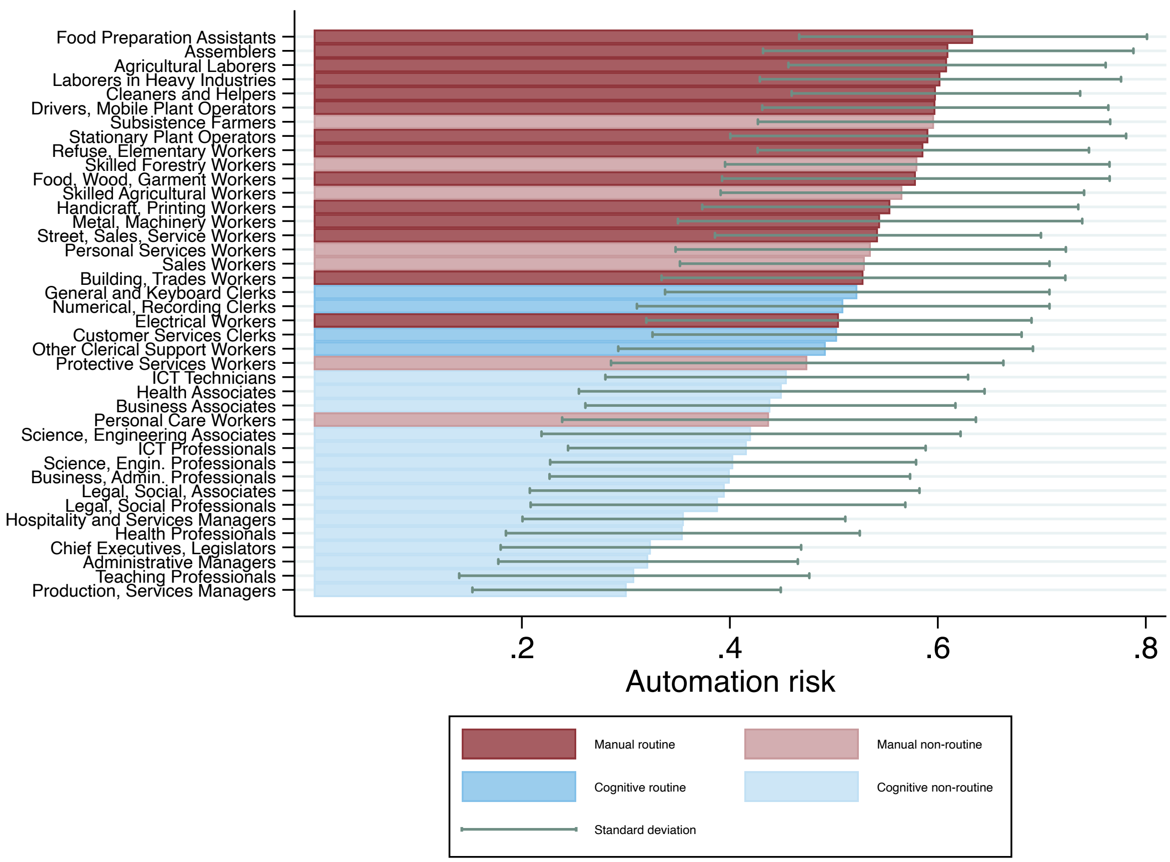
3.2. Descriptive Patterns in Automation Risk

Figure 1 illustrates automation risk across and within occupations. This automation risk reflects an individual worker's probability of automation based on his/her job tasks. A value of zero suggests that the worker's tasks make the job entirely secure from automation, while a value of one means that the worker is fully at risk of being automated. We observe a familiar pattern across occupations: Cognitive non-routine occupations exhibit the lowest average automation risk; for example, less than one-third of the tasks performed by production and service managers or teaching professionals are fully automatable. In contrast, manual routine occupations face the highest risk of automation, with food preparation assistants at the extreme end—where the likelihood of their tasks being fully automated is as high as 61 percent. Across all occupations and countries in our sample, the average automation risk lies at 46 percent.

More notably, however, there is considerable variation in automation risk *within* occupations, indicating that workers in the same occupation perform tasks with varying degrees of automatability. For instance, food preparation assistants at the fifth percentile of automation risk are as automatable as the average production and service manager. In other words, five percent of workers in the most automatable occupation face an automation risk as low as the average risk in the least automatable occupation. Similarly, when we consider production and service managers with a higher-than-average automation risk—say, those at the 75th percentile—eleven percent of food preparation assistants have a comparable automation risk. These examples illustrate a substantial overlap in the automation risk faced by individual workers, even when comparing occupations with the lowest and highest average risks.

Another way to show the amount of variation in within-occupation automation risk in Figure 1 is by examining the standard deviations. We find that the standard deviation of automation risk in the occupation with the highest average automation risk, food preparation assistants, is 17 pp, while it is 14 pp in the occupation with the lowest average automation risk, production and service managers. Generally, substantial within-occupation variation in automation risk is evident across *all* occupations, indicating that this variation is not dependent on an occupation's average automation risk. In Figure A.1, we plot the densities of our individual-level automation risk across occupations and reveal that there is a substantial mass of workers in each occupation with an automation risk lying outside the mode of their respective occupations. These comparisons underscore that focusing solely on occupational averages would obscure significant heterogeneity and potentially lead to misleading conclusions about workers' actual automation risk based on the tasks they perform.[29]

Figure 1: Automation Risk Across and Within Occupations



Notes: The figure shows the average automation risk by two-digit ISCO occupation across all countries in our sample. Whiskers indicate one standard deviation from the mean (see Section 3.1 for details on the construction of the automation risk measure). Classification of occupations by routine intensity according to Cortes (2016).
Data source: PIAAC.

[29]Our individual-level automation risk measure is a strong predictor of workers' wages, underscoring its economic relevance. Figure A.2 shows that workers in jobs less susceptible to automation tend to earn higher wages, even after controlling for country, industry, and occupation fixed effects. Thus, even within occupations, there is a negative relationship between automation risk and wages.

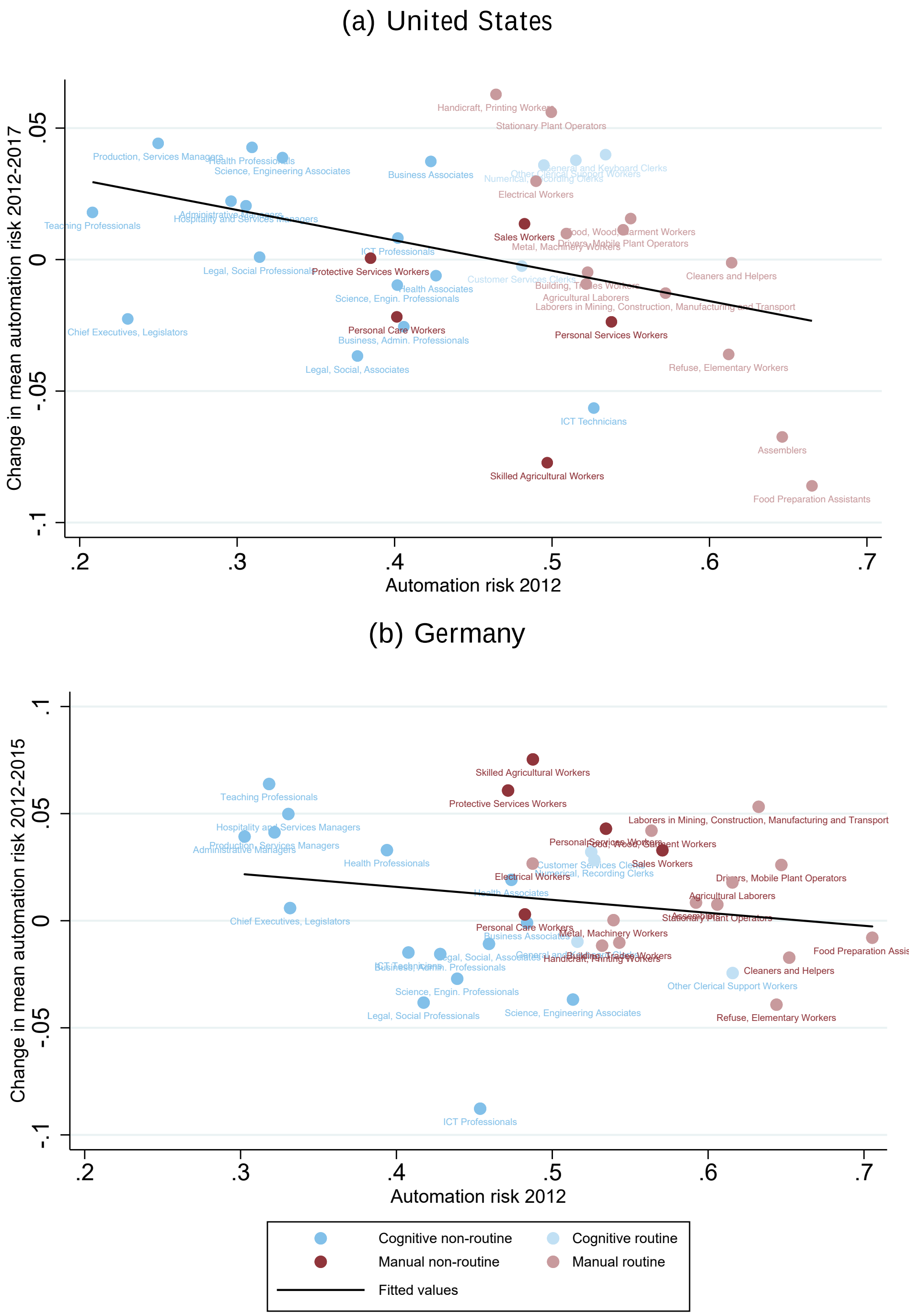
We also document notable shifts in the composition of tasks over time. For this analysis, we take advantage of the fact that both the United States and Germany were surveyed twice in PIAAC (see Section 2.2). Across both countries, Figure 2 shows that occupations with a higher average automation risk in 2012 were subject to a more substantial decrease in automation risk over time. Since the same individuals were re-surveyed in Germany, we can also examine changes in automation risk at the individual level.[30] Table B.6 indicates that individuals' automation risk in 2012 is negatively correlated with the subsequent change in automation risk (column 1), supporting the occupation-level evidence from Figure 2. Interestingly, this negative relationship becomes even stronger within occupations (column 2), suggesting that workers in the same occupation converge in their automation risk. Including a comprehensive set of individual-level controls for 2012 (see Section 4) hardly has any effect on the results (column 3).

The observation that automation risk is converging both across and within occupations can help to explain recent findings in the literature. For example, Bachmann et al. (2022) and Boehm et al. (2024) demonstrate that, despite the decline of routine-intensive jobs in Germany, the wages of workers who remain in these occupations have not decreased. This suggests that workers who continue to be employed in more automatable, declining occupations may be positively selected based on their productivity. Moreover, those who remain in these occupations may be performing tasks that are less susceptible to automation, and thus cannot be replaced by current technologies. Similarly, Battisti et al. (2023) show that technological and organizational changes within German firms lead to a decline in the share of routine jobs, but do not necessarily result in higher unemployment or reduced wage growth for affected workers. Instead, many workers transition to more abstract roles within the same firm, particularly in environments with strong training programs and union presence. Further supporting the relevance of within-occupational transitions, Atalay et al. (2020) highlight that there has been a substantial shift from routine tasks to non-routine interactive and analytical tasks in the United States between 1950 and 2000, with much of this shift occurring within narrowly defined job titles.

Overall, our descriptive findings indicate that task demands—and consequently, automation risk—are evolving within occupations, highlighting the importance for workers to adapt to these changing requirements. This paper investigates whether participation in job training enables workers to perform tasks that are complementary to automation technologies, thereby reducing their risk of automation and increasing their wages.

[30] The United States sampled a different population for their 2017 survey.

Figure 2: Change in Automation Risk over Time



Notes: The figure plots the average automation risk in 2012 against the change in average automation risk between 2012 and 2017 for the United States (Panel A) and between 2012 and 2015 for Germany (Panel B), by two-digit ISCO occupation. For details on the construction of the automation risk measure, see Section 3.1.
Data source: PIAAC United States (2012 and 2017 waves), PIAAC Germany (2012 and 2015 waves).

4. Empirical Strategy

4.1. Basic Empirical Model

The relationship between job training and labor market outcomes can be modeled using the following OLS regression:

$$Y_i = \alpha + \beta_1 job\_training_i + \varepsilon_i, \tag{1}$$

where Y_i represents the outcome of interest for individual i. Our two main outcome variables are the automation risk and log hourly wages. The key explanatory variable is the indicator $job\_training_i$, which equals one if a respondent participated in job-related training—such as organized sessions or training by supervisors or co-workers—within the twelve months prior to the survey, and zero otherwise (Section 2).

In this specification, β_1 is the coefficient of interest, capturing the association between job training and our outcomes. However, estimating β_1 using the naive approach outlined in Equation 1 only yields correlational evidence due to potential bias from omitted variables.[31] Specifically, we are concerned about selection into training, that is, the possibility of unobserved factors influencing both the likelihood of receiving training and the outcomes. For example, if workers with higher innate ability are more likely to receive job training, then the estimate of β_1 for wages would be upward biased since more able workers also tend to earn higher wages. Similarly, if more able workers worked in less automatable jobs, β_1 would be downward biased in the analysis of automation risk. This means that a simple comparison between workers with and without training is likely to overestimate the causal effect of training if the selection problem is not adequately addressed; this is particularly challenging in cross-sectional data.

We employ several strategies to address these concerns.[32] First, we leverage our rich survey data to include an extensive set of control variables, including a direct measure of ability. Second, we exploit within-occupation variation and apply entropy balancing

[31] Reverse causality is mitigated in our setting as individuals are asked about training measures that they completed in the twelve months preceding the outcome assessment.

[32] The existing literature addresses selection bias through the following approaches: The first involves instrumental variable approaches, yet this method struggles with the validity of the instruments (Bartel, 1995). The second, more common, approach uses fixed-effects regressions to control for unobserved individual-level heterogeneity (see e.g., Greenhalgh and Stewart, 1987; Lynch, 1992; Parent, 1999; Frazis and Loewenstein, 2005). Third, a few studies (e.g., Leuven and Oosterbeek, 2008; Goerlitz, 2011) adopt a quasi-experimental approach by using random reasons for non-participation in training among those participants who planned to participate to create treatment and control groups. They find that conventional methods are likely to overestimate the returns to training.

(Hainmueller, 2012), a reweighting technique that ensures perfect covariate balance between the treatment group (those who received training) and the control group (those who did not).[33] We also conduct a large series of robustness checks using alternative estimation approaches, including individual fixed effects models, which account for time-invariant unobserved heterogeneity (see Section 5.2 for details).

4.2. Selection into Job Training

Using PIAAC's detailed background questionnaire and the availability of our key variables at the individual level, we augment the basic OLS model by estimating the following specification:

$$Y_{icoj} = \alpha + \beta_1 job\_training_i + \beta_2 numeracy_i + \mathbf{X}_i\gamma + \delta_c + \zeta_o + \eta_j + \varepsilon_{icoj}, \qquad (2)$$

where Y_{icoj} represents the outcome of interest for individual i who lives in country c and works in occupation o and industry j. We include country fixed effects (δ_c) to account for differences in job training provision and the general quality of job training programs across countries. Additionally, two-digit industry fixed effects (η_j) are incorporated to control for variation in training frequency and effectiveness across industries. As outlined in Section 2.1, the frequency of job training varies by country and occupation. To account for within-occupation and within-country correlation, we estimate Equation 2 using two-way clustered standard errors at the occupation and country levels. This represents the most conservative approach, as alternative assumptions about the variance-covariance matrix yield smaller standard errors (see Table B.7).

A key innovation of our study is the use of an individual-level automation risk measure, based on detailed task data rather than aggregate occupational classifications (for details, see Section 3.1). This is crucial, as occupations can vary significantly in their task composition, and hence in their average automation risk (e.g., Acemoglu and Autor, 2011; Goos et al., 2014).[34] At the same time, both the demand for training (Lergetporer et al., 2023) and the take-up of training (Nedelkoska and Quintini, 2018) varies considerably across occupations. By exploiting within-occupation variation due to the inclusion of

[33] Heckman et al. (1997, 1998) and Dehejia and Wahba (2002) suggest matching estimators to create counterfactual comparison groups. Smith and Todd (2005) evaluate potential non-experimental estimators of training effectiveness and conclude that among these estimators, a matching difference-in-difference estimator performs best. The entropy balancing weights replace weights that give each country the same weight.

[34] This variation is illustrated in Figure 1 and Figure A.1.

occupation fixed effects (ζ_o), we account for unobserved occupation-specific factors that simultaneously influence both automation risk and the likelihood of receiving job training.

Another innovation of our analysis is the introduction of numeracy skills as a control variable in the job training literature. Numeracy skills provide a more accurate measure of human capital compared to previously used proxies such as years of schooling (Lynch, 1992; Arulampalam and Booth, 1997; Leuven and Oosterbeek, 1999; Bassanini et al., 2007), which have several notable limitations (for a discussion, see Hanushek and Woessmann, 2008). For instance, educational quality varies across time and countries, which is not captured by simply counting the number of school years acquired. Moreover, educational attainment measures are coarse, as individuals within the same attainment category often vary greatly when it comes to their actual human capital (Langer and Wiederhold, 2023). Furthermore, they reflect human capital only at the end of formal education, failing to account for changes in human capital throughout an individual's working life. These limitations are particularly problematic in an international setting and in the context of rapidly changing labor markets where the relevance of specific skills may evolve over time.[35]

The model also includes a broad set of control variables included in the vector $\mathbf{X}_i$, which have frequently been applied in the training literature (Oosterbeek, 1996, 1998; Lynch and Black, 1998; Grund and Martin, 2012). These covariates encompass standard socio-economic factors potentially related to training participation, such as educational attainment, age, gender, migration status, parental education, and whether the respondent has children. Since workers in larger firms and those who are full-time employed generally receive more training, we further include controls for firm size and full-time employment status (Fouarge and Schils, 2009).[36] We also account for self-organized training—activities

[35] Note that numeracy skills are assessed concurrently with our outcome measures in PIAAC. If numeracy skills also improve as a result of job training, our estimates of training effectiveness should be interpreted as lower bounds. To test this, we estimate the effect of training on numeracy skills residualized for literacy skills—that is, the portion of numeracy skills not explained by a worker's general ability (Table B.8, column 1). The estimate is close to zero and not statistically different from zero, suggesting that job training does not affect genuine numeracy skills. Contrarily, we find positive and significant estimates for residualized digital skills (Table B.8, column 2), suggesting that job training improves digital skills beyond what can be attributed to fixed general ability.

