How do Workers Learn? Theory and Evidence on the Roots of Lifecycle Human Capital Accumulation*

Xiao Ma — *Peking University*

Alejandro Nakab — *Universidad Torcuato Di Tella*

Daniela Vidart — *University of Connecticut*

First version: September 2020; **This version**: October 2022

*Please click **here** for the most recent version*

Abstract

How do the sources of worker learning change over the lifecycle, and how do these changes affect on-the-job human capital accumulation and wage growth? We use detailed worker qualification data from Germany and the US to document that internal learning (learning through colleagues) decreases with worker experience, while external learning (on-the-job training) has an inverted U-shape in worker experience. To shed light on these findings, we build an analytical model where the incentives to learn from each of these two sources evolve throughout the lifecycle due to shifts in the relative position of the worker in the human capital distribution. We embed this two-source learning mechanism in a quantitative Burdett and Mortensen search framework where firms and workers jointly fund learning investments. The model equilibrium replicates our empirical lifecycle results, and shows that more productive firms provide better learning environments by offering greater variety in learning options. Counterfactual analyses imply that the two sources are highly complementary, that internal learning is more critical for young workers' wage growth although external learning is more critical for overall lifetime wage growth, and that the lifecycle increase in wage dispersion is driven by external learning.

*Email: xiaoma@phbs.pku.edu.cn, anakab@utdt.edu and daniela.vidart@uconn.edu.

We would like to thank Titan Alon, Victoria Gregory, David Lagakos, Jeremy Lise, Simon Mongey, Marc Muendler, Joseph Mullins, Mariacristina de Nardi, Tommaso Porzio, Valerie Ramey, Todd Schoellman, Kjetil Storesletten, and Johannes Wieland for their helpful comments. We are also grateful for the insightful comments of participants at seminars at UCSD, CU Boulder, the Univ. of Hawaii at Manoa, Univ. Torcuato di Tella, IADB, NUS, UIUC, PHBS Macro and Finance Workshop, BSE Summer Forum, SED, Univ. of Minnesota, and ASU.

1 Introduction

Ever since Becker (1962), the economics literature on human capital has recognized the importance of on-the-job learning for understanding the dynamics and dispersion of lifecycle earnings (Rubinstein and Weiss (2006)). Work in this literature has identified several key inputs driving on-the-job human capital acquisition, including on-the-job training (Acemoglu (1997); Acemoglu and Pischke (1998), Moen and Rosén (2004), Ma et al. (2020)), learning-by-doing (Bagger et al. (2014), Gregory (2019)), and coworker quality (Nix (2017), Herkenhoff et al. (2018), Jarosch et al. (2019)).[1] This literature to date has focused on studying each source of human capital accumulation individually, and thus has not yet considered how different inputs interact and jointly shape on-the-job skill acquisition. In this paper, we study the shared scope of different sources of learning to influence lifetime human capital and wage dynamics.

Motivated by the literature and data, we focus on two sources of learning: internal learning (or learning through colleagues), which draws on firms' internal knowledge, and thus crucially depends on coworker quality and firm structure; and external learning (or external on-the-job training), which draws on external knowledge, and is thus potentially less sensitive to these firm-level aspects, but may depend on broader institutional aspects. Distinguishing between these two sources of learning and studying their shared scope to influence lifetime wage dynamics is important for several reasons. First, work in the labor literature suggests that the productivity and earnings gains of on-the-job training can vary greatly depending on the type of learning opportunity provided.[2] Second, given that these two sources draw from separate knowledge pools and are affected by fundamentally different processes, their relevance may vary with worker and firm characteristics. This implies, for instance, that the effectiveness of policies promoting internal or external learning, such as apprenticeship or re-training programs, respectively, may vary greatly depending on the age, sex, occupation, or other characteristics of the workers targeted, along with the size, industry, and performance of the firms involved. Finally, complementarities between different sources of learning may exist, amplifying the role of each source to explain lifecycle wage growth, and suggesting that policies that encourage multiple sources of learning may be superior.

[1] Other inputs explored in the literature include formal schooling (Ben-Porath (1967)), knowledge hierarchies, (Garicano (2000); Garicano and Rossi-Hansberg (2004, 2006), Caicedo et al. (2019)), complementary goods (Manuelli and Seshadri (2014)), and managerial inputs (Burstein and Monge-Naranjo (2007)).

[2] See Heckman et al. (1999), Kluve (2010), Card et al. (2018), McKenzie (2017) and What Works - Centre for Local Economic Growth (2016) for reviews on this evidence.