[36] More precisely, $\mathbf{X}_i'$ includes the following variables: years of schooling, age in four categories (25–34, 35–44, 45–54, and 55–65 years), gender, migration status in three categories (first-generation and second-generation migrant and native), parental education in three categories (neither parent has attained upper secondary, at least one parent has attained secondary or post-secondary/non-tertiary, and at least one parent has attained tertiary), an indicator whether the respondent has children, the age group of the oldest child in four categories (0–2, 3–5, 6–12, and 13+ years), an indicator of full-time employment, and firm size as measured by the number of employees in five categories (1–10, 11–50, 51–250, 251–1000,

initiated by the worker rather than the employer—to control for differences in motivation and effort between workers who generally participate in training and those who do not. This variable contains all training measures that are not job-related (e.g., self-organized training or seminar participation).

4.3. Entropy Balancing

To address potential biases in estimating the effect of job training on labor market outcomes, we apply entropy balancing (Hainmueller, 2012). Entropy balancing is a non-parametric pre-processing technique designed to achieve exact covariate balance between treated and control groups in observational studies. This method ensures that the covariate distributions, i.e., mean and variance, in the control group, are adjusted to match those in the treatment group, minimizing biases due to observable characteristics.[37]

The procedure works by assigning weights to the control group observations (i.e., workers who did not participate in job training in the previous twelve months) so that the weighted distribution of covariates mirrors that of the treatment group (i.e., workers who participated in job training). Specifically, entropy balancing involves minimizing a loss function that penalizes deviations from the desired covariate balance, effectively realigning the distribution of covariates in the control group to closely resemble that of the treated group. This process results in a reweighted control group that is identical to the treatment group in terms of the mean and variance of observable characteristics, differing only in terms of the treatment status after reweighting.

In our analysis, we balance the first and second moments of all variables specified in Equation 2 between treatment and control workers. The core identifying assumption for the causal interpretation of our results is that selection into job training is based on observables, meaning that all relevant variables influencing training participation are included in the model.[38] While we cannot directly test this assumption, the inclusion of numeracy skills—a strong proxy for individual ability—combined with a comprehensive set of covariates and the use of within-occupation variation gives us confidence that we have adequately accounted for the primary factors influencing selection into training.

1000+ employees). In rare cases, variables have missing values. For example, approximately one percent of respondents lack data on years of schooling. For continuous control variables, we impute missing values with the variable mean and add a variable indicating missing values. For discrete control variables, we handle missing values by creating an additional "missing" category.

[37] We implement entropy balancing by using the *ebalance* command in Stata (Hainmueller and Xu, 2013).

[38] See Cunningham (2021) for a recent discussion.

Table B.9 provides the full balancing table for our covariates. Before weighting, workers without training tend to have lower numeracy skills, are, on average, younger, less educated, and more likely to have parents with lower levels of education. They are also less likely to be employed full-time and work more often in smaller firms. After applying entropy balancing, however, the covariate distributions between workers with and without training are perfectly aligned across all these dimensions, demonstrating that the balancing exercise successfully eliminated the substantial initial differences between the two groups.

5. Results

5.1. Training and Automation

This section investigates the relationship between job training and individual-level automation risk, our main outcome of interest. Table 1 presents the main findings. The columns progressively introduce more rigorous specifications to assess the association between job training and automation risk. Column (1) shows the raw correlation within countries and industries, revealing that job training is associated with an 8.4 pp reduction in automation risk. However, the coefficient on training decreases to −5.6 pp when occupation fixed effects are included to control for unobserved occupation-specific characteristics that might influence both training and automation risk (column 2). The inclusion of occupation fixed effects also significantly enhances the model's explanatory power, as indicated by the increase in the R^2 from 0.11 to 0.22—essentially doubling the share of explained variation in automation risk. This suggests that a considerable amount of the variation in both occupational training and automation risk is at the between-occupation level, which we can account for with our measure of automation risk at the individual level by including occupation fixed effects. Importantly, even when analyzing within-occupation variation, the training coefficient remains sizable and precisely estimated.[39]

Including numeracy skills to control for unobserved ability in column (3) slightly reduces the size of the job training estimate, though not by much. Notably, when the full set of socio-demographic and work-related control variables is added in column (4) and entropy balancing is applied in column (5), the estimated coefficient on job training

[39] In the main specification, we control for occupations at the two-digit level, as this information is consistently available across all PIAAC countries. Table B.10 presents estimates that include occupation fixed effects at the more granular four-digit level for the 30 countries where such detailed occupational information is available. Reassuringly, our estimates remain robust when controlling for occupational differences at such detailed level.

remains nearly identical to that in column (3).[40] This consistency suggests that occupational selection and numeracy skills effectively act as a "sufficient statistic" for other socio-demographic and work-related differences between workers with and without training.[41]

Table 1: Training and Automation Risk

	Automation risk				
	(1)	(2)	(3)	(4)	(5)
Job training	-0.0839***	-0.0559***	-0.0511***	-0.0430***	-0.0467***
	(0.0017)	(0.0081)	(0.0068)	(0.0051)	(0.0069)
Numeracy skills			-0.0229***	-0.0114***	-0.0129***
			(0.0045)	(0.0020)	(0.0027)
Occupation FE		X	X	X	X
Further controls				X	X
Entropy balancing					X
R^2	0.11	0.22	0.22	0.24	0.20
Observations	91,470	91,470	91,470	91,470	91,470

Notes: Ordinary least squares estimation in columns (1)–(4) with weights such that each country has the same weight, least squares estimation with weights from entropy balancing in column (5). Dependent variable: individual-level automation risk, ranging from 0 (indicating a low probability that a worker is fully automated) to 1 (indicating a high probability that a worker is fully automated); automation risk is predicted using items on task use at work from PIAAC. Sample: employees aged 25–65 years with information on automation risk and log hourly wages. Job Training: binary variable indicating whether the respondent participated in on-the-job training or job-related training in the 12 months prior to the survey. Numeracy skills are standardized to unit standard deviation across countries. Further controls include: years of schooling, age group in four categories (25–34, 35–44, 45–54, 55–65), gender, migration status in three categories (first-generation migrant, second-generation migrant, native), parental education in three categories (neither parent has attained upper secondary, at least one parent has attained secondary and post-secondary/non-tertiary, at least one parent has attained tertiary), an indicator whether the respondent has children, age group of the oldest child in four categories (0–2, 3–5, 6–12, 13+), an indicator for participation in non-job-related training (i.e., open/distance education, seminars/workshops, or private lessons), an indicator for full-time employment, and firm size measured by the number of employees in five categories (1–10, 11–50, 51–250, 251–1000, 1000+). Occupation fixed effects at the two-digit ISCO level. All regressions also control for industry (two-digit ISIC) and country fixed effects. All control variables were used for the entropy balancing in column (5). R^2 refers to within-country R^2. Standard errors shown in parentheses are two-way clustered at the country and occupation level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Data source: PIAAC.

In the most rigorous specification presented in column (5) of Table 1, we find that job training reduces the automation risk by 4.7 pp. In terms of magnitude, this coefficient corresponds to approximately one-fourth of a standard deviation in individual automation risk within our sample or ten percent of the mean automation risk. To further contextualize the magnitude of the training effect, the estimated training effect is equivalent to the

[40] The coefficients in columns (4) and (5) are not statistically different from the coefficient in column (3) at conventional levels. Columns (1) through (4) use weights that give each country the same weight.

[41] Our main analysis restricts the sample to respondents that do not have missing information on both automation risk and wages (see Section 2). However, our results are very similar in a more complete sample that also has respondents with missing wage information (Table B.11).

difference in automation risk between ICT professionals (0.40) and ICT technicians (0.44) or between business administration professionals (0.39) and business administration associates (0.43), where professional occupations in the same field are typically characterized by a higher degree of complex tasks and consequently lower automation risk. To provide an example for the magnitude of the training effect within the same occupation, training would reduce the automation risk of metal and machinery workers (who are at the median of the automation risk distribution) from 58.4 percent to 53.7 percent, a decrease of nine percentiles.

Since PIAAC is an individual-level survey, our ability to control for firm characteristics is limited. However, we can approximate firm fixed effects in the data by adding region×industry×firm size fixed effects. Table B.12 presents estimates using this approach in the 30 countries where detailed regional data are available. Reassuringly, the estimates remain robust even in this more stringent specification. Thus, we are confident that our identifying assumption—namely, that our entropy balancing approach adequately accounts for all variables influencing selection into training—is valid. Below, we provide additional robustness checks on this assumption, using coefficient bounds for unobservable selection as well as country-specific repeated cross-sectional and panel data.

5.2. Robustness

While we utilize rich individual-level data to account for selection into training, PIAAC's cross-sectional data do not provide information on pre-training automation risk. This limitation leaves room for potential selection bias if workers with either higher (or lower) pre-training automation risk are more likely to receive training. For example, firms might prefer to invest in training for workers already engaged in less automatable tasks, as the cost of additional training for these workers might be lower. Conversely, firms might choose to train those with a higher pre-training automation risk in order to ensure their skills remain aligned with evolving production technologies.

In this section, we conduct several robustness checks to demonstrate that unobserved selection into training is unlikely to bias our results. Below, we first apply Oster bounds to assess the influence of unobserved variables in the international analysis. Furthermore, we exploit the repeated U.S. PIAAC data to impute workers' past automation risk at the individual level and construct a matched pseudo-panel. Finally, we leverage the German PIAAC panel data to estimate value-added and panel models. Each of these approaches consistently shows that training significantly reduces the risk of automation.

Unobservable Selection and Coefficient Bounds. In order to address concerns about selection on unobservables, we apply the coefficient bounds methodology proposed by Oster (2019). Intuitively, this approach allows us to assess how the estimated training coefficient would change if we could account for unobserved variables. The method is based on two key parameters: the relative degree of selection on unobservables (denoted as δ)[42] and the maximum possible R-squared (R_{max}) that could be achieved if all relevant variables, including unobserved ones, were included in the model.

The intuition behind the Oster bounds lies in the idea that coefficient stability, when introducing additional controls in a hypothetical "long" regression that includes unobservables, can provide insights into the potential bias from unobserved variables. If the coefficient remains stable when additional controls are included, it suggests that the unobserved variables do not meaningfully bias the estimate. By scaling the observed movements in the coefficient with changes in R-squared, the Oster bounds provide an estimate of how much the coefficient would change when accounting for unobservables, given a specified hypothetical maximum R-squared from adding observables and unobservables, R_{max}, as a factor of $\tilde{R}$, the R-squared of our baseline regression in column (5) of Table 1.

In our analysis, the coefficients on job training remain sizable and highly statistically significant for a range of plausible assumptions on δ and R_{max} (Table B.13). For instance, when following the recommendation in Oster (2019) and the application in Chen (2021) to set $\delta = 1$ and $R_{max} = 1.4$, the training coefficient is still at 0.0366, with a standard error of 0.0018. Even when δ is set to 1.2—assuming unobservable factors are 20 percent more influential than observable ones—the training coefficient only slightly decreases to 0.0338, with a standard error of 0.0020. This coefficient stability indicates that our training estimates are robust to potential omitted variable bias from unobserved factors.

U.S.-Specific Analysis. Additionally, we make use of the repeated cross-sectional PIAAC data for the United States in the years 2012 and 2017. In this U.S. analysis, we use the no-training group from 2012 to impute an individual-level past automation risk for both training and no-training workers in the 2017 sample. To achieve this, we impute time-varying variables, such as the respondent's age or the age of their oldest child, back to their 2012 values.[43] Specifically, we regress automation risk on all covariates included in our baseline model (see column 5 of Table 1), but only for the no-training group in 2012.

[42] For instance, $\delta = 1$ assumes that unobservable factors were as influential as observable ones in determining the outcome.

[43] Note that this analysis assumes that workers did not change occupations, industries, or employers between 2012 and 2017.

We then apply the coefficients from this regression to predict the past automation risk of individuals observed in the 2017 sample. Results are shown in Panel A of Table 2. When we add the imputed past automation risk as an individual-level control in column (2), the training coefficient remains virtually unchanged compared to the baseline estimate in column (1).

Additionally, the availability of two waves of PIAAC data in the U.S. allows us to construct a matched pseudo-panel. We draw inspiration from Kleven et al. (2024), who estimated earning losses for women following childbirth using cross-sectional data in the absence of individual-level panel data. In their approach, Kleven et al. (2024) leveraged repeated cross-sections to construct pseudo-panel data by matching individuals across survey waves based on observable characteristics. Matched individuals in earlier cross-sections serve as surrogate past observations for those in later cross-sections. Applying this method to our context, we match respondents from the 2017 PIAAC wave to respondents in the 2012 wave. The matching is performed within cells defined by fixed characteristics of gender, age group, parental and migration background, level of education, and quintiles of numeracy skills.[44] Ultimately, we successfully matched 1,396 observations from the 2017 U.S. sample to corresponding observations from 2012, which serve as surrogate past observations. Using this matched pseudo-panel, we can add individual fixed effects to the analysis (i.e., fixed effects for the matched group). Panel B of Table 2 presents the results. The training estimate using our matched pseudo-panel of −3.7 pp (column 2) is very similar to the cross-sectional estimate in the 2017 U.S. sample (column 1)[45]. Notably, the estimate is also consistent with the cross-sectional estimate that includes imputed past automation risk as an individual-level control, shown in column (2) of Panel A.

Germany-Specific Analysis. We can even go beyond the pseudo-panel approach using panel data for Germany, which allows researchers to track changes in training participa-

[44] To ensure a sufficient number of matches within these exact cells, we use the 2012 Canadian PIAAC sample due to its larger number of observations, as Canada is the only PIAAC country that surveyed enough respondents to achieve representativeness at the sub-national level. We conduct two checks to confirm that both countries are comparable in the no-training sample. First, conditional on all characteristics that we include in our entropy balancing, this risk of automation for workers without job training does not differ between the Canadian and U.S. samples in 2012. Second, in a regression of individual-level automation risk on covariates and country fixed effects, the country fixed effect for Canada is not statistically significant. That is, differences in automation risk between the Canadian and U.S. samples are due to compositional differences with respect to, say, occupations, and not due to fundamental differences in automation risk between both countries.

[45] This refers to the estimate in the 2017 sample that could be matched to corresponding observations from 2012.

Table 2: U.S. Evidence Linking Two PIAAC Waves

Panel A: Controlling for Imputed Past Automation Risk

	Automation risk (2017)	
	(1)	(2)
Job training	-0.0403	-0.0409
	(0.0101)	(0.0087)
Numeracy skills	0.0063	0.0104
	(0.0056)	(0.0061)
Imputed automation risk (2012)		1.0879
		(0.6498)
Controls	X	X
R^2	0.33	0.35
Observations	1,554	1,554

Panel B: Pseudo Panel

	Automation risk	
	(1)	(2)
Job training	-0.0457	-0.0372
	(0.0106)	(0.0081)
Controls	X	Time-varying
Match/Worker FE		X
R^2	0.34	0.45
Observations	1,396	2,792

Notes: Dependent variable: individual-level automation risk, ranging from 0 (indicating a low probability that a worker is fully automated) to 1 (indicating a high probability that a worker is fully automated); automation risk is predicted using items on task use at work from PIAAC. Sample: employees aged 25–65 years in the United States, wave 2017 (Panel A) and waves 2012 and 2017 (Panel B). Job Training: binary variable indicating whether the respondent participated in on-the-job training or job-related training in the 12 months prior to the survey. Panel A: Imputed automation risk 2012: calculated using the coefficients obtained from a regression that includes all the covariates used in entropy balancing (see main text) within the group that did not receive any training in the 2012 sample. Age in 2012 for the 2017 sample is imputed from the 2017 values. Panel B: Sample in pseudo-panel consists of individuals surveyed in the 2017 U.S. survey wave and matched individuals in U.S. and Canada in 2012 survey wave. Matching based on the set of fixed characteristics in our control variables: gender, age group in four categories (25–34, 35–44, 45–54, 55–65), level of education in four categories (less than 9, 9-12, 13-16, more than 16 years of education), migration status in three categories (first-generation migrant, second-generation migrant, native), parental education in three categories (neither parent has attained upper secondary, at least one parent has attained secondary and post-secondary/non-tertiary, at least one parent has attained tertiary, quintile of numeracy skills. Column (2) includes survey year fixed effects, all other potentially time-varying controls as well as fixed effects for the matched worker pairs. Standard errors shown in parentheses are clustered at the occupation level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Data source: PIAAC United States (2012 and 2017 waves).

tion and job tasks over time at the individual level. By leveraging these data, we can directly observe how training influences automation risk when controlling for past values of automation risk, training, and even earnings; this mitigates concerns about selection bias that may arise in the cross-sectional data.

In Table 3, we estimate value-added models using both the 2012 and 2015 automation risk measures (see Section 3).[46] Specifically, we regress the 2015 automation risk on job training in the prior year, controlling for the automation risk in 2012, along with other covariates measured in 2012. All models apply entropy balancing, ensuring exact balancing between the training and no-training groups on the control variables included in the respective specification.[47] The results indicate that job training is strongly negatively correlated with automation risk in 2015, consistent with our main findings from the international analysis. Notably, the training estimate is only subject to a small reduction when moving from the model with standard pre-training controls (column 1) to the model that also includes pre-training automation risk (column 2). The training estimate remains virtually unchanged after additionally controlling for prior training participation (column 3) and even when accounting for pre-training hourly wages (column 4). This consistency in the magnitude of the training coefficient, even after controlling for variables that could reasonably be associated with selection into training (e.g., based on automation risk or wage level), suggests that our approach of leveraging within-occupation variation and utilizing rich individual controls, including tested numeracy skills, effectively addresses selection concerns. This further bolsters confidence in the findings from the international analysis.

In the full-control model presented in column (4) of Table 3, job training reduces automation risk by 6 pp. The panel dimension of the German PIAAC data allows us to benchmark this estimate in several ways. First, the reduction in automation risk through training is nearly twice the average change in automation risk observed between 2012 and 2015 (3.3 pp). Second, training proves to be more effective than switching occupations in mitigating automation risk. When replacing the training variable with an indicator for having changed two-digit occupations between 2012 and 2015 in the model from column (4), the coefficient on switching occupation is -0.0260 ($p = 0.004$). This indicates that training is more than twice as effective as changing occupations when it comes to reducing automation risk (results not shown).

[46] See Section 3.1 for details on how the automation measure is constructed in each wave.

[47] Standard errors are clustered at the two-digit occupation level (42 clusters) for consistency with the international analysis. Clustering at the four-digit occupation level (304 clusters) yields very similar results.

Table 3: Training and Automation Risk: Value-Added Approach for Germany

	Automation risk (2015)			
	(1)	(2)	(3)	(4)
Job training (2014)	-0.0762***	-0.0627***	-0.0590***	-0.0599***
	(0.0112)	(0.0107)	(0.0102)	(0.0103)
Automation risk (2012)		0.3210***	0.2904***	0.2924***
		(0.0399)	(0.0403)	(0.0424)
Job training (2012)			-0.0425***	-0.0426***
			(0.0120)	(0.0121)
Log hourly wage (2012)				-0.0251
				(0.0175)
Numeracy skills (2012)	-0.0228*	-0.0168	-0.0131	-0.0099
	(0.0116)	(0.0108)	(0.0121)	(0.0121)
Occupation FE (2012)	X	X	X	X
Further controls (2012)	X	X	X	X
Entropy balancing	X	X	X	X
R^2	0.38	0.42	0.42	0.43
Observations	1,585	1,585	1,585	1,585

Notes: Ordinary least squares estimation with weights from entropy balancing. Dependent variable: individual-level automation risk, elicited in the 2015 PIAAC survey. Automation risk ranges from 0 (indicating a low probability that a worker is fully automated) to 1 (indicating a high probability that a worker is fully automated); automation risk is predicted using items on task use at work from PIAAC. Sample: employees aged 25–65 years in Germany. Job training, elicited in the 2015 PIAAC survey, refers to the year 2014. All control variables were elicited in the 2012 PIAAC survey. Occupation fixed effects are measured at the two-digit ISCO level. Numeracy skills are standardized to unit standard deviation within Germany. Further controls include: years of schooling, age group in four categories (25–34, 35–44, 45–54, 55–65), gender, migration status in three categories (first-generation migrant, second-generation migrant, native), parental education in three categories (neither parent has attained upper secondary, at least one parent has attained secondary and post-secondary/non-tertiary, at least one parent has attained tertiary), an indicator whether the respondent has children, age group of the oldest child in four categories (0–2, 3–5, 6–12, 13+), an indicator for participation in non-job-related training (e.g., self-organized training or seminar participation), an indicator for full-time employment, firm size measured by number of employees in five categories (1–10, 11–50, 51–250, 251–1000, 1000+), and industry fixed effects (two-digit ISIC). Standard errors shown in parentheses are clustered at the occupation level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Data source: PIAAC Germany (2012 and 2015 waves).

Furthermore, observing workers' occupations in both 2012 and 2015 in the German PIAAC data allows us to test a key assumption underlying our international analysis—namely, that training is unrelated to switching occupations. Our individual-level measure of automation risk allows us to include occupation fixed effects to estimate the effect of job training within the same occupation, thereby controlling for differences in training participation and automation risk between occupations. However, if workers were more likely to receive training after switching occupations, the within-occupation comparison would be compromised. In such a scenario, the no-training workers in the observed occupation would not serve as an appropriate (no-training) counterfactual for workers who switched occupations and thus received training. For a valid within-occupational comparison of training and no-training groups, it would either be necessary to know the training participants' previous (pre-training) occupation or be sure that training and occupational switching are unrelated. While we cannot track the occupations of workers prior to training in the cross-sectional PIAAC data, the German PIAAC panel provides information on participants' occupations in both 2012 and 2015. This allows us to exploit information on workers' occupation and training participation within individuals over time. We observe that switching occupation between 2012 and 2015 is completely unrelated to training participation in 2015 ($\rho = 0.0084$). This finding supports the validity of our international analysis, confirming that training and occupational switching are indeed unrelated, corroborating the validity of our within-occupation comparisons.

Finally, we estimate a panel model with individual and survey year fixed effects in the German PIAAC data, allowing for a more rigorous analysis that controls for unobserved time-invariant characteristics of individuals as well as time-specific effects. In this demanding model, presented in Table 4, we continue to find that job training reduces automation risk, reinforcing the robustness of our main findings. However, it is important to note that the coefficient on training in this fixed effects model is considerably smaller and less precisely estimated compared to the value-added model. This reduction in magnitude and precision likely reflects the limited amount of over-time variation in both automation risk and training within individuals (original survey and follow-up are just 3.5 years apart), making it more challenging to detect significant training effects. Nonetheless, the consistency of our findings across different models and sources of variation is reassuring and supports the overall robustness of our findings.

Training Effectiveness Over Time. Since the United States and Germany participated in PIAAC twice, we have the opportunity to examine how the effectiveness of job training in reducing automation risk has changed over time. Table B.14 leverages this time dimension by estimating our baseline model (from column 5 of Table 1) by wave, separately for the United States (columns 1–3) and Germany (columns 4–6). The results indicate that training has become more effective in reducing automation risk over time. Consistently in both countries, the effectiveness of training roughly doubles between the two waves.[48] While the observed increase in training effectiveness could be due to sampling variation

[48] Note that in the 2015 PIAAC wave in Germany, information on children and non-work-related training was not collected. To ensure comparability across specifications, these control variables are excluded in all regressions for Germany.

Table 4: Training and Automation Risk: Panel Analysis for Germany

	Automation risk			
	(1)	(2)	(3)	(4)
Job training	-0.0145	-0.0151	-0.0149	-0.0129
	(0.0088)	(0.0090)	(0.0090)	(0.0089)
Individual FE	X	X	X	X
Survey year FE	X	X	X	X
Industry FE		X	X	X
Occupation FE			X	X
Further controls				X
R^2	0.72	0.73	0.75	0.75
Individuals	1,869	1,869	1,869	1,869
Observations	3,738	3,738	3,738	3,738

Notes: Fixed effects panel estimation, using the 2012 and 2015 waves of PIAAC Germany. Dependent variable: individual-level automation risk, ranging from 0 (indicating a low probability that a worker is fully automated) to 1 (indicating a high probability that a worker is fully automated); automation risk is predicted using items on task use at work from PIAAC. Sample: employees aged 25–65 years in Germany (restrictions refer to the values in the 2012 survey). Industry fixed effects are measured at the two-digit ISIC level, occupation fixed effects are measured at the two-digit ISCO level. Further controls are time-varying and include: age, an indicator for full-time employment, and firm size measured by number of employees in five categories (1–10, 11–50, 51–250, 251–1000, 1000+). Observations are not weighted. Standard errors shown in parentheses are clustered at the individual level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Data source: PIAAC Germany (2012 and 2015 waves).

between the two survey waves (e.g., Meager, 2019), it may also reflect true temporal heterogeneity driven by factors such as shifts in labor market dynamics or changes in training content over time in response to an evolving technological landscape. For example, demographic changes that lead to a shrinking labor force (e.g., Acemoglu and Restrepo, 2022), declining labor force participation (e.g., Dotsey et al., 2017), or a tighter labor market due to the United States' extended recovery from the Great Recession (e.g., Cunningham, 2018), could incentivize employers to invest in more effective training programs that help incumbent employees adapt to changing task demands. Additionally, the content of training may have evolved, with newer technologies requiring different task inputs (e.g., Atalay et al., 2020) that are less susceptible to automation.[49]

[49] We do not observe notable differences in the duration or financing of training in the United States between 2012 and 2017.

5.3. Mechanisms

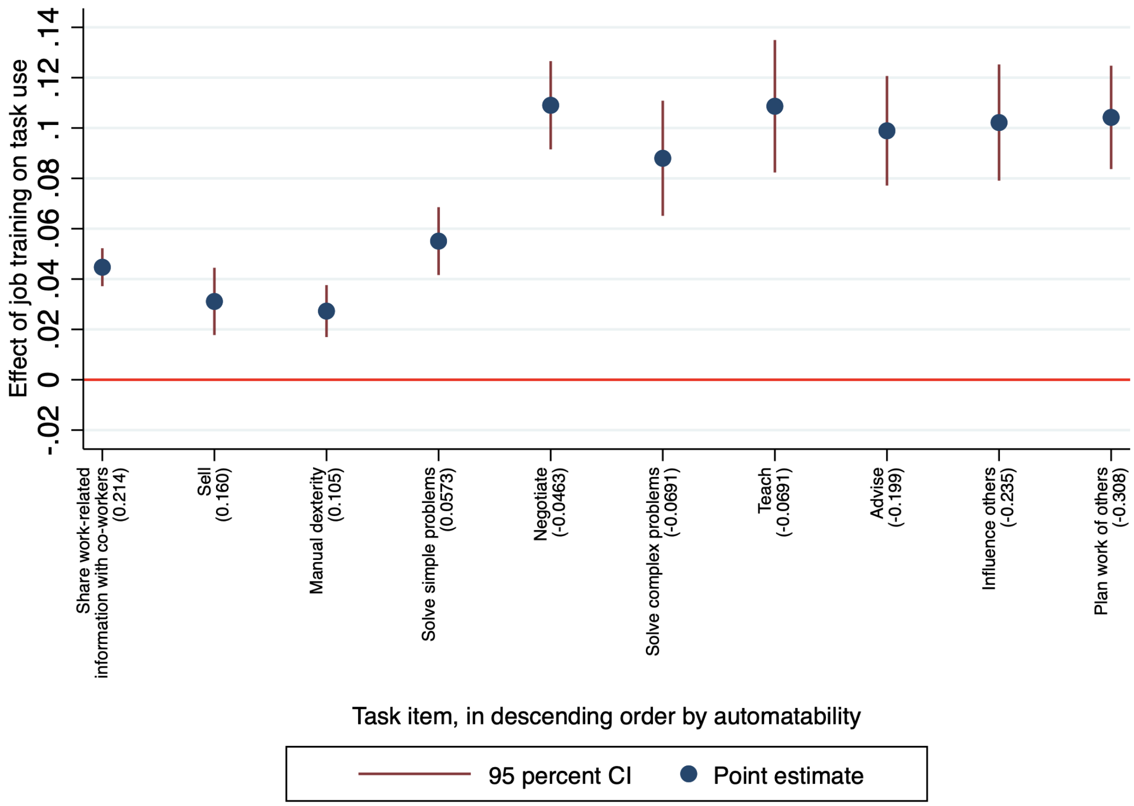
Understanding the mechanisms through which job training reduces automation risk is crucial for developing effective policies. In this section, we first identify which specific job tasks are enhanced by training, and thus reduce the likelihood of automation. We also investigate the role of digital skills as a potential channel through which training impacts the risk of automation.

Job Tasks. Our composite measure of automation risk in Table 1 is derived from various tasks that workers perform in their jobs. To better understand the role of training in influencing these tasks, we estimate our baseline specification using individual job tasks as outcome. Results are presented in Figure 3, where task items are ordered by their contribution to automation risk.[50] The results indicate that workers who receive training engage in *all* tasks more frequently than those without training. However, the increase in task intensity is particularly pronounced for tasks associated with a lower risk of automation. For instance, training participation increases the likelihood that workers frequently engage in complex problem-solving (by 8.8 pp) and in interaction-intensive activities, such as influencing others (by 10.2 pp) and negotiating (by 10.9 pp). This finding aligns with the work of Deming (2017), who argues that social interaction tasks are difficult to automate, making them increasingly valuable in the labor market. These tasks are also likely to complement automation technology (Acemoglu and Autor, 2011), further reducing automation risk. For example, workers who participated in training subsequently perform manual dexterity tasks (associated with a higher automation risk) only 2.6 pp more often than workers without training, while they perform planning the work of others (associated with a lower automation risk) 10.2 pp more frequently.

Digital Skills. Next, we investigate the role of digital skills as a potential channel through which training impacts the risk of automation. Digital skills are particularly relevant as a mechanism for several reasons. First, as the modern economy becomes increasingly digitized, workers with strong digital skills are better equipped to adapt to new technologies and perform tasks that are complementary to automation as opposed to those that can

[50] The task items in PIAAC are measured by the frequency of task use on a Likert scale ranging from one (never) to five (every day). To facilitate interpretation, we dichotomize this variable for the analysis in Figure 3, defining a binary task variable equal to one if the respondent performs the task at least once a week (values four and five of the original task item). The binary task variable takes a value of zero if a task is performed less than once a week (values one, two, or three). This allows us to obtain OLS estimates analogous to those in Table 1, which are more straightforward to interpret compared to coefficients from an ordered logit model typically applied to categorical Likert responses (Cameron and Trivedi, 2005). The results in Figure 3 are robust to different specifications of the task-use cutoff.

Figure 3: Training and Task Use



Notes: The figure shows how job training affects job tasks related to automation. Estimates are based on our baseline specification in column (5) of Table 1, using individual job tasks instead of the composite automation measure as outcome. The values of the responses to the task items elicited in PIAAC represent the frequency of task use given on a Likert scale ranging from 1 (never) to 5 (every day). For ease of interpretation, we binarize the task items (1: task is performed at least once a week; 0: task is performed less than once a week). Training coefficients are shown in descending order by automatability; automatability weights (i.e., factor loadings from Table B.3) are shown on the horizontal axis in parentheses. 95 percent confidence intervals are based on standard errors two-way clustered at the country and occupation level.
Data source: PIAAC.

be substituted by it (e.g., Krueger, 1993; Acemoglu and Pischke, 1998; Caunedo et al., 2023). This means that digital skills can help workers transition into roles that are less likely to be automated, such as those involving complex problem-solving, decision-making, and social interaction—areas where human oversight continues to be essential.[51] Second, digital skills enable workers to leverage automation technologies more effectively, allowing them to take on new, technology-enhanced roles. For instance, workers who are proficient in digital tools may be able to use advanced software for planning, coordination, and

[51] For instance, digital technologies might enable workers to effectively manage and plan the work of others using computer-aided tools.

analysis, which not only improves their productivity, but also reduces their exposure to routine tasks at higher risk of automation.

To further explore the role of digital skills as a mechanism, Table 5 examines how training enhances these skills, potentially enabling workers to perform more complex, less automatable tasks. Using the same empirical strategy as in Table 1, column (1) of Table 5 shows that workers who receive training have digital skills that are 0.30 standard deviations higher than those of workers without training. While the coefficient decreases to 0.21 when occupation fixed effects are included in column (2), the most significant reduction occurs in column (3) when numeracy skills are added as a control. This is not surprising, given the high correlation between numeracy and digital skills ($\rho = 0.74$). In the most demanding specification, which incorporates additional controls and applies entropy balancing (column 5 of Table 5), job training is associated with an increase in digital skills of 0.051 standard deviations. This effect size is roughly 15 percent of the average difference in digital skills between an ICT professional (standardized digital skill score of 0.847) and a business and administration professional (0.472). Alternatively, it represents about 30 percent of the difference in digital skills between workers aged 25 to 34 (standardized digital skill score of 0.241) and those aged 35 to 44 (0.085). These findings suggest that enhancing digital skills through job training may enable workers to perform more complex tasks less susceptible to automation.

The PIAAC assessment of digital skills was optional, allowing countries to opt out entirely. In participating countries, only those participants who demonstrated basic computer skills and confidence in computer usage were assessed (Section 2.1). This implies sample selection since participation in the digital skill assessment is potentially correlated with unobserved factors such as ability, motivation, and effort. In Table B.15, we present results after imputing missing digital skills in various ways. Column (1) replicates the preferred specification from column (5) of Table 5. In columns (2)–(4), we impute missing digital skills with zero, the global minimum, and the country-specific minimum, respectively. As anticipated, the training estimate increases in these more inclusive samples, as individuals with imputed digital skills are less likely to engage in job training. Consistently across all imputation methods, the training estimate more than doubles. Therefore, we interpret the training estimate in the baseline sample without imputed digital skills as a lower bound of the true effect of job training on digital skills.

Additionally, we consider the mere fact of whether an individual participated in the digital skill assessment as indicative of their basic digital skills; individuals without any computer experience or those who failed a simple initial computer test are likely to pos-

Table 5: Training and Digital Skills

	Digital skills				
	(1)	(2)	(3)	(4)	(5)
Job training	0.2999***	0.2116***	0.0879***	0.0770***	0.0509***
	(0.0198)	(0.0188)	(0.0122)	(0.0108)	(0.0088)
Numeracy skills			0.8213***	0.7778***	0.7762***
			(0.0139)	(0.0126)	(0.0112)
Occupation FE		X	X	X	X
Further controls				X	X
Entropy balancing					X
R^2	0.08	0.14	0.54	0.58	0.59
Observations	72,180	72,180	72,180	72,180	72,180

Notes: Ordinary least squares estimation in columns (1)–(4), least squares estimation with weights from entropy balancing in column (5). Dependent variable: digital skills standardized to standard deviation 1 across countries. Sample: employees aged 25–65 years with information on digital skills, automation risk, and log hourly wages. Job Training: binary variable indicating whether the respondent participated in on-the-job training or job-related training in the 12 months prior to the survey. Controls: numeracy skills (standardized to unit standard deviation across countries), years of schooling, age group in four categories (25–34, 35–44, 45–54, 55–65), gender, migration status in three categories (first-generation migrant, second-generation migrant, native), parental education in three categories (neither parent has attained upper secondary, at least one parent has attained secondary and post-secondary/non-tertiary, at least one parent has attained tertiary), an indicator whether the respondent has children, age group of the oldest child in four categories (0–2, 3–5, 6–12, 13+), an indicator for participation in non-job-related training (i.e., open/distance education, seminars/workshops, or private lessons), an indicator for full-time employment, and firm size measured by the number of employees in five categories (1–10, 11–50, 51–250, 251–1000, 1000+). Occupation fixed effects at the two-digit ISCO level. All regressions also control for industry (two-digit ISIC) and country fixed effects. All control variables were used for the entropy balancing in column (5). R^2 refers to within-country R^2. Standard errors shown in parentheses are two-way clustered at the country and occupation level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Data source: PIAAC.

sess very limited to no digital skills. Intriguingly, we find that job training significantly increases the probability of individuals having at least basic digital skills by 4 pp in column (5) of Table B.16, corresponding to 5 percent of the average probability of having basic digital skills (78 percent).

5.4. Training and Wages

In this section, we analyze the impact of job training on wages. Higher wages typically reflect increased productivity, which can result from enhanced skills, better job performance, and greater adaptability to technological changes. Understanding the relationship between training and wages provides insights into how training contributes towards individual labor market success.