In the empirical section of the paper, we first document two novel facts that speak to the importance of our two sources of learning from both firm and worker perspectives. First, using enterprise survey data from Europe, we show that both internal and external sources of learning are widely provided by firms to their workers; and that larger (and thus more productive) firms offer greater variety in learning options by providing their workers with more opportunities to engage in both internal and external learning. Second, we use detailed worker qualification data from Germany and the United States to show that both sources of learning are important to workers, and have markedly different lifecycle patterns. In particular, we document that (1) the prevalence of internal learning decreases with worker experience; and (2) the prevalence of external learning has an inverted U-shape in worker experience. These lifecycle patterns are robust to considering alternate worker experience definitions, controlling for industry, occupation and demographics, and decomposing the data across several worker and firm-level characteristics such as educational level, gender and firm size.

We then build a benchmark model that sheds light on these lifecycle findings by examining how the incentives to accumulate human capital through each source of learning evolve throughout the lifecycle. The benchmark model is rich enough to match our lifecycle facts and generate several predictions we can test in the data, yet simple enough to yield analytical results for the dynamics of worker learning. The model follows a Blanchard–Yaari "perpetual youth" structure and features two sectors: a final goods production sector, and a training sector providing external learning services. Production in each of these sectors is respectively carried out by production workers and trainers whose productivity is determined by their level of human capital. This human capital stock follows a ladder structure with a discrete number of steps. Workers in the production sector can choose to spend their time either working, or attempting to climb the human capital ladder via internal or external learning. Learning from each of these sources follows from random meetings with coworkers and trainers respectively, and is contingent on matching with a coworker or trainer with a higher human capital than the own. Both forms of learning carry a foregone production cost, but external learning carries an additional cost from the purchase of training services.

Incentives to engage in each source of skill acquisition evolve throughout workers' lifecycles as they accumulate human capital. In particular, changes in the relative position of the worker in the human capital distribution mediate the supply of coworkers and trainers that can be learnt from, and lead to distinct lifecycle patterns of learning. Consistent with our data, young workers in the model disproportionately rely on coworkers to learn, since coworker learning is relatively cheap and the proportion of coworkers with a larger stock of knowledge is high. As workers age this proportion declines, leading to a switch to external learning since trainers tend to have higher average human capital levels than production workers. As human capital continues to accumulate, the opportunity cost of learning rises, and progressively more individuals reach the last human capital level ladder step, leading the portion of external learners to decline.

We then test for the existence of the benchmark model's key predictions in the data. First, we provide evidence showing that the distribution of trainers across the human capital state-space is skewed left relative to production workers. Second, we present evidence matching our theory's key learning predictions: (1) the portion of individuals that do not engage in the two sources of learning explored rises with human capital; and (2) the average human capital is lowest for individuals engaging in internal learning, followed by individuals engaging in external learning, and highest for trainers. Finally, we present evidence on a natural implication of our theory: workers whose jobs require the use of new and innovative techniques rely more on external learning and thus knowledge that is not currently available in the firm.

We then build a quantitative version of the model that considers rich interactions of firms and workers in learning investments in order to quantify the importance of internal and external sources of learning to human capital accumulation and wage growth, and understand the role of both sources of learning in shaping firms' learning environments. To this end, we embed the two-source human capital ladder mechanism formalized in the benchmark model within a Burdett-Mortensen framework where firms and workers jointly choose and fund learning investments. Similar to the analytical framework, this model follows a Blanchard–Yaari structure and features a training sector and a production sector. We assume that the training sector is frictionless, while the production sector is characterized by labor market frictions and firm heterogeneity. Firms in the production sector post vacancies and wage rates per efficiency unit of labor to attract both unemployed individuals and workers from other firms, and meet their matches by random search. After matching, workers and firms jointly decide and pay for internal and external learning investments.