Table 6 presents the relationship between job training and log hourly wages. As with our analysis of automation risk, column (1) demonstrates the association between training and wages, controlling only for country and industry fixed effects. In this specification, workers who have received training earn 20.8 percent more than workers without training. However, similar to the findings in Table 1, the coefficient drops substantially after we include occupation fixed effects in column (2), and it further decreases when numeracy skills are added as control in column (3). This suggests that a considerable portion of the initial wage premium associated with training is explained by occupational sorting and selection into training based on ability. When additional worker and firm characteristics are included in column (4), the training coefficient only changes slightly, indicating that these factors have a limited impact on the estimated effect of training. After applying entropy balancing in column (5), we find that job training increases wages by 8.2 percent. While previous country-specific analyses of training effects on wages reveal highly heterogeneous effect sizes—varying by country, period of analysis, and estimation approach (see Online Appendix A in Ma et al., 2024)—our estimate falls roughly in the middle of this range. The estimated 8 percent wage increase is sizable, comparable to the wage gains associated with an additional year of schooling in industrialized countries (see Table A.2 in Hanushek et al., 2015).[52]

These results corroborate the notion that job training enhances workers' productivity by shifting their task composition toward less automatable, more technology-complementary activities. This shift increases workers' marginal product of labor, leading to higher wages. But how important is a reduced risk of automation for the wage gains from training? One simple way to address this question is to add automation risk as an additional control when examining the relationship between training and wages. In column (6), we observe a reduction in the estimated wage effect of training by about one-fifth. This suggests that a sizable portion of the wage gains from training can be attributed to its role in lowering workers' risk of automation.

6. Heterogeneity in Training Effectiveness

This section investigates the effects of job training on automation risk and wages, with a focus on differences across countries, socio-demographic groups, and types of training. We begin by exploring how the effectiveness of training varies by country, considering diverse institutional frameworks and cultural contexts. We then examine heterogeneity in training effects by age and gender, which may influence how much workers are exposed to technological change. Finally, we analyze the impact of different types of training to understand which forms of training are most effective in reducing automation risk and increasing wages.

[52] The results remain virtually unchanged when we include the full sample without restricting it to workers with information on automation, as there are very few missing values related to job tasks and the corresponding risk of automation.

Table 6: Training and Wages

	Log hourly wages					
	(1)	(2)	(3)	(4)	(5)	(6)
Job training	0.2082***	0.1336***	0.1131***	0.0890***	0.0824***	0.0679***
	(0.0175)	(0.0160)	(0.0122)	(0.0078)	(0.0117)	(0.0084)
Numeracy skills			0.0972***	0.0617***	0.0716***	0.0676***
			(0.0126)	(0.0069)	(0.0095)	(0.0085)
Automation risk						-0.3116***
						(0.0449)
Occupation FE		X	X	X	X	X
Further controls				X	X	X
Entropy balancing					X	X
R^2	0.16	0.27	0.29	0.36	0.35	0.36
Observations	91,470	91,470	91,470	91,470	91,470	91,470

Notes: Ordinary least squares estimation in columns (1)–(4), least squares estimation with weights from entropy balancing in column (5). Dependent variable: log hourly wages, excluding bonuses for wage and salary workers. Sample: employees aged 25–65 years with information on automation risk and log hourly wages. Job training: binary variable indicating whether the respondent participated in on-the-job training or job-related training in the 12 months prior to the survey. Numeracy skills are standardized to unit standard deviation across countries. Further controls: years of schooling, years of schooling, age group in four categories (25–34, 35–44, 45–54, 55–65), gender, migration status in three categories (first-generation migrant, second-generation migrant, native), parental education in three categories (neither parent has attained upper secondary, at least one parent has attained secondary and post-secondary/non-tertiary, at least one parent has attained tertiary), an indicator whether the respondent has children, age group of the oldest child in four categories (0–2, 3–5, 6–12, 13+), an indicator for participation in non-job-related training (i.e., open/distance education, seminars/workshops, or private lessons), an indicator for full-time employment, and firm size measured by the number of employees in five categories (1–10, 11–50, 51–250, 251–1000, 1000+). Occupation fixed effects at the two-digit ISCO level. All regressions also control for industry (two-digit ISIC) and country fixed effects. All control variables were used for the entropy balancing in columns (5) and (6). R^2 refers to within-country R^2. Standard errors shown in parentheses are two-way clustered at the country and occupation level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Data source: PIAAC.

6.1. Heterogeneity by Country

Examining the returns to training by country is crucial, as it allows us to assess the consistency of our findings across diverse institutional frameworks and cultural contexts. Our international dataset is uniquely suited for this purpose, enabling us to explore how training impacts workers in different countries and to verify whether the effects we observe hold true across a wide range of settings.

Figure A.3 and Figure A.4 show how job training affects automation risk and log hourly wages for each country individually, using our main estimation model outlined in Equation 2. We observe substantial cross-country heterogeneity in the effectiveness of training. Specifically, the reduction in automation risk due to training ranges from just 1 pp in Turkey to as much as 7 pp in Canada. Similarly, the effect of training on wages varies widely, from a modest 1 percent increase in France to a substantial 21 percent increase in Mexico.[53] These variations highlight the critical role of institutional and cultural contexts in shaping the effectiveness of training programs.[54] Despite this cross-country variation in training returns, the effect of training on both automation risk and wages is statistically significant in almost all countries in our sample.[55] This consistency underscores the external validity of our findings across diverse economic and institutional environments.

6.2. Heterogeneity by Gender and Age

Next, we examine heterogeneity in the effectiveness of job training effects by age and gender. This analysis reveals how training interventions can differentially benefit specific demographic groups that may be more vulnerable to technological changes in the labor market. As labor forces in industrialized countries age, it becomes increasingly important to equip older workers with the skills necessary to stay competitive. Similarly, understanding gender differences in training effectiveness can guide firms and policymakers in using training as a tool to mitigate gender disparities in the workplace.

Figure A.5 illustrates how training effectiveness varies by gender during the life cycle. The differences in training effectiveness between men and women tend to widen with age, indicating that training becomes increasingly beneficial for women relative to men as they grow older. For automation risk (left panel), training has similar effects for men and women at younger ages (25 to 34). However, as individuals age, training effectiveness diminishes for men while it increases for women, leading to a growing gender divergence. Specifically, in the 45–54 age group, the effect of training on automation risk is 1.37 pp larger for women compared to men ($p = 0.032$). In the 55–65 age group, this gender gap is even at 1.5 pp, though it is not statistically significant ($p = 0.161$). This pattern suggests that older women gain more from training in terms of reducing automation risk than their male counterparts.

The results for wages (right panel) reveal a similar gender trend. While training returns increase for both men and women from age 35 onward, the growth is steeper

[53] Bassanini et al. (2007) find wage returns to training across European countries ranging from 3.7 to 21.6 percent.

[54] In unreported analyses, we also examined whether differences in training effects on automation risk are systematically related to features of country economies, such as labor market regulations and pension generosity. However, we found little evidence to support this.

[55] The training coefficient is not statistically significant at the 5 percent level in Czech Republic, Kazakhstan, Slovakia, Spain, and Turkey for automation risk and in Chile, France, New Zealand, and Turkey for wages.

for women, further highlighting the greater benefits of training for women later in their careers.

Potential explanations for the observed gender-age heterogeneities include older women's specialization in automation-resistant tasks, compensatory skill investments following earlier career interruptions, and their greater representation in automation-complementary sectors (e.g., healthcare and education). However, a deeper exploration of these factors is beyond the scope of this paper.

6.3. Heterogeneity by Education Level

Figure A.6 shows the effect of job training on automation risk and wages by education level, categorizing workers as low (lower secondary degree or less), medium (upper secondary and post-secondary degree), or high (tertiary degree) educated. Notably, we find no significant differences in training effectiveness across these education levels. While the literature on skill- and task-biased technological change suggests that technology typically complements higher-educated workers and substitutes lower-educated ones (e.g., Acemoglu and Autor, 2011; Goos et al., 2014), our results highlight the potential of job training to bridge outcome gaps across education groups.

6.4. Heterogeneity by Type of Training

Finally, we examine heterogeneity in training effectiveness by its duration and financing. Dividing the sample into quintiles based on training duration, ranging from less than one day in the first quintile to over 14 days in the fifth, Figure A.7 demonstrates that the effectiveness of training on automation risk and wages generally increases with duration. However, training lasting more than 14 days, while still significant and substantial, is less effective than training in the 6- to 14-day range.[56] The overall positive relationship between training duration and returns suggests that the benefits of training are driven by skill acquisition and content rather than signaling effects alone.

Regarding financing, Figure A.8 indicates that job training is most effective for reducing automation risk and increasing wages when fully or partly financed by the employer (approximately 70 percent of workers in our sample), compared to training not financed by the employer or offered free of charge. Consistent with the findings on training duration, this suggests that productivity gains from training are more likely when employers anticipate a return on their investment in training outcomes.

[56] This may reflect selection effects. Workers with lower levels of education generally receive shorter training (Table B.1). However, workers in the highest quintile of training duration tend to have lower education levels and numeracy skills compared to those in the third and fourth quintile.

7. Conclusions

Automation has become a transformative force in the labor market, posing significant challenges to workers whose jobs are heavily characterized by routine tasks. While the risk of automation is largely determined by the specific tasks a worker performs, previous literature has almost exclusively relied on occupation-level measures of automation risk, thereby neglecting the significant variation within occupations. Our study addresses this limitation by leveraging rich task data to construct an individual-level measure of automation risk, capturing the heterogeneity in automation risk among workers within the same occupation.

Using data from 37 industrialized countries, and more than 90,000 workers, we examine how job training can help workers adapt to evolving task requirements and reduce their susceptibility to automation. To tackle the potential endogeneity of training participation, we apply entropy balancing on a comprehensive set of worker characteristics, including tested numeracy skills to control for unobserved ability. Drawing on within-country, within-industry, and within-occupation variation to identify training effects, our analysis overcomes the identification challenges that typically plague cross-country studies.

Our results show that job training significantly reduces automation risk, with training participants experiencing a reduction in automation risk of 4.7 pp—equivalent to a ten percent decrease in the average automation risk. Additionally, workers who undergo training receive approximately eight percent higher wages compared to those without training. While the positive wage effects of training are well-documented in the literature, our study is the first to demonstrate that approximately one-fifth of these wage gains can be directly attributed to the reduction in automation risk. This underscores the importance that reduced susceptibility to automation through training plays in improving workers' labor market prospects.

Training effectively mitigates automation risk and enhances wages in nearly all countries, underscoring the external validity of our findings. Examining individual-level heterogeneity, we find that training effectiveness increases over the career for women, contradicting the perception that older workers struggle to acquire skills that complement new technologies. Moreover, women generally benefit more from training than men, suggesting its potential to help reduce gender disparities in the labor market. Additionally, we observe that training effectiveness tends to increase with training duration, and is highest when training is fully or partly financed by the employer. These patterns highlight the importance of both training content and employer investment in maximizing training gains.

While our study provides insights into the role of job training in reducing automation risk, it also raises important questions that warrant further investigation. A key limitation of our analysis is the lack of detailed information on the specific content of training programs. Understanding which type of training is more effective is crucial for tailoring more effective policy interventions. Future research should aim to collect data on the types of training provided, particularly focusing on how these programs align with the evolving demands of the labor market. Finally, while our study emphasizes the importance of reducing the risk of automation, future work should consider how training interacts with other forms of human capital investment, such as formal education and lifelong learning, to provide a more comprehensive understanding of how workers can adapt to and thrive in an increasingly automated world.

References

Acemoglu, D. and D. Autor (2011). Skills, Tasks and Technologies: Implications for Employment and Earnings. *Handbook of Labor Economics 4B*, 1043–1171.

Acemoglu, D. and J.-S. Pischke (1998). Why Do Firms Train? Theory and Evidence. *Quarterly Journal of Economics 113*(1), 79–119.

Acemoglu, D. and P. Restrepo (2022). Demographics and Automation. *Review of Economic Studies 89*(1), 1–44.

Adachi, D., D. Kawaguchi, and Y. U. Saito (2024). Robots and Employment: Evidence from Japan, 1978–2017. *Journal of Labor Economics 42*(2), 591–634.

Adhvaryu, A., N. Kala, and A. Nyshadham (2023). Returns to On-the-Job Soft Skills Training. *Journal of Political Economy 131*(8), 2165–2208.

Akerman, A., I. Gaarder, and M. Mogstad (2015). The Skill Complementarity of Broadband Internet. *Quarterly Journal of Economics 130*(4), 1781–1824.

Arntz, M., T. Gregory, and U. Zierahn (2016). The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis. *OECD Social, Employment and Migration Working Papers 189*.

Arntz, M., T. Gregory, and U. Zierahn (2017). Revisiting the Risk of Automation. *Economics Letters 159*, 157–160.

Arulampalam, W. and A. Booth (1997). Who Gets Over the Training Hurdle? A Study of the Training Experiences of Young Men and Women in Britain. *Journal of Population Economics 10*(2), 197–217.

Atalay, E., P. Phongthiengtham, S. Sotelo, and D. Tannenbaum (2020). The Evolution of Work in the United States. *American Economic Journal: Applied Economics 12*(2), 1–34.

Autor, D. H., F. Levy, and R. J. Murnane (2003). The Skill Content of Recent Technological Change: An Empirical Exploration. *Quarterly Journal of Economics 118*(4), 1279–1333.

Bachmann, R., G. Demir, C. Green, and A. Uhlendorff (2022). The Role of Within-Occupation Task Changes in Wage Development. *IZA Discussion Papers* (No. 15647).

Bartel, A. P. (1995). Training, Wage Growth, and Job Performance: Evidence from a Company Database. *Journal of Labor Economics 13*(3), 401–425.

Bassanini, A., A. Booth, G. Brunello, M. De Paola, and E. Leuven (2007). Workplace Training in Europe. In Brunello, G., Garibaldi, P., and Wasmer, E. (Eds.), *Education and Training in Europe*, pp. 388. Oxford, New York: Oxford University Press.

Battisti, M., C. Dustmann, and U. Schoenberg (2023). Technological and Organizational Change and the Careers of Workers. *Journal of the European Economic Association 21*(4), 1551–1594.

Black, S. E. and A. Spitz-Oener (2010). Explaining Women's Success: Technological Change and the Skill Content of Women's Work. *The Review of Economics and Statistics 92*(1), 187–194.

Blundell, R., L. Dearden, C. Meghir, and B. Sianesi (1999). Human Capital Investment: The Returns from Education and Training to the Individual, the Firm and the Economy. *Fiscal Studies 20*(1), 1–23.

Boehm, M. J., H.-M. v. Gaudecker, and F. Schran (2024). Occupation Growth, Skill Prices, and Wage Inequality. *Journal of Labor Economics 42*(1), 201–243.

Braxton, J. C. and B. Taska (2023). Technological Change and the Consequences of Job Loss. *American Economic Review 113*(2), 279–316.

Brunello, G., D. Rueckert, C. T. Weiss, and P. Wruuck (2023). Advanced Digital Technologies and Investment in Employee Training: Complements or Substitutes? *IZA Discussion Papers* (No. 15936).

Brzeski, C. and I. Burk (2015). Die Roboter kommen. *Folgen der Automatisierung fuer den deutschen Arbeitsmarkt. INGDiBa Economic Research 30*.

Cameron, A. and P. Trivedi (2005). *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press.

Card, D., J. Kluve, and A. Weber (2010). Active Labour Market Policy Evaluations: A Meta-Analysis. *The Economic Journal 120*(548), 452–477.

Card, D., J. Kluve, and A. Weber (2018). What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations. *Journal of the European Economic Association 16*(3), 894–931.

Caunedo, J., D. Jaume, and E. Keller (2023). Occupational Exposure to Capital-Embodied Technical Change. *American Economic Review 113*(6), 1642–1685.

Chen, Y. (2021). Team-Specific Human Capital and Team Performance: Evidence from Doctors. *American Economic Review 111*(12), 3923–3962.

Cortes, G. M. (2016). Where Have the Middle-Wage Workers Gone? A Study of Polarization Using Panel Data. *Journal of Labor Economics 34*(1), 63–105.

Cunningham, E. (2018). Great Recession, Great Recovery? Trends from the Current Population Survey: Monthly Labor Review: U.S. Bureau of Labor Statistics.

Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press.

De Grip, A. and J. Sauermann (2012). The Effects of Training on Own and Co-Worker Productivity: Evidence from a Field Experiment. *The Economic Journal 122*(560), 376–399.

De Grip, A. and J. Sauermann (2013). The Effect of Training on Productivity: The Transfer of On-The-Job Training from the Perspective of Economics. *Educational Research Review 8*, 28–36.

Dearden, L., H. Reed, and J. Van Reenen (2006). The Impact of Training on Productivity and Wages: Evidence from British Panel Data. *Oxford Bulletin of Economics and Statistics 68*(4), 397–421.

Dehejia, R. H. and S. Wahba (2002). Propensity Score-Matching Methods for Nonexperimental Causal Studies. *The Review of Economics and Statistics 84*(1), 151–161.

Deming, D. J. (2017). The Growing Importance of Social Skills in the Labor Market. *Quarterly Journal of Economics 132*(4), 1593–1640.

Deming, D. J. and K. Noray (2020). Earnings Dynamics, Changing Job Skills, and STEM Careers. *Quarterly Journal of Economics 135*(4), 1965–2005.

Dotsey, M., S. Fujita, and L. Rudanko (2017). Where Is Everybody? The Shrinking Labor Force Participation Rate. *Economic Insights 2*(4), 17–24.

Falck, O., A. Heimisch-Roecker, and S. Wiederhold (2021). Returns to ICT Skills. *Research Policy 50*(7), Article 104064.

Falck, O., V. Lindlacher, and S. Wiederhold (2022). Elderly Left Behind? How Older Workers Can Participate in the Modern Labor Market. *EconPol Forum 23*(05), 16–19.

Feigenbaum, J. and D. P. Gross (2024). Answering the Call of Automation: How the Labor Market Adjusted to Mechanizing Telephone Operation. *Quarterly Journal of Economics 139*(3), 1879–1939.

Fialho, P., G. Quintini, and M. Vandeweyer (2019). Returns to Different Forms of Job Related Training: Factoring in Informal Learning. Technical report, OECD, Paris.

Fouarge, D. and T. Schils (2009). The Effect of Early Retirement Incentives on the Training Participation of Older Workers. *LABOUR 23*(s1), 85–109.

Frazis, H. and M. A. Loewenstein (2005). Reexamining the Returns to Training Functional Form, Magnitude, and Interpretation. *Journal of Human Resources XL*(2), 453–476.

Frey, C. B. and M. A. Osborne (2017). The Future of Employment: How Susceptible Are Jobs to Computerisation? *Technological Forecasting and Social Change 114*, 254–280.

Goerlitz, K. (2011). Continuous Training and Wages: An Empirical Analysis Using a Comparison-Group Approach. *Economics of Education Review 30*(4), 691–701.

Goerlitz, K. and M. Tamm (2016). The Returns to Voucher-Financed Training on Wages, Employment and Job Tasks. *Economics of Education Review 52*, 51–62.

Goldin, C. and L. Katz (1998). The Origins of Technology-Skill Complementarity. *Quarterly Journal of Economics 113*(3), 693–732.

Goos, M., A. Manning, and A. Salomons (2014). Explaining Job Polarization: Routine-Biased Technological Change and Offshoring. *American Economic Review 104*(8), 2509–2526.

Goux, D. and E. Maurin (2000). Returns to Firm-Provided Training: Evidence from French Worker–Firm Matched Data. *Labour Economics 7*(1), 1–19.

Greenhalgh, C. and M. Stewart (1987). The Effects and Determinants of Training. *Oxford Bulletin of Economics and Statistics 49*(2), 171–190.

Grund, C. and J. Martin (2012). Determinants of Further Training - Evidence for Germany. *The International Journal of Human Resource Management 23*(17), 3536–3558.

Hainmueller, J. (2012). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis 20*(1), 25–46.

Hainmueller, J. and Y. Xu (2013). ebalance: A Stata Package for Entropy Balancing. *Journal of Statistical Software 54*(7), 1–18.

Hanushek, E. A., G. Schwerdt, S. Wiederhold, and L. Woessmann (2015). Returns to Skills around the World: Evidence from PIAAC. *European Economic Review 73*, 103–130.

Hanushek, E. A., G. Schwerdt, S. Wiederhold, and L. Woessmann (2017). Coping with Change: International Differences in the Returns to Skills. *Economics Letters 153*(C), 15–19.

Hanushek, E. A. and L. Woessmann (2008). The Role of Cognitive Skills in Economic Development. *Journal of Economic Literature 46*(3), 607–668.

Heckman, J., H. Ichimura, J. Smith, and P. Todd (1998). Characterizing Selection Bias Using Experimental Data. *Econometrica 66*(5), 1017–1098.

Heckman, J. J., H. Ichimura, and P. E. Todd (1997). Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *The Review of Economic Studies 64*(4), 605–654.

Heß, P., S. Janssen, and U. Leber (2023). The Effect of Automation Technology on Workers' Training Participation. *Economics of Education Review 96*, 102438.

Hidalgo, D., H. Oosterbeek, and D. Webbink (2014). The Impact of Training Vouchers on Low-Skilled Workers. *Labour Economics 31*, 117–128.

Hirvonen, J., A. Stenhammar, and J. Tuhkuri (2022). New Evidence on the Effect of Technology on Employment and Skill Demand. *ETLA Working Papers No. 93*.

Hujer, R., S. L. Thomsen, and C. Zeiss (2006). The Effects of Vocational Training Programmes on the Duration of Unemployment in Eastern Germany. *Allgemeines Statistisches Archiv 90*(2), 299–321.

Kleven, H., C. Landais, and G. Leite-Mariante (2024). The Child Penalty Atlas. *The Review of Economic Studies*, rdae104.

Kleven, H., C. Landais, J. Posch, A. Steinhauer, and J. Zweimueller (2019). Child Penalties across Countries: Evidence and Explanations. *AEA Papers and Proceedings 109*, 122–126.

Kluve, J. (2010). The Effectiveness of European Active Labor Market Programs. *Labour Economics 17*(6), 904–918.

Konings, J. and S. Vanormelingen (2015). The Impact of Training on Productivity and Wages: Firm-Level Evidence. *The Review of Economics and Statistics 97*(2), 485–497.

Krueger, A. B. (1993). How Computers Have Changed the Wage Structure: Evidence from Microdata, 1984-1989. *Quarterly Journal of Economics 108*(1), 33–60.

Krusell, P., L. E. Ohanian, J.-V. Rios-Rull, and G. L. Violante (2000). Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis. *Econometrica 68*(5), 1029–1053.

LaLonde, R. J. (1986). Evaluating the Econometric Evaluations of Training Programs with Experimental Data. *American Economic Review 76*(4), 604–620.

Langer, C. and S. Wiederhold (2023). The Value of Early-Career Skills. *CESifo Working Paper No. 10288*.

Lechner, M. (1999). Earnings and Employment Effects of Continuous Off-the-Job Training in East Germany After Unification. *Journal of Business & Economic Statistics 17*(1), 74–90.

Lergetporer, P., K. Wedel, and K. Werner (2023). Automation Potential of Occupations and Workers' Labor Market Expectations. *IZA Discussion Papers No. 16687*.

Leuven, E. (2005). The Economics of Private Sector Training: A Survey of the Literature. *Journal of Economic Surveys 19*(1), 91–111.

Leuven, E. and H. Oosterbeek (1999). The Demand and Supply of Work-Related Training: Evidence from Four Countries. In *Research in Labor Economics*, Volume 18, pp. 303–330. Emerald Group Publishing Limited.

Leuven, E. and H. Oosterbeek (2008). An Alternative Approach to Estimate the Wage Returns to Private-Sector Training. *Journal of Applied Econometrics 23*(4), 423–434.

Lynch, L. M. (1992). Private-Sector Training and the Earnings of Young Workers. *American Economic Review 82*(1), 299–312.

Lynch, L. M. and S. E. Black (1998). Beyond the Incidence of Employer-Provided Training. *ILR Review 52*(1), 64–81.

Ma, X., A. Nakab, and D. Vidart (2024). Human Capital Investment and Development: The Role of On-the-Job Training. *Journal of Political Economy Macroeconomics 2*(1), 107–148.

Martin, S., C. Lechner, C. Kleinert, and B. Rammstedt (2021). Literacy Skills Predict Probability of Refusal in Follow-Up Wave: Evidence from Two Longitudinal Assessment Surveys. *International Journal of Social Research Methodology 24*(4), 413–428.

McCall, B., J. Smith, and C. Wunsch (2016). Government-Sponsored Vocational Education for Adults. In E. A. Hanushek, S. Machin, and L. Woessmann (Eds.), *Handbook of the Economics of Education*, Volume 5, pp. 479–652. Elsevier.

Meager, R. (2019). Understanding the Average Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of Seven Randomized Experiments. *American Economic Journal: Applied Economics 11*(1), 57–91.

Mokyr, J., C. Vickers, and N. L. Ziebarth (2015). The History of Technological Anxiety and the Future of Economic Growth: Is This Time Di erent? *Journal of Economic Perspectives 29*(3), 31–50.

Nedelkoska, L. and G. Quintini (2018). Automation, Skills Use and Training. OECD Social, Employment and Migration Working Papers 202, OECD, Paris.

OECD (2013). *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*. Paris: Organisation for Economic Co-operation and Development.

OECD (2016). *Skills Matter: Further Results from the Survey of Adult Skills*. Paris: Organisation for Economic Co-operation and Development.

OECD (2017). *OECD Skills Outlook 2017: Skills and Global Value Chains*. Paris: Organisation for Economic Co-operation and Development.

OECD (2019). *Technical Report of the Survey of Adult Skills (PIAAC) (3rd Edition).* Paris: Organisation for Economic Co-operation and Development.

Oosterbeek, H. (1996). A Decomposition of Training Probabilities. *Applied Economics 28*(7), 799–805. Number: 7.

Oosterbeek, H. (1998). Unravelling Supply and Demand Factors in Work-Related Training. *Oxford Economic Papers 50*(2), 266–283.

Oster, E. (2019). Unobservable Selection and Coe cient Stability: Theory and Evidence. *Journal of Business & Economic Statistics 37*(2), 187–204.

Pajarinen, M. and P. Rouvinen (2014). Computerization Threatens One Third of Finnish Employment. ETLA Brief 22, The Research Institute of the Finnish Economy.

Parent, D. (1999). Wages and Mobility: The Impact of Employer-Provided Training. *Journal of Labor Economics 17*(2), 298–317.

Pischke, J.-S. (2001). Continuous Training in Germany. *Journal of Population Economics 14*(3), 523–548.

Rammstedt, B., S. Martin, A. Zabal, C. Carstensen, and J. Schupp (2017). The PIAAC Longitudinal Study in Germany: Rationale and Design. *Large-scale Assessments in Education 5*(1), 4.

Ruhose, J., S. L. Thomsen, and I. Weilage (2019). The Benefits of Adult Learning: Work-Related Training, Social Capital, and Earnings. *Economics of Education Review 72*, 166–186.

Salas-Velasco, M. (2009). Beyond Lectures and Tutorials: Formal On-the-Job Training Received by Young European University Graduates. *Research in Economics 63*(3), 200–211.

Schmidpeter, B. and R. Winter-Ebmer (2021). Automation, Unemployment, and the Role of Labor Market Training. *European Economic Review 137*, 103808.

Schwerdt, G., D. Messer, L. Woessmann, and S. C. Wolter (2012). The Impact of an Adult Education Voucher Program: Evidence from a Randomized Field Experiment. *Journal of Public Economics 96*(7), 569–583.

Smith, J. A. and P. E. Todd (2005). Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators? *Journal of Econometrics 125*(1), 305–353.

Zabal, A., S. Martin, and B. Rammstedt (2017). *PIAAC-L Data Collection 2015: Technical Report*, Volume 2017/29 of *GESIS Papers*. Köln: GESIS - Leibniz-Institut für Sozialwissenschaften.

Online Appendix

for

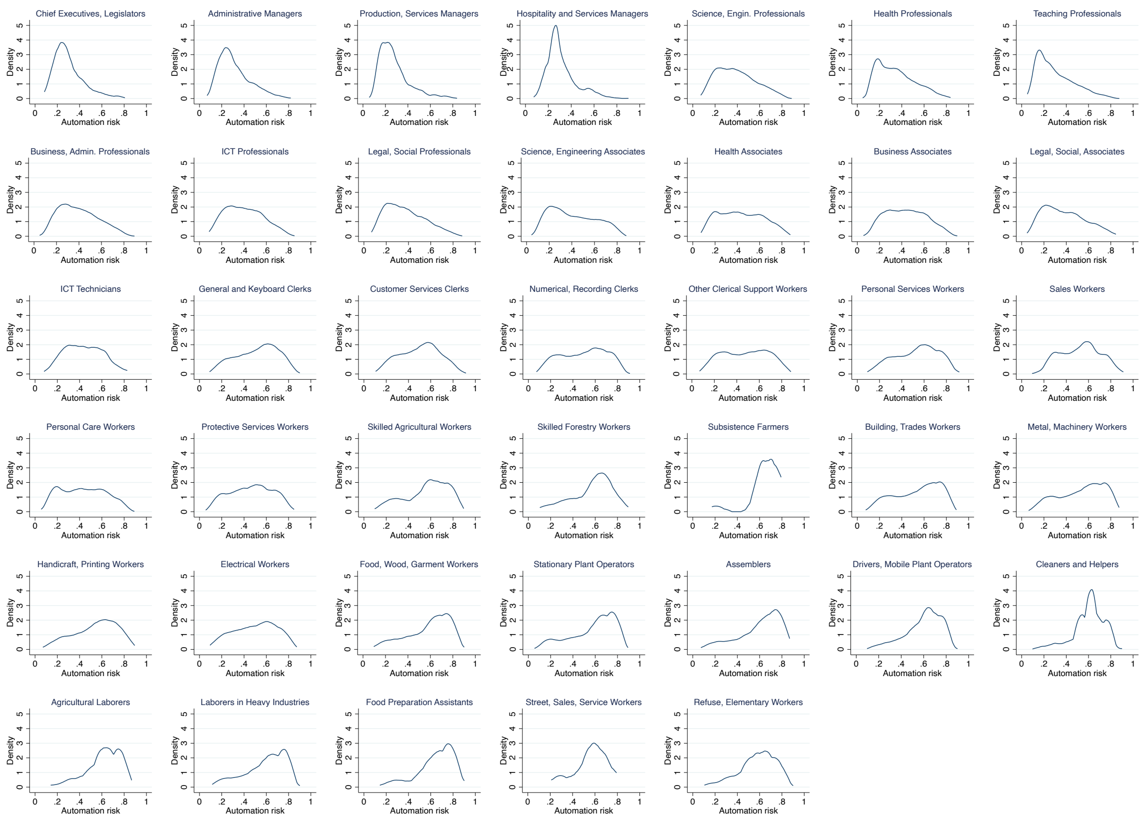
"Training, Automation, and Wages: International Worker-Level Evidence"

by

Oliver Falck, Yuchen Guo, Christina Langer, Valentin Lindlacher & Simon Wiederhold

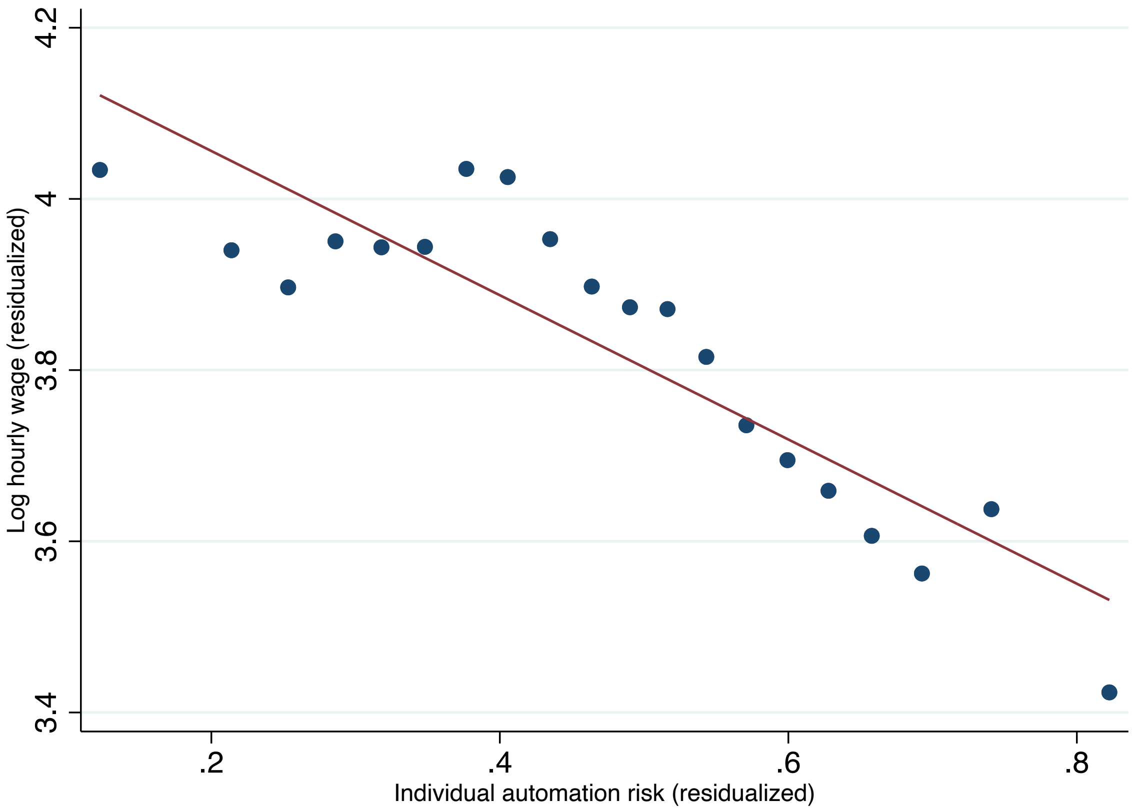
Appendix A. Additional Figures

Figure A.1: Densities of Automation Risk by Occupation



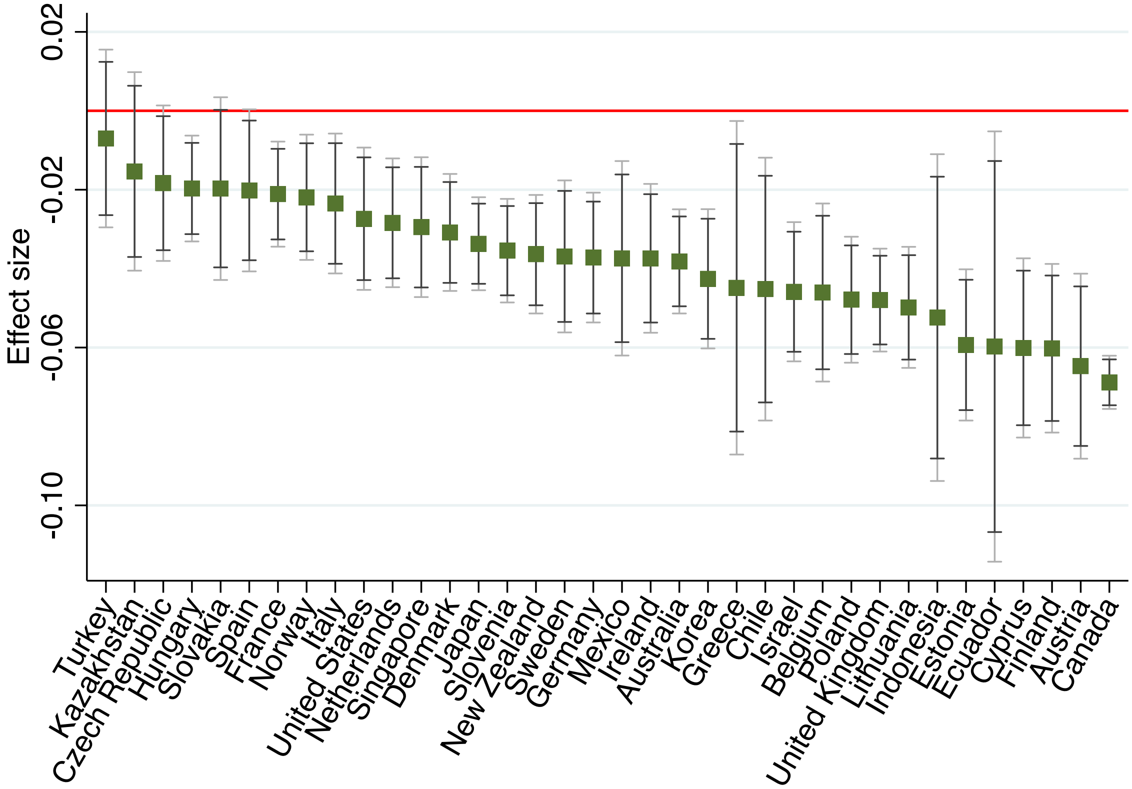
Notes: The figure shows the density of the individual-level automation risk by occupation at the two-digit ISCO occupation level across all countries in our sample. The automation risk ranges from 0 (indicating a low probability that a worker is fully automated) to 1 (indicating a high probability that a worker is fully automated); automation risk is predicted using items on task use at work from PIAAC.
Data source: PIAAC.