We calibrate the model to the United States economy, and find that the stationary equilibrium of the model replicates the lifecycle results found empirically (non-targeted). The model's stationary equilibrium also highlights the importance of firms' learning environments in the formation of human capital. At all levels of human capital, workers in more productive firms spend more time in both internal and external learning, and thus climb the human capital ladder faster. This finding matches our empirical evidence showing that workers in larger European firms spend significantly more hours engaging in both sources of learning, and is also consistent with evidence found by Engbom (2017) and Arellano-Bover (2020) showing that workers in more productive firms exhibit faster rates of skill acquisition, and evidence found by Gregory (2019) showing that having different forms of training available is important for firms' learning environments. In our model, more productive firms invest more in both types of learning since they exhibit both larger returns to skill acquisition (due to supermodularity of the production function), and a better pool of coworkers to learn from (due to positive assortative matching between firms and workers).[3,4]

To assess the importance of internal and external learning in the formation of human capital, wage growth, and wage dispersion, we perform a counterfactual analyses where we subsequently shut down each of these two sources of skill acquisition and examine how the stationary equilibrium changes. We find that both internal and external learning contribute largely and roughly equally to workers' human capital: without external learning, workers' human capital decreases by 30%, whereas without internal learning, workers' human capital decreases by 29%. In addition, we find that the two sources of learning are highly complementary in the aggregate, since the existence of each source of learning improves the pool of workers and thus the probability of learning from the other source.

Finally, we find that although internal learning is more critical for young workers' wage growth, external learning is more critical for overall lifetime wage growth. Given that internal learning is cheaper than external learning, young workers' wages grow considerably slower in the scenario without internal learning than that without external learning. However, since the returns to internal learning are depleted more quickly, older workers' wages grow considerably slower in the scenario without external learning than that without internal learning, significantly depressing overall lifetime wage growth. These mechanisms also give rise to important differences in the evolution of wage dispersion in the two counterfactual

[3] This positive assortative matching pattern emerges in our framework due to (1) the more favorable learning environments prevalent in more productive firms which allow workers to climb the human capital ladder faster; and (2) on-the-job search, which helps more productive firms poach employed workers who tend to be more skilled than the unemployed.

[4] The importance of coworkers is further confirmed when we simulate data for a panel of 10,000 workers for 40 years using our model, and find a positive correlation between the future wage realizations of workers, and the wage of their coworkers that is similar in magnitude to that found by Herkenhoff et al. (2018). Similar to their results, we also find that these results are particularly marked for workers who are paid less than their coworkers.

scenarios, and imply that the lifecycle increase in wage dispersion is driven by external learning in our model. Without external learning, the dispersion in human capital and wages remains low throughout the lifecycle as workers learn from and catch up to colleagues whose learning opportunities have been exhausted. Without internal learning, since skill acquisition is more expensive, only a small number of workers climb the human capital ladder, causing the dispersion in human capital and wages to rise throughout the lifecycle.

The paper is organized as follows. In Section 2 we present a literature review. In Section 3, we describe the data and methods used for the empirical assessment of the facts presented. In Section 4 we present the benchmark model and perform an empirical assessment of its testable predictions. In Section 5 we present the quantitative model, calibration, and results. We conclude in Section 6.

2 Literature Review

Our paper is most closely related to the literature exploring the importance of on-the-job skill acquisition on human capital accumulation and wage growth. Our theory provides a unified structure to jointly consider the roles of internal and external sources of learning, and thus relates to different strands within this literature. First, our paper relates to the literature that has explored the role of peers in knowledge diffusion within coworker and production teams (Garicano (2000), Garicano and Rossi-Hansberg (2004, 2006), Azoulay et al. (2010), Luttmer (2014), Nix (2017), Herkenhoff et al. (2018), Jarosch et al. (2019), Caicedo et al. (2019)) and within the population at large (Jovanovic (2014), Lucas and Moll (2014), Perla and Tonetti (2014), de la Croix et al. (2016), Benhabib et al. (2021)). Our paper is particularly related to de la Croix et al. (2016), who explore the role of old-to-young knowledge transmission mechanisms such as guilds or journeymanship in the dissemination of knowledge in Europe. Similar to them, our model features old-to-young knowledge transmission both through external training and coworkers. However, relative to this and other papers in the literature, our paper makes an explicit distinction between sources of learning that draw on knowledge from inside and outside the firm (internal v. external), highlighting particularly the importance of the finite distribution of coworker human capital, and consequent limiting nature of coworker learning in explaining lifecycle human capital acquisition patterns. Moreover, through its focus on the importance of external on-the-job training on skill acquisition, this paper relates to both the theory on general training investments, first proposed by Becker (1964), and later developed by others (Acemoglu (1997), Acemoglu and

Pischke (1998), Acemoglu and Pischke (1999), Autor (2001), Moen and Rosén (2004)).