Figure A.2: Automation Risk and Wages



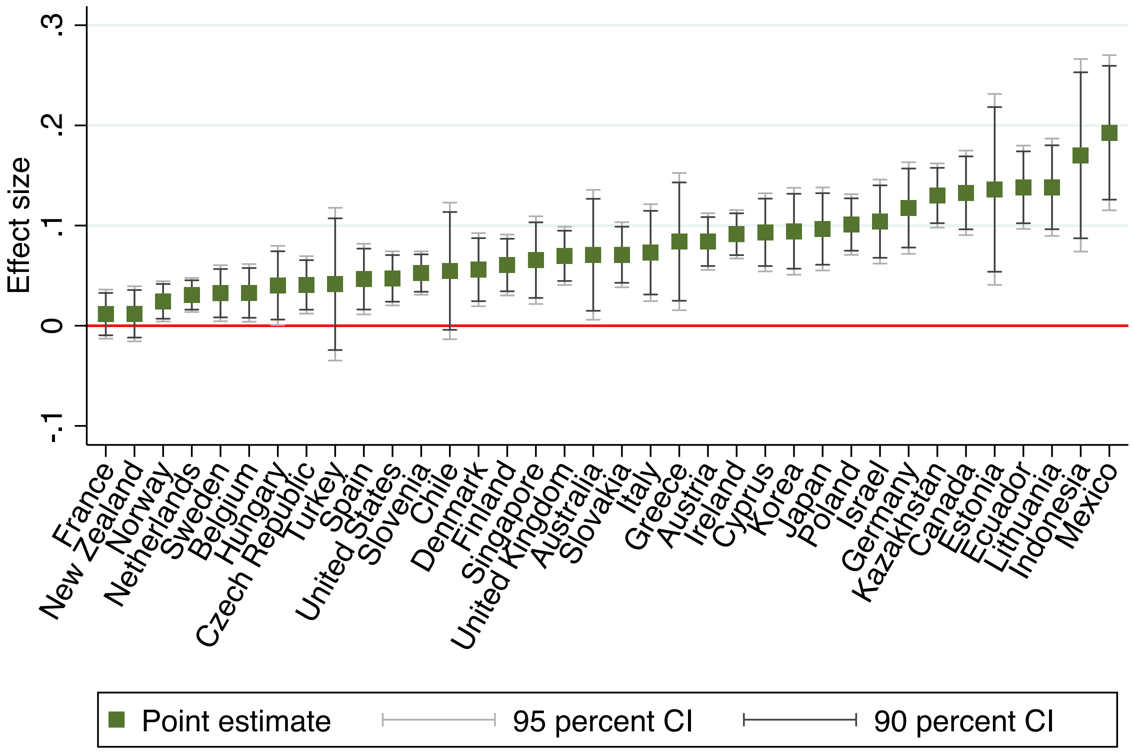
Notes: The figure displays a binned scatter plot showing the relationship between automation risk and log hourly wages. The automation risk is measured at the individual level, ranging from 0 (indicating a low probability that a worker is fully automated) to 1 (indicating a high probability that a worker is fully automated); automation risk is predicted using items on task use at work from PIAAC. To construct the figure, we divided the average automation risk into 20 ranked equal-sized groups and plotted the mean of the log hourly wages against the mean of average automation risk in each bin. The figure shows the residualized relationship after accounting for country, industry (two-digit ISIC), and occupation (two-digit ISCO) fixed effects. Best-fit line shown in red.
Data source: PIAAC.

Figure A.3: Training and Automation Risk by Country



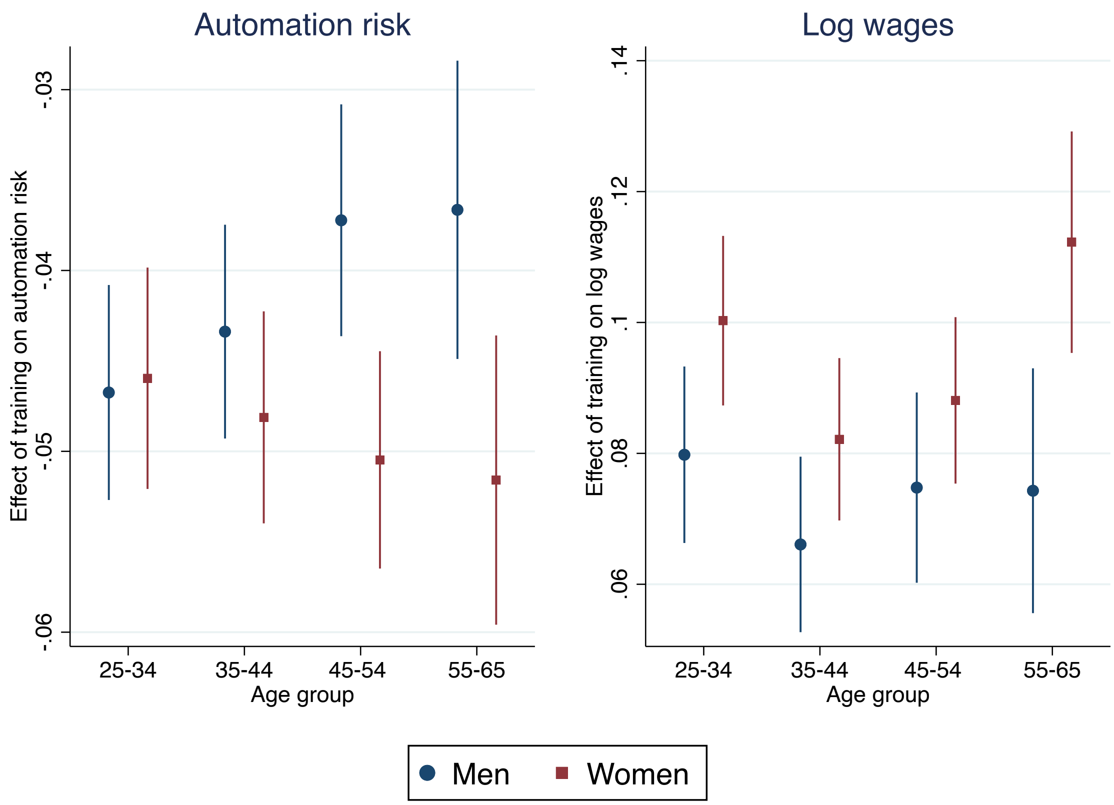
Notes: The figure shows the effect of training on automation risk separately for each PIAAC country. Least squares estimation with weights from entropy balancing. Dependent variable: individual-level automation risk, ranging from 0 (indicating a low probability that a worker is fully automated) to 1 (indicating a high probability that a worker is fully automated); automation risk is predicted using items on task use at work from PIAAC. Sample: employees aged 25–65 years with information on automation risk and log hourly wages. Job Training: binary variable indicating whether the respondent participated in on-the-job training or job-related training in the 12 months prior to the survey. Included controls: numeracy skills, years of schooling, age group in four categories (25–34, 35–44, 45–54, 55–65), gender, migration status in three categories (first-generation migrant, second-generation migrant, native), parental education in three categories (neither parent has attained upper secondary, at least one parent has attained secondary and post-secondary/non-tertiary, at least one parent has attained tertiary), an indicator whether the respondent has children, age group of the oldest child in four categories (0–2, 3–5, 6–12, 13+), an indicator for participation in non-job-related training (i.e., open/distance education, seminars/workshops, or private lessons), an indicator for full-time employment, and firm size measured by the number of employees in five categories (1–10, 11–50, 51–250, 251–1000, 1000+). Industry fixed effects at the two-digit ISIC level and occupation fixed effects at the two-digit ISCO level. All control variables were used for the entropy balancing. 95 percent confidence intervals are based on standard errors clustered at the occupation level.
Data source: PIAAC.

Figure A.4: Training and Wages by Country



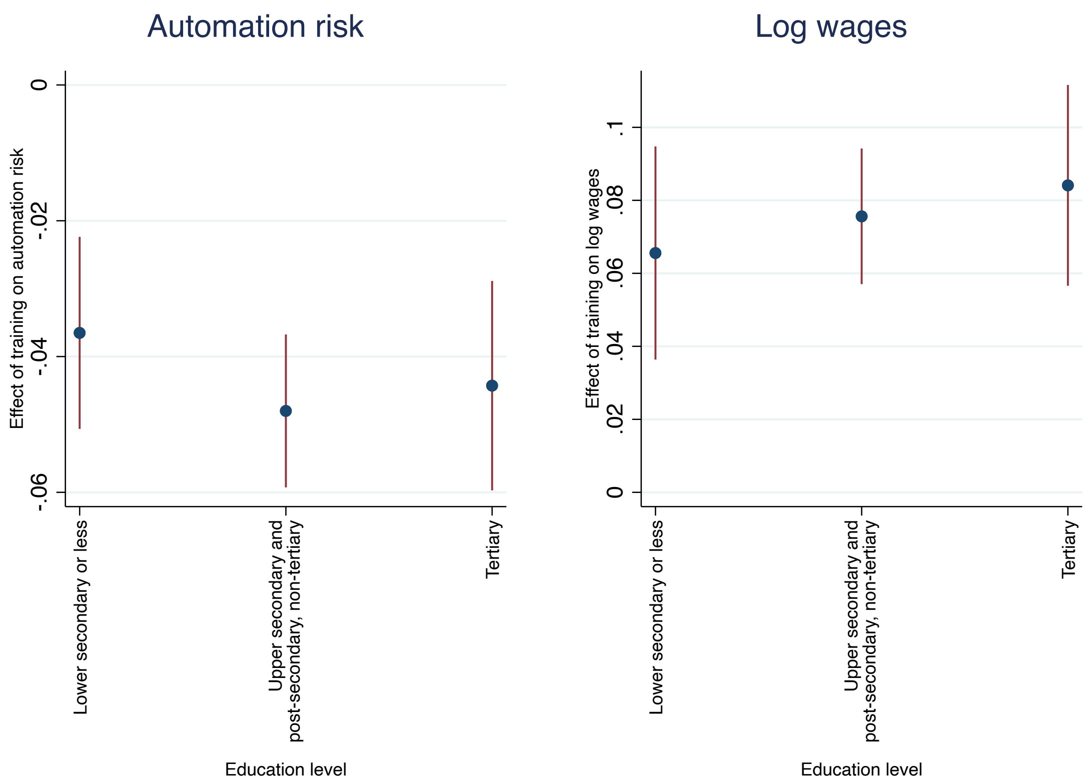
Notes: The figure shows the effect of training on wages separately for each PIAAC country. Least squares estimation with weights from entropy balancing. Dependent variable: log hourly wages. Sample: employees aged 25–65 years with information on automation risk and log hourly wages. Job Training: binary variable indicating whether the respondent participated in on-the-job training or job-related training in the 12 months prior to the survey. Included controls: numeracy skills, years of schooling, age group in four categories (25–34, 35–44, 45–54, 55–65), gender, migration status in three categories (first-generation migrant, second-generation migrant, native), parental education in three categories (neither parent has attained upper secondary, at least one parent has attained secondary and post-secondary/non-tertiary, at least one parent has attained tertiary), an indicator whether the respondent has children, age group of the oldest child in four categories (0–2, 3–5, 6–12, 13+), an indicator for participation in non-job-related training (i.e., open/distance education, seminars/workshops, or private lessons), an indicator for full-time employment, and firm size measured by the number of employees in five categories (1–10, 11–50, 51–250, 251–1000, 1000+). Industry fixed effects at the two-digit ISIC level and occupation fixed effects at the two-digit ISCO level. All control variables were used for the entropy balancing. 95 percent confidence intervals are based on standard errors clustered at the occupation level.
Data source: PIAAC.

Figure A.5: Heterogeneity by Age and Gender



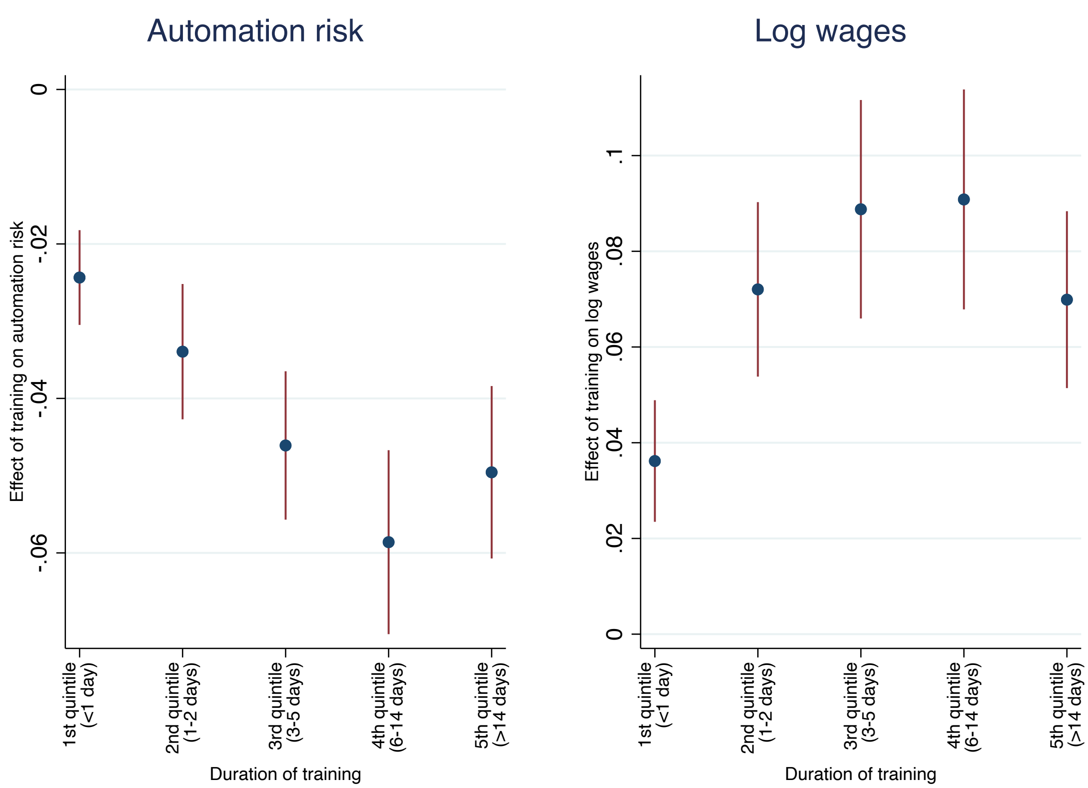
Notes: Least squares estimation with weights from entropy balancing in both figures. Separate estimations by gender. Dependent variable in left figure: individual-level automation risk, ranging from 0 (indicating a low probability that a worker is fully automated) to 1 (indicating a high probability that a worker is fully automated); automation risk is predicted using items on task use at work from PIAAC. Dependent variable in right figure: log hourly wages. Sample: employees aged 25–65 years with information on automation risk and log hourly wages. Job Training: binary variable indicating whether the respondent participated in on-the-job training or job-related training in the 12 months prior to the survey. Included controls are numeracy skills, years of schooling, age group in four categories (25–34, 35–44, 45–54, 55–65), gender, migration status in three categories (first-generation migrant, second-generation migrant, native), parental education in three categories (neither parent has attained upper secondary, at least one parent has attained secondary and post-secondary/non-tertiary, at least one parent has attained tertiary), an indicator whether the respondent has children, age group of the oldest child in four categories (0–2, 3–5, 6–12, 13+), an indicator for participation in non-job-related training (i.e., open/distance education, seminars/workshops, or private lessons), an indicator for full-time employment, and firm size measured by the number of employees in five categories (1–10, 11–50, 51–250, 251–1000, 1000+). Industry fixed effects at the two-digit ISIC level and occupation fixed effects at the two-digit ISCO level. All regressions also control for country fixed effects. All control variables were used for the entropy balancing. R^2 refers to within-country R^2. 95 percent confidence intervals are based on standard errors two-way clustered at the country and occupation level. □□□$p < 0.01$,□□ $p < 0.05$,□ $p < 0.1$.
Data source: PIAAC.

Figure A.6: Heterogeneity by Education Level



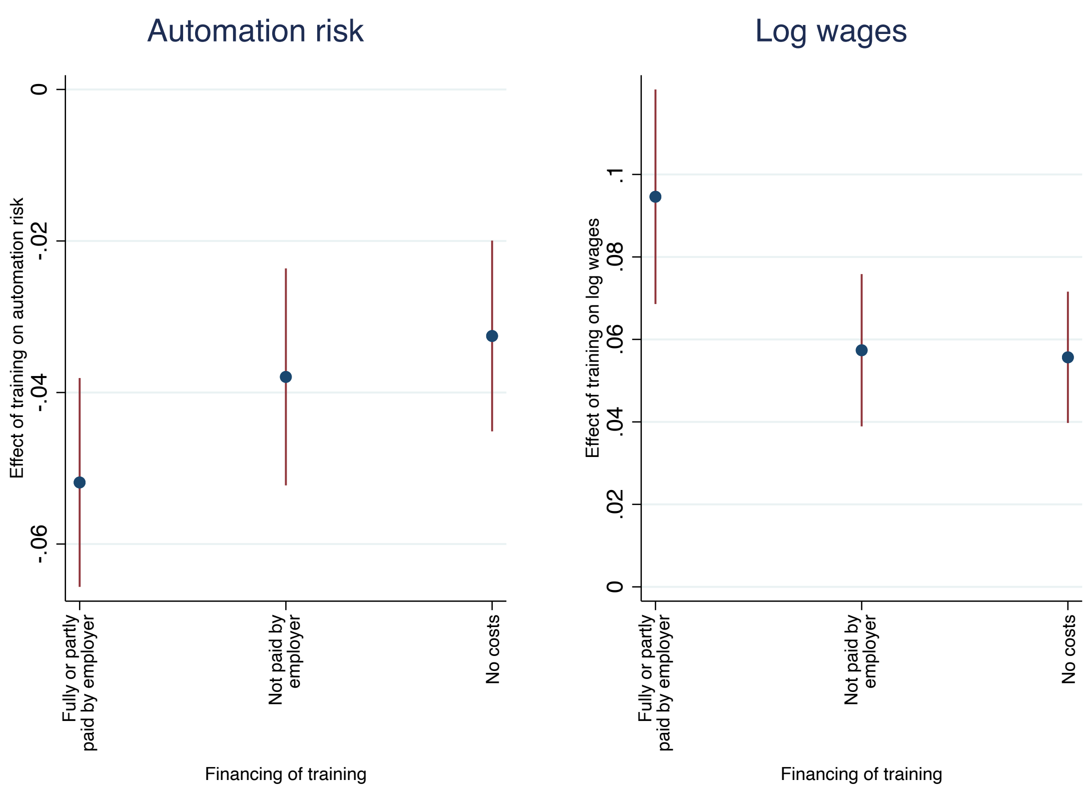
Notes: Least squares estimation with weights from entropy balancing in both figures. Separate estimations by highest educational degree obtained. Dependent variable in left figure: individual-level automation risk, ranging from 0 (indicating a low probability that a worker is fully automated) to 1 (indicating a high probability that a worker is fully automated); automation risk is predicted using items on task use at work from PIAAC. Dependent variable in right figure: log hourly wages. Sample: employees aged 25–65 years with information on automation risk and log hourly wages. Job Training: binary variable indicating whether the respondent participated in on-the-job training or job-related training in the 12 months prior to the survey. Included controls are numeracy skills, years of schooling, age group in four categories (25–34, 35–44, 45–54, 55–65), gender, migration status in three categories (first-generation migrant, second-generation migrant, native), parental education in three categories (neither parent has attained upper secondary, at least one parent has attained secondary and post-secondary/non-tertiary, at least one parent has attained tertiary), an indicator whether the respondent has children, age group of the oldest child in four categories (0–2, 3–5, 6–12, 13+), an indicator for participation in non-job-related training (i.e., open/distance education, seminars/workshops, or private lessons), an indicator for full-time employment, and firm size measured by the number of employees in five categories (1–10, 11–50, 51–250, 251–1000, 1000+). Industry fixed effects at the two-digit ISIC level and occupation fixed effects at the two-digit ISCO level. All regressions also control for country fixed effects. All control variables were used for the entropy balancing. 95 percent confidence intervals are based on standard errors two-way clustered at the country and occupation level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Data source: PIAAC.

Figure A.7: Heterogeneity by Duration of Training



Notes: Least squares estimation with weights from entropy balancing in both figures. Separate estimations by duration of training. Dependent variable in left figure: individual-level automation risk, ranging from 0 (indicating a low probability that a worker is fully automated) to 1 (indicating a high probability that a worker is fully automated); automation risk is predicted using items on task use at work from PIAAC. Dependent variable in right figure: log hourly wages. Sample: employees aged 25–65 years with information on automation risk and log hourly wages. Job Training: binary variable indicating whether the respondent participated in on-the-job training or job-related training in the 12 months prior to the survey. Included controls are numeracy skills, years of schooling, age group in four categories (25–34, 35–44, 45–54, 55–65), gender, migration status in three categories (first-generation migrant, second-generation migrant, native), parental education in three categories (neither parent has attained upper secondary, at least one parent has attained secondary and post-secondary/non-tertiary, at least one parent has attained tertiary), an indicator whether the respondent has children, age group of the oldest child in four categories (0–2, 3–5, 6–12, 13+), an indicator for participation in non-job-related training (i.e., open/distance education, seminars/workshops, or private lessons), an indicator for full-time employment, and firm size measured by the number of employees in five categories (1–10, 11–50, 51–250, 251–1000, 1000+). Industry fixed effects at the two-digit ISIC level and occupation fixed effects at the two-digit ISCO level. All regressions also control for country fixed effects. All control variables were used for the entropy balancing. 95 percent confidence intervals are based on standard errors two-way clustered at the country and occupation level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Data source: PIAAC.

Figure A.8: Heterogeneity by Financing of Training



Notes: Least squares estimation with weights from entropy balancing in both figures. Separate estimations by how the training was financed. Dependent variable in left figure: individual-level automation risk, ranging from 0 (indicating a low probability that a worker is fully automated) to 1 (indicating a high probability that a worker is fully automated); automation risk is predicted using items on task use at work from PIAAC. Dependent variable in right figure: log hourly wages. Sample: employees aged 25–65 years with information on automation risk and log hourly wages. Job Training: binary variable indicating whether the respondent participated in on-the-job training or job-related training in the 12 months prior to the survey. Included controls are numeracy skills, years of schooling, age group in four categories (25–34, 35–44, 45–54, 55–65), gender, migration status in three categories (first-generation migrant, second-generation migrant, native), parental education in three categories (neither parent has attained upper secondary, at least one parent has attained secondary and post-secondary/non-tertiary, at least one parent has attained tertiary), an indicator whether the respondent has children, age group of the oldest child in four categories (0–2, 3–5, 6–12, 13+), an indicator for participation in non-job-related training (i.e., open/distance education, seminars/workshops, or private lessons), an indicator for full-time employment, and firm size measured by the number of employees in five categories (1–10, 11–50, 51–250, 251–1000, 1000+). Industry fixed effects at the two-digit ISIC level and occupation fixed effects at the two-digit ISCO level. All regressions also control for country fixed effects. All control variables were used for the entropy balancing. 95 percent confidence intervals are based on standard errors two-way clustered at the country and occupation level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Data source: PIAAC.

Appendix B. Additional Tables

Table B.1: Job Training Characteristics By Worker Demographics

Subgroup	**Share Participating in Training**	**Duration of Training**		**Share by Financing of Training**		
	Mean	Median	25th–75th percentile	Fully or partly paid by employer	Not paid by employer	No costs
All workers	54.63	4 days	1–12 days	68.86	16.65	14.49
Age group						
25–34	56.16	5 days	1–14 days	64.62	19.44	15.94
35–44	56.78	4 days	1–12 days	69.70	16.23	14.07
45–54	54.55	4 days	1–10 days	71.19	15.27	13.53
55–65	48.81	3 days	1–8 days	70.96	14.73	14.31
Gender						
Male	54.02	4 days	1–12 days	72.39	14.26	13.35
Female	55.21	4 days	1–11 days	65.64	18.82	15.54
Education level						
Lower secondary or less	31.70	2 days	4 hours–7 days	67.13	15.54	17.33
Upper secondary and post-secondary, non-tertiary	46.83	3 days	1–8 days	70.39	15.13	14.48
Tertiary	69.31	5 days	2–15 days	68.16	17.76	14.07
Firm size						
1–10 employees	38.60	3 days	1–10 days	62.21	22.41	15.38
11–50 employees	53.92	3 days	1–10 days	68.43	17.20	14.36
51–250 employees	61.22	4 days	1–12 days	70.49	14.99	14.52
251–1000 employees	65.69	5 days	1–14 days	72.44	13.26	14.30
> 1000 employees	71.48	5 days	2–14 days	73.00	13.69	13.30

Notes: This table presents descriptive statistics on training participation, duration, and financing across worker demographics, with data pooled from all countries.
Data source: PIAAC.

Table B.2: Automation Risk: Description of Included Tasks

PIAAC Variable	**Variable Description**
Simple problems	The next question is about "problem solving" tasks you do in your current job. Think of "problem solving" as what happens when you are faced with a new or difficult situation which requires you to think for a while about what to do next. How often are you usually faced by relatively simple problems that take no more than 5 minutes to find a good solution?
Complex problems	And how often are you usually confronted with more complex problems that take at least 30 minutes to find a good solution? The 30 minutes only refers to the time needed to THINK of a solution, not the time needed to carry it out.
	How often does your current job usually involve ...
Communicate	sharing work-related information with co-workers?
Teach	instructing, training or teaching people, individually or in groups?
Sell	selling a product or selling a service?
Advise	providing advice?
Plan work for others	planning the activities of others?
Influence others	working to persuade or influence people?
Negotiate	negotiating with people either inside or outside your firm or organization?
Manual dexterity	using skill or accuracy with your hands or fingers?

Notes: The table provides an overview of the job tasks in PIAAC used to construct the automation risk measure.

Table B.3: Factor Loadings Automation Risk: International Analysis

Task	Logit Coefficient
Plan work of others	-0.308***
	(0.0234)
Influence others	-0.235***
	(0.0267)
Advise	-0.199***
	(0.0270)
Teach	-0.0691***
	(0.0255)
Complex problems	-0.0691**
	(0.0297)
Negotiate	-0.0463*
	(0.0255)
Simple problems	0.0573*
	(0.0309)
Manual dexterity	0.105***
	(0.0220)
Sell	0.160***
	(0.0206)
Sharing information	0.214***
	(0.0260)
Pseudo R^2	0.137
Observations	4,656

Notes: The table shows factor loadings from Nedelkoska and Quintini (2018), Table 4.3, derived from a logistic regression at the individual level. Dependent variable: Occupational automation risk from Frey and Osborne (2017), based on expert surveys assessing engineering bottlenecks (70 four-digit occupations). Occupations in which all tasks can be automated receive a value of 1, while those with only partially automatable tasks receive a value of 0. Independent variables: PIAAC task items corresponding to engineering bottlenecks identified by Frey and Osborne (2017). PIAAC assesses the frequency of task use, with responses recorded on a Likert scale ranging from 1 (never) to 5 (every day). Task items are ordered by degree of automatability. Coefficients are estimated on the Canadian PIAAC sample, which offers a detailed categorization of 4-digit ISCO occupations and provides the largest sample size among all participating PIAAC countries. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Data source: PIAAC Canada.

Table B.4: Job Tasks in German PIAAC Data

Label	Question
Problem Solving - Solve Difficult Problems	How often do you have to solve difficult problems as part of your job?
Routine - Deal with Unexpected Situations	How often do you have to react to situations at work that you could not foresee?
Interaction - Check Work of Others	How often do you have to check the quality of other people's work as part of your job?
Learning - Learn Something New	How often do you have to learn new things at work?
Autonomy - Personally Involved in Strategic Decisions	How often are you personally involved in important strategic decisions in your company, such as those regarding products, services, staffing, or finances?
Autonomy - Look for New Tasks for Yourself	How often do you have the opportunity to find new tasks for yourself at work?
Routine - Do New Things	How often do you have to do things at work that you have not done before?
Autonomy - Organize Your Own Work	How often can you organize your work yourself?
Routine - Need to Familiarize Oneself with Tasks	How often do you get tasks at work that require you to familiarize yourself with them first?
Autonomy - Quality of Your Work is Monitored	How often is the quality of your work monitored?
Routine - Task Diversity	How often do the tasks you have to complete at work change?
Autonomy - Can Determine Own Working Pace	How often can you set your own work pace?
Routine - Work Days Very Similar	How often is one workday very similar to another one for you?
Routine - Carry Out Short, Repetitive Tasks	How often do you have to carry out short, repetitive tasks in your daily work?
Routine - Get Detailed Specifications for Tasks	How often are you told how to do your job down to the last detail?

Notes: The table presents job tasks from the German PIAAC analysis. Task use items from the 2015 PIAAC survey, which capture the frequency of task use. Responses are recorded on a Likert scale ranging from 1 (very rarely or never) to 5 (always or very often).
Data source: PIAAC Germany (2015 wave).

Table B.5: Automation Risk in German PIAAC Data

Task	Logit Coefficient
Solve difficult problems	-0.3760***
	(0.0516)
Deal with unexpected situations	-0.2100***
	(0.0486)
Check work of others	-0.1786***
	(0.0333)
Learn something new	-0.1586***
	(0.0531)
Personally involved in strategic decisions	-0.1478***
	(0.0330)
Look for new tasks for yourself	-0.1212***
	(0.0400)
Do new things	-0.0813
	(0.0590)
Organize your own work	-0.0760*
	(0.0445)
Need to familiarize oneself with tasks	0.0104
	(0.0571)
Quality of your work is monitored	0.0466
	(0.0351)
Task diversity	0.0641
	(0.0488)
Can determine own working pace	0.0839**
	(0.0414)
Work days very similar	0.1022***
	(0.0390)
Carry out short, repetitive tasks	0.1371***
	(0.0499)
Get detailed specifications for tasks	0.1999***
	(0.0373)
Pseudo R^2	0.162
Observations	3,068

Notes: The table presents factor loadings from the German PIAAC analysis, derived from a logistic regression at the individual level. Dependent variable: Indicator of high automation risk in 2012, defined at the occupational level. To construct this variable, we first collapse our individual-level automation risk measure from the 2012 survey (see Section 3.1) to the 4-digit ISCO occupation level. We then generate a binary indicator, assigning a value of 1 if the aggregated automation risk is 0.5 or higher, and 0 if it is below 0.5. Independent variables: Task use items from the 2015 PIAAC survey, which capture the frequency of task use. Responses are recorded on a Likert scale ranging from 1 (very rarely or never) to 5 (always or very often). Task items are ordered by degree of automatability. Standard errors shown in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Data source: PIAAC Germany (2012 and 2015 waves).

Table B.6: Change in Automation Risk over Time in Germany

	Change in automation risk 2012–2015		
	(1)	(2)	(3)
Automation risk (2012)	-0.3830	-0.6226	-0.6704
	(0.0498)	(0.0300)	(0.0309)
Occupation FE (2012)		X	X
Further controls (2012)			X
R^2	0.11	0.30	0.43
Observations	1,713	1,713	1,713

Notes: Ordinary least squares estimation. Dependent variable: change in individual-level automation risk between 2012 and 2015. Automation risk ranges from 0 (indicating a low probability that a worker is fully automated) to 1 (indicating a high probability that a worker is fully automated); automation risk is predicted using items on task use at work from PIAAC. Sample: employees aged 25–65 years in Germany. Occupation fixed effects are measured at the two-digit ISCO level. Further controls refer to 2012 and include numeracy skills, years of schooling, age group in four categories (25–34, 35–44, 45–54, 55–65), gender, migration status in three categories (first-generation migrant, second-generation migrant, native), parental education in three categories (neither parent has attained upper secondary, at least one parent has attained secondary and post-secondary/non-tertiary, at least one parent has attained tertiary), an indicator whether the respondent has children, age group of the oldest child in four categories (0–2, 3–5, 6–12, 13+), an indicator for participation in non-job-related training (e.g., self-organized training or seminar participation), an indicator for full-time employment, firm size measured by number of employees in five categories (1–10, 11–50, 51–250, 251–1000, 1000+), and industry fixed effects (two-digit ISIC). Standard errors shown in parentheses are clustered at the occupation level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Data source: PIAAC Germany (2012 and 2015 waves).

Table B.7: Training and Automation Risk: Different Levels of Clustering Standard Errors

	Automation risk				
	(1)	(2)	(3)	(4)	(5)
Job training	-0.0467***	-0.0467***	-0.0467***	-0.0467***	-0.0467***
	(0.0018)	(0.0048)	(0.0068)	(0.0047)	(0.0069)
Numeracy skills	-0.0129***	-0.0129***	-0.0129***	-0.0129***	-0.0129***
	(0.0012)	(0.0024)	(0.0022)	(0.0018)	(0.0027)
Occupation FE	X	X	X	X	X
Further controls	X	X	X	X	X
Entropy balancing	X	X	X	X	X
R^2	0.20	0.20	0.20	0.20	0.20
Observations	91,470	91,470	91,470	91,470	91,470
Clustering	None (robust s.e.)	Country level	Occupation level	Country × occupation level	Two-way country and occupation level (baseline)

Notes: Least squares estimation with weights from entropy balancing. Dependent variable: individual-level automation risk, ranging from 0 (indicating a low probability that a worker is fully automated) to 1 (indicating a high probability that a worker is fully automated); automation risk is predicted using items on task use at work from PIAAC. Sample: employees aged 25–65 years with information on automation risk and log hourly wages. Job Training: binary variable indicating whether the respondent participated in on-the-job training or job-related training in the 12 months prior to the survey. Numeracy skills are standardized to unit standard deviation across countries. Further controls include: years of schooling, age group in four categories (25–34, 35–44, 45–54, 55–65), gender, migration status in three categories (first-generation migrant, second-generation migrant, native), parental education in three categories (neither parent has attained upper secondary, at least one parent has attained secondary and post-secondary/non-tertiary, at least one parent has attained tertiary), an indicator whether the respondent has children, age group of the oldest child in four categories (0–2, 3–5, 6–12, 13+), an indicator for participation in non-job-related training (i.e., open/distance education, seminars/workshops, or private lessons), an indicator for full-time employment, and firm size measured by the number of employees in five categories (1–10, 11–50, 51–250, 251–1000, 1000+). Occupation fixed effects at the two-digit ISCO level. All regressions also control for industry (two-digit ISIC) and country fixed effects. All control variables were used for the entropy balancing. R^2 refers to within-country R^2. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Data source: PIAAC.

Table B.8: Training and Residualized Skills

	Residualized numeracy skills	**Residualized digital skills**
	(1)	(2)
Job training	0.0031	0.0272
	(0.0069)	(0.0092)
Occupation FE	X	X
Further controls	X	X
Entropy balancing	X	X
R^2	0.08	0.08
Observations	72,180	72,180

Notes: Ordinary least squares estimation with weights from entropy balancing. Dependent variable: Numeracy and digital skills, standardized to unit standard deviation across countries, each residualized for literacy skills. Sample: employees aged 25–65 years with information on automation risk and log hourly wages. Job Training: binary variable indicating whether the respondent participated in on-the-job training or job-related training in the 12 months prior to the survey. Further controls include: years of schooling, age group in four categories (25–34, 35–44, 45–54, 55–65), gender, migration status in three categories (first-generation migrant, second-generation migrant, native), parental education in three categories (neither parent has attained upper secondary, at least one parent has attained secondary and post-secondary/non-tertiary, at least one parent has attained tertiary), an indicator whether the respondent has children, age group of the oldest child in four categories (0–2, 3–5, 6–12, 13+), an indicator for participation in non-job-related training (i.e., open/distance education, seminars/workshops, or private lessons), an indicator for full-time employment, and firm size measured by the number of employees in five categories (1–10, 11–50, 51–250, 251–1000, 1000+). Occupation fixed effects at the two-digit ISCO level. All regressions also control for industry (two-digit ISIC) and country fixed effects. All control variables were used for the entropy balancing. R^2 refers to within-country R^2. Standard errors shown in parentheses are two-way clustered at the country and occupation level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Data source: PIAAC.