Our paper also relates to the vast literature exploring the interaction between learning and lifecycle dynamics. First, our paper relates to the literature that examines the effects of work-related human capital acquisition on earnings. Much of this literature has focused on disentangling the role of learning from search dynamics and shocks on earnings growth (Bunzel et al. (1999), Rubinstein and Weiss (2006), Huggett et al. (2011), Barlevy (2008), Yamaguchi (2010), Burdett et al. (2011), Bowlus and Liu (2013), Bagger et al. (2014), Gregory (2019), Karahan et al. (2022)). This contrasts with our goal, which is to disentangle the contributions of different sources of learning to human capital accumulation and earnings growth. Second, and given that we add coworker- and external training-based learning options where human capital acquisition may be enhanced when production is given up, our paper relates to the literature highlighting the trade-off between learning and work, following seminal papers on on-the-job learning of Ben-Porath (1967), Heckman (1976), and Rosen (1976). This contrasts with several recent papers which examine the role of on-the-job human capital accumulation on knowledge diffusion or earnings growth (Lucas (2009), Bagger et al. (2014), Gregory (2019)). These papers model on-the-job human capital accumulation via learning-by-doing, and thus do not consider multiple sources of learning, or a tradeoff between learning and work.[5]

By examining how the provision of different sources of learning shapes firms' learning environments, our paper also relates to the literature that considers the role of firms and firm-level characteristics on workers' human capital accumulation and earnings dynamics (Gregory (2019), Arellano-Bover (2020), Engbom (2021), Friedrich et al. (2021), Jarosch (2022), Engbom et al. (2022)). This literature has focused on showing that there is substantial heterogeneity in firms' promotion of human capital accumulation, and that this is an important determinant of lifecycle earning dynamics. However, the drivers of firms' learning environments are still poorly understood. In our paper we contribute in this direction by providing empirical and theoretical evidence of a concrete driver of success in firms' learning environments: variety in learning opportunities stemming from internal and external learning sources.[6]

[5] Our paper also relates to the literature examining the link between post-schooling human capital accumulation and growth (Manuelli and Seshadri (2014), Lagakos et al. (2018), Ma et al. (2020)).

[6] In addition, given the importance of matching between different firm and worker types in shaping learning environments in our theory, our paper also relates to the literature on sorting in frictional labor markets (Shimer and Smith (2000), Teulings and Gautier (2004), Eeckhout and Kircher (2011), Lise et al. (2016), Hagedorn et al. (2017), De Melo (2018), Bagger and Lentz (2019)). In our model, sorting arises from

Our paper also relates to the vast labor literature examining the impacts of on-the-job learning opportunities on worker earnings and employment (see Heckman et al. (1999), Kluve (2010), Card et al. (2018), McKenzie (2017) and What Works - Centre for Local Economic Growth (2016) for reviews on this evidence), and particularly the literature showing that the productivity and earnings gains of on-the-job training can vary greatly depending on the type of learning opportunity provided. One key distinction highlighted in these studies arises from comparing in-firm to classroom-based on-the-job learning opportunities, which broadly matches our internal and external categories of learning, respectively. For instance, Fitzenberger and Völter (2007) find that an on-the-job training program designed to improve professional skills through medium-term in-classroom courses in East-Germany increased the probability of employment of participants, while re-training and training programs conducted in practice studios did not have any effects. Several other studies also find significant differences between in-firm and in-classroom learning opportunities, and also along other dimensions such as length of learning programs, type of skills targeted (basic versus advanced), private versus public provision, etc (see What Works - Centre for Local Economic Growth (2016) for a review).

3 Data and Empirical Findings

We now turn our attention to analyzing the importance and lifecycle patterns of different sources of learning. For this purpose, we use enterprise data from Europe, detailed worker qualification data from Germany, and adult education data from the United States. In this section, we first describe the data, and then proceed to document two facts about the importance of different sources of learning from firms' and workers' perspectives. We include further details on data sources and analysis in Appendix A.

3.1 Data

To document our first fact on the provision of internal and external sources of learning by firms, we rely on data from the European Union's Continuing Vocational Training (EU-CVT) enterprise survey. This survey collects information from enterprises across the European Union, and focuses on enterprises' investments in continuing vocational training of their staff, and provides information on the types, content and volume of continuing training,

incentives to human capital accumulation, and on-the job search. This contrasts with most of this literature, which focuses on sorting arising from worker and firm type complementarities.

enterprises' own training resources and use of external training providers, costs of continuing training, and initial vocational training. Due to data availability, we rely on three of the five waves of EU-CVT conducted in 2005, 2010, and 2015, labeled as CVT3, CVT4 and CVT5. These three surveys provide a sample of 78,000, 101,000 and 111,000 enterprises, respectively, from across all EU member states and Norway. For further details on this data please see Appendix A.1.