Table B.9: Balancing

	Training		No Training		No Training (Entropy Weighted)	Difference (2)–(4)	Difference (2)–(5)
	N	Mean/(SE)	N	Mean/(SE)	Mean/(SE)	p-value	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Numeracy skills	49,968	0.283	41,502	-0.178	0.283	0.000	1.000
		(0.004)		(0.005)	(0.004)		
Years of schooling	49,968	14.162	41,502	12.486	14.162	0.000	1.000
		(0.012)		(0.014)	(0.012)		
Age							
25–34 years	49,968	0.282	41,502	0.265	0.282	0.000	1.000
		(0.002)		(0.002)	(0.002)		
35–44 years	49,968	0.296	41,502	0.271	0.296	0.000	1.000
		(0.002)		(0.002)	(0.002)		
45–54 years	49,968	0.267	41,502	0.268	0.267	0.778	1.000
		(0.002)		(0.002)	(0.002)		
55–65 years	49,968	0.156	41,502	0.197	0.156	0.000	1.000
		(0.002)		(0.002)	(0.002)		
Female	49,968	0.517	41,502	0.505	0.517	0.000	1.000
		(0.002)		(0.002)	(0.002)		
Migration status							
Native	49,968	0.854	41,502	0.849	0.854	0.018	1.000
		(0.002)		(0.002)	(0.002)		
First-generation migrant	49,968	0.111	41,502	0.118	0.111	0.001	1.000
		(0.001)		(0.002)	(0.001)		
Second-generation migrant	49,968	0.035	41,502	0.033	0.035	0.332	1.000
		(0.001)		(0.001)	(0.001)		
Parental education							
Neither parent with	49,968	0.295	41,502	0.413	0.295	0.000	1.000
upper secondary education		(0.002)		(0.002)	(0.002)		
At least one parent with	49,968	0.373	41,502	0.348	0.373	0.000	1.000
upper secondary education		(0.002)		(0.002)	(0.002)		
At least one parent with	49,968	0.293	41,502	0.184	0.293	0.000	1.000
tertiary education		(0.002)		(0.002)	(0.002)		
Age of oldest child							
No children	49,968	0.828	41,502	0.807	0.828	0.000	1.000
		(0.002)		(0.002)	(0.002)		
Below 3 years	49,968	0.038	41,502	0.037	0.038	0.283	1.000
		(0.001)		(0.001)	(0.001)		
3–5 years	49,968	0.027	41,502	0.027	0.027	0.991	1.000
		(0.001)		(0.001)	(0.001)		
6–12 years	49,968	0.037	41,502	0.039	0.037	0.111	1.000
		(0.001)		(0.001)	(0.001)		
13 years or more	49,968	0.070	41,502	0.090	0.070	0.000	1.000
		(0.001)		(0.001)	(0.001)		
Training (other)	49,968	0.043	41,502	0.077	0.043	0.000	1.000
		(0.001)		(0.001)	(0.001)		
Full-time employment	49,968	0.885	41,502	0.827	0.885	0.000	1.000
		(0.001)		(0.002)	(0.001)		
Firm size							
1–10 employees	49,968	0.175	41,502	0.335	0.175	0.000	1.000
		(0.002)		(0.002)	(0.002)		
11–50 employees	49,968	0.301	41,502	0.310	0.301	0.005	1.000
		(0.002)		(0.002)	(0.002)		
51–250 employees	49,968	0.268	41,502	0.205	0.268	0.000	1.000
		(0.002)		(0.002)	(0.002)		
251–1000 employees	49,968	0.145	41,502	0.091	0.145	0.000	1.000
		(0.002)		(0.001)	(0.002)		
> 1000 employees	49,968	0.107	41,502	0.051	0.107	0.000	1.000
		(0.001)		(0.001)	(0.001)		

Notes: Balancing table showing covariate means and standard deviations (in parentheses) in the training group (columns 1 and 2), the no-training group (columns 3 and 4), and the no-training group after entropy weighting (column 5). Entropy weighting follows Hainmueller (2012). Entropy weighting also includes fixed effects for occupations (two-digit ISCO level), industries (two-digit ISIC level), and countries (not shown in the balancing table for expositional reasons).
Data source: PIAAC.

Table B.10: Training and Automation Risk: Occupation Fixed Effects at Four-Digit Level

	Automation risk				
	(1)	(2)	(3)	(4)	(5)
Job training	-0.0807***	-0.0427***	-0.0404***	-0.0356***	-0.0348***
	(0.0051)	(0.0031)	(0.0029)	(0.0025)	(0.0022)
Numeracy skills			-0.0155***	-0.0082***	-0.0071**
			(0.0023)	(0.0021)	(0.0034)
Occupation FE		X	X	X	X
Further controls				X	X
Entropy balancing					X
R^2	0.11	0.27	0.28	0.29	0.27
Observations	48,877	48,764	48,764	48,764	48,764

Notes: Ordinary least squares estimation in columns (1)–(4) with weights such that each country has the same weight, least squares estimation with weights from entropy balancing in column (5). Dependent variable: individual-level automation risk, ranging from 0 (indicating a low probability that a worker is fully automated) to 1 (indicating a high probability that a worker is fully automated); automation risk is predicted using items on task use at work from PIAAC. Sample: employees aged 25–65 years with information on automation risk and log hourly wages. Sample is restricted to countries that report occupations at the four-digit level. Job Training: binary variable indicating whether the respondent participated in on-the-job training or job-related training in the 12 months prior to the survey. Numeracy skills are standardized to unit standard deviation across countries. Further controls include: years of schooling, age group in four categories (25–34, 35–44, 45–54, 55–65), gender, migration status in three categories (first-generation migrant, second-generation migrant, native), parental education in three categories (neither parent has attained upper secondary, at least one parent has attained secondary and post-secondary/non-tertiary, at least one parent has attained tertiary), an indicator whether the respondent has children, age group of the oldest child in four categories (0–2, 3–5, 6–12, 13+), an indicator for participation in non-job-related training (i.e., open/distance education, seminars/workshops, or private lessons), an indicator for full-time employment, and firm size measured by the number of employees in five categories (1–10, 11–50, 51–250, 251–1000, 1000+). Occupation fixed effects at the four-digit ISCO level. All regressions also control for industry (two-digit ISIC) and country fixed effects. All control variables were used for the entropy balancing in column (5). R^2 refers to within-country R^2. Standard errors shown in parentheses are two-way clustered at the country and occupation level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Data source: PIAAC.

Table B.11: Training and Automation Risk: Unrestricted Sample

	Automation risk				
	(1)	(2)	(3)	(4)	(5)
Job training	-0.0829***	-0.0556***	-0.0508***	-0.0458***	-0.0460***
	(0.0060)	(0.0081)	(0.0068)	(0.0064)	(0.0068)
Numeracy skills			-0.0219***	-0.0166***	-0.0126***
			(0.0040)	(0.0034)	(0.0026)
Occupation FE		X	X	X	X
Further controls				X	X
Entropy balancing					X
R^2	0.11	0.21	0.22	0.23	0.20
Observations	101,949	101,949	101,949	101,949	101,949

Notes: The table replicates Table 1 when not restricting the sample to individuals with non-missing wage information. Ordinary least squares estimation in columns (1)–(4) with weights such that each country has the same weight, least squares estimation with weights from entropy balancing in column (5). Dependent variable: individual-level automation risk, ranging from 0 (indicating a low probability that a worker is fully automated) to 1 (indicating a high probability that a worker is fully automated); automation risk is predicted using items on task use at work from PIAAC. Sample: employees aged 25–65 years. Job Training: binary variable indicating whether the respondent participated in on-the-job training or job-related training in the 12 months prior to the survey. Controls: numeracy skills (standardized to unit standard deviation across countries), years of schooling, age group in four categories (25–34, 35–44, 45–54, 55–65), gender, migration status in three categories (first-generation migrant, second-generation migrant, native), parental education in three categories (neither parent has attained upper secondary, at least one parent has attained secondary and post-secondary/non-tertiary, at least one parent has attained tertiary), an indicator whether the respondent has children, age group of the oldest child in four categories (0–2, 3–5, 6–12, 13+), an indicator for participation in non-job-related training (i.e., open/distance education, seminars/workshops, or private lessons), an indicator for full-time employment, and firm size measured by the number of employees in five categories (1–10, 11–50, 51–250, 251–1000, 1000+). Occupation fixed effects at the two-digit ISCO level. All regressions also control for industry (two-digit ISIC) and country fixed effects. All control variables were used for the entropy balancing in column (5). R^2 refers to within-country R^2. Standard errors shown in parentheses are two-way clustered at the country and occupation level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Data source: PIAAC.

Table B.12: Training and Automation Risk: Region × Industry × Firm Size Fixed Effects

	Automation risk	
	(1)	(2)
Job training	-0.0384	-0.0333
	(0.0022)	(0.0040)
Numeracy skills	-0.0087	-0.0090
	(0.0015)	(0.0016)
Occupation FE	X	X
Further controls	X	X
Entropy balancing	X	X
Region × industry × firm size FE		X
R^2	0.24	0.55
Observations	57,874	57,874

Notes: Ordinary least squares estimation with weights from entropy balancing. Dependent variable: individual-level automation risk, ranging from 0 (indicating a low probability that a worker is fully automated) to 1 (indicating a high probability that a worker is fully automated); automation risk is predicted using items on task use at work from PIAAC. Sample: employees aged 25–65 years with information on automation risk and log hourly wages. Job Training: binary variable indicating whether the respondent participated in on-the-job training or job-related training in the 12 months prior to the survey. Numeracy skills are standardized to unit standard deviation across countries. Column (1) replicates column (5) in Table 1 for countries that provide regional information in PIAAC. Column (2) includes region × industry × firm size fixed effects as a proxy for firm fixed effects. Regional information in PIAAC is available at the two-digit territorial level, i.e., the first administrative tier of sub-national government (e.g., federal states in Germany). Regional information at the two-digit territorial level is not available for 7 out of 37 countries in our sample: Australia, Austria, Canada, Finland, Norway, Turkey, and the United States. Industry fixed effects are at the two-digit ISIC level, whereas firm size is measured by the number of employees in five categories (1–10, 11–50, 51–250, 251–1000, 1000+). Further controls include: years of schooling, age group in four categories (25–34, 35–44, 45–54, 55–65), gender, migration status in three categories (first-generation migrant, second-generation migrant, native), parental education in three categories (neither parent has attained upper secondary, at least one parent has attained secondary and post-secondary/non-tertiary, at least one parent has attained tertiary), an indicator whether the respondent has children, age group of the oldest child in four categories (0–2, 3–5, 6–12, 13+), an indicator for participation in non-job-related training (i.e., open/distance education, seminars/workshops, or private lessons), and an indicator for full-time employment. Occupation fixed effects at the two-digit ISCO level. All regressions also control for industry (two-digit ISIC) and country fixed effects. All control variables included in the respective specification were used for the entropy balancing. R^2 refers to within-country R^2. Standard errors shown in parentheses are two-way clustered at the country and occupation level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Data source: PIAAC.

Table B.13: Training and Automation Risk: Coefficient Bounds

	δ = 0.8	δ = 1	δ = 1.2
R_{max} = 1.3$\tilde{R}$	-0.0441	-0.0427	-0.0413
	(0.0016)	(0.0012)	(0.0016)
R_{max} = 1.4$\tilde{R}$	-0.0393	-0.0366	-0.0338
	(0.0019)	(0.0018)	(0.0020)
R_{max} = 1.5$\tilde{R}$	-0.0344	-0.0304	-0.0262
	(0.0015)	(0.0015)	(0.0017)
Observations	91,470	91,470	91,470

Notes: This table shows coefficient bounds following Oster (2019) of our baseline job training estimate (see Table 1, column 5) for different assumptions regarding R_{max} and δ. R_{max} is the hypothetical maximum R-squared that can be explained by observables and unobservables, as a factor of $\tilde{R}$, the R-squared of our baseline regression with observable covariates in column (5) of Table 1 ($\tilde{R} = 0.20$). We follow Chen (2021) in relying on the within-$\tilde{R}$ (i.e., after partialling out country fixed effects) to estimate all values. Following the recommendation in Oster (2019), we provide coefficient bounds for $R_{max} = 1.4\tilde{R}$ and additional checks for the cases $R_{max} = 1.3\tilde{R}$ and $R_{max} = 1.5\tilde{R}$. δ indicates the degree of selection on unobservables relative to observables; if $\delta = 1$, unobservable factors are as influential as observable ones in determining the outcome. Again following Chen (2021), we include the case of $\delta = 1$, and additionally provide coefficient bounds for $\delta = 0.8$ and $\delta = 1.2$. Standard errors shown in parentheses are two-way clustered at the country and occupation level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Data source: PIAAC.

Table B.14: Change in Training Effectiveness over Time in the United States and Germany

	Automation risk					
	United States			Germany		
	Pooled	2012 Only	2017 Only	Pooled	2012 Only	2015 Only
	(1)	(2)	(3)	(4)	(5)	(6)
Job training	-0.0319***	-0.0271***	-0.0397***	-0.0469***	-0.0357***	-0.0726***
	(0.0033)	(0.0096)	(0.0119)	(0.0073)	(0.0092)	(0.0098)
Numeracy skills	0.0029	0.0007	0.0097	-0.0190***	-0.0192***	-0.0152*
	(0.0035)	(0.0071)	(0.0071)	(0.0044)	(0.0043)	(0.0087)
R^2	0.32	0.37	0.34	0.30	0.31	0.41
Observations	4,367	2,430	1,937	4,976	3,069	1,907

Notes: Least squares estimation with weights from entropy balancing. Dependent variable: individual-level automation risk, ranging from 0 (indicating a low probability that a worker is fully automated) to 1 (indicating a high probability that a worker is fully automated); automation risk in each wave is predicted using items on task use at work from PIAAC. Sample: employees aged 25–65 years in the United States (columns 1–3) and in Germany (columns 4–6); restrictions are applied in each survey wave separately. Job Training: binary variable indicating whether the respondent participated in on-the-job training or job-related training in the 12 months prior to the survey; the 2015 PIAAC wave in Germany elicits participation in job training in 2014. Numeracy skills are standardized to unit standard deviation across countries. Numeracy skills are standardized to unit standard deviation within Germany. Further controls include: years of schooling, years of schooling, age group in four categories (25–34, 35–44, 45–54, 55–65), gender, migration status in three categories (first-generation migrant, second-generation migrant, native), parental education in three categories (neither parent has attained upper secondary, at least one parent has attained secondary and post-secondary/non-tertiary, at least one parent has attained tertiary), an indicator whether the respondent has children, age group of the oldest child in four categories (0–2, 3–5, 6–12, 13+), an indicator for participation in non-job-related training (i.e., open/distance education, seminars/workshops, or private lessons), an indicator for full-time employment, and firm size measured by the number of employees in five categories (1–10, 11–50, 51–250, 251–1000, 1000+). Analysis for Germany excludes controls for child-related variables and non-work-related training, as these data were not collected in the 2015 PIAAC wave. Occupation fixed effects at the two-digit ISCO level. All regressions also control for industry (two-digit ISIC) and country fixed effects. Entropy balancing is based on all control variables included in the respective specification. Standard errors shown in parentheses are clustered at the occupation level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Data source: PIAAC United States (2012 and 2017 waves); PIAAC Germany (2012 and 2015 waves).

Table B.15: Training and Digital Skills, Imputed Digital Skills for Missing Data

	Digital skills			
	Baseline	Missings imputed with		
		zeros	global minimum	country minimum
	(1)	(2)	(3)	(4)
Job Training	0.0509***	0.1046***	0.1045***	0.0949***
	(0.0088)	(0.0135)	(0.0132)	(0.0132)
Numeracy Skills	0.7762***	0.3490***	0.3899***	0.4413***
	(0.0112)	(0.0187)	(0.0186)	(0.0234)
Observations	72,180	91,470	91,470	91,470
R^2	0.59	0.32	0.32	0.41
Occupation FE	X	X	X	X
Further controls	X	X	X	X
Entropy balancing	X	X	X	X

Notes: Least squares estimation with weights from entropy balancing. Dependent variable: digital skills standardized to standard deviation 1 across countries with different imputations. Sample: employees aged 25–65 years with information on automation risk and log hourly wages; column (1) replicates column (5) of Table 5, that is, it is restricted to employees with non-missing digital skills. Job Training: binary variable indicating whether the respondent participated in on-the-job training or job-related training in the 12 months prior to the survey. Controls: numeracy skills (standardized to unit standard deviation across countries), years of schooling, age group in four categories (25–34, 35–44, 45–54, 55–65), gender, migration status in three categories (first-generation migrant, second-generation migrant, native), parental education in three categories (neither parent has attained upper secondary, at least one parent has attained secondary and post-secondary/non-tertiary, at least one parent has attained tertiary), an indicator whether the respondent has children, age group of the oldest child in four categories (0–2, 3–5, 6–12, 13+), an indicator for participation in non-job-related training (i.e., open/distance education, seminars/workshops, or private lessons), an indicator for full-time employment, and firm size measured by the number of employees in five categories (1–10, 11–50, 51–250, 251–1000, 1000+). Industry fixed effects at the two-digit ISIC level and occupation fixed effects at the two-digit ISCO level. All regressions also control for country fixed effects. All control variables were used for the entropy balancing. R^2 refers to within-country R^2. Standard errors shown in parentheses are two-way clustered at the country and occupation level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Data source: PIAAC.

Table B.16: Training and Basic Digital Skills

	Basic digital skills				
	(1)	(2)	(3)	(4)	(5)
Job training	0.1366***	0.0988***	0.0776***	0.0647***	0.0401***
	(0.0123)	(0.0122)	(0.0092)	(0.0080)	(0.0059)
Numeracy skills			0.1004***	0.0767***	0.0540***
			(0.0100)	(0.0086)	(0.0078)
Occupation FE		X	X	X	X
Further controls				X	X
Entropy balancing					X
R^2	0.08	0.14	0.18	0.22	0.15
Observations	91,470	91,470	91,470	91,470	91,470

Notes: Ordinary least squares estimation in columns (1)–(4) with weights such that each country has the same weight, least squares estimation with weights from entropy balancing in column (5). Dependent variable: Binary indicator for basic digital skills, which takes a value of one if the respondent was able to participate in PIAAC in a computer-based mode, zero otherwise. There are three reasons for why individuals may lack basic digital skills in PIAAC (see Falck et al., 2021, 2022): (i) individuals had no prior computer experience; (ii) individuals failed a computer core test, which assessed basic digital competencies such as using a keyboard/mouse or scrolling through a web page; (iii) individuals refused to take part in the computer-based assessment. Sample: employees aged 25–65 years with information on automation risk and log hourly wages. Job Training: binary variable indicating whether the respondent participated in on-the-job training or job-related training in the 12 months prior to the survey. Controls: numeracy skills (standardized to unit standard deviation across countries), years of schooling, age group in four categories (25–34, 35–44, 45–54, 55–65), gender, migration status in three categories (first-generation migrant, second-generation migrant, native), parental education in three categories (neither parent has attained upper secondary, at least one parent has attained secondary and post-secondary/non-tertiary, at least one parent has attained tertiary), an indicator whether the respondent has children, age group of the oldest child in four categories (0–2, 3–5, 6–12, 13+), an indicator for participation in non-job-related training (i.e., open/distance education, seminars/workshops, or private lessons), an indicator for full-time employment, and firm size measured by the number of employees in five categories (1–10, 11–50, 51–250, 251–1000, 1000+). Occupation fixed effects at the two-digit ISCO level. All regressions also control for industry (two-digit ISIC) and country fixed effects. All control variables were used for the entropy balancing in column (5). R^2 refers to within-country R^2. Standard errors shown in parentheses are two-way clustered at the country and occupation level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.
Data source: PIAAC.