To document our second set of facts regarding the importance of different sources of learning for workers throughout the lifecycle, we use detailed worker qualification data from Germany, and adult education data from the United States. The German data we use spans across 7 waves conducted in 1979, 1985, 1999, 2006, 2012 and 2018. This data was collected by the BIBB (Bundesinstitut fr Berufsbildung, Bonn), a federal agency devoted to vocational education, in conjunction with the IAB (in 1979, 1985, 1991 and 1999) and BAuA (in 2006, 2012 and 2018). This data comprises several questions about worker qualifications and working conditions in Germany. All surveys include measures of on-the-job skill acquisition, formal education, and occupational skill requirements. The data collection strategy was designed to cover a representative sample of 20,000 to 35,000 members of the German labor-force. The survey is repeated every 6 years to a different set of subjects, yielding a repeated cross-section structure. For further details on this data please see Appendix A.2.

The US data corresponds to data from the 2016 wave of the National Household Education Survey (NHES), and specifically the module on Adult Training and Education (ATES), which contains detailed adult education data including formal education and on-the-job skill acquisition, along with detailed employment information and respondent background characteristics. The data for the ATES collection focused on non-institutionalized adults ages 16—65 not enrolled in grade 12 or below, and comprised 47,744 individuals which are representative of the US population at large. The ATES survey was first deployed as part of the 2016 NHES Survey, and has not been deployed again in more recent waves. For further details on this data please see Appendix A.3.

3.2 Fact 1: Firms Provide Both Internal and External Learning; Larger Firms Offer Greater Variety

Using the EU-CVT data, we first show that a large contingent of firms in the European Union offer their workers both learning opportunities that rely on resources and knowledge pools that are internal to the firm, along with learning opportunities that rely on resources and knowledge pools that are external to the firm. In addition, we show that larger (and thus more productive) firms provide better learning environments by offering greater variety in learning options.

We distinguish between internal and external learning opportunities by relying on information on the location and instructor affiliation of CVT activities. CVT encompasses educational or training activities that are planned in advance, organized, or supported with the specific goal of learning. The survey explicitly distinguishes between "Internal CVT Courses" and "External CVT Courses" by separating courses, seminars or activities that take place inside firms and employ internal trainers from those that occur outside firms or employ external trainers. In addition, the survey also measures "Other types of CVT Activities", which are geared towards learning and are typically connected the active workplace. These are often characterized by self-organization by a group of learners within the firm. Within these activities, we consider four learning sources which can be categorized into either internal or external learning: Participation in conferences and lectures, Guided-on-the job training, Job rotation, and Learning or quality circles.[7] We categorize firms that invest in "External Learning" as those offering External CVT Courses and/or Other types of CVT Activities in the form of Conferences, Workshops or Lectures, and firms that invest in "Internal Learning" as those offering Internal CVT Courses, and/or the remaining categories in Other types of CVT Activities. In Table B.1 we show the proportion of firms for which workers participate in Internal and External CVT Courses, along with Other types of CVT Activities, for all EU countries. We find that a large portion of firms in all countries surveyed offer Internal and External CVT Courses, along with Other types of CVT Activities.

3.2.1 Fact 1a: Firms Provide Both Internal and External Learning

In Table 3.1, we show the proportion of firms that invest in either External or Internal Learning, and both External and Internal learning activities simultaneously. We find that a large contingent of firms surveyed invest in both External and Internal learning activities simultaneously. In particular, 41% of the firms surveyed offer both External and Internal learning opportunities to their employees.[8] This shows that in the context of Europe, both internal and external learning sources are part of the learning opportunities offered by firms to their

[7] We provide more detailed definitions and characteristics of each source of learning in Appendix A.1.

[8] We show robustness to these patterns in Appendix B. Specifically, we show that all of the following are substantial: (1) the share of time workers spend in CVT courses (internal or external) (Table B.2); (2) the share of workers engaging in Other CVT Activities (Table B.3); and (3) the share of firms providing internal and external learning activities of different kinds (Table B.4).

workers, and thus suggests that both sources are important from firms' perspectives.

Table 3.1: Share of firms Providing Internal & External Learning Activities

		External Learning 0	External Learning 1
Internal Learning	0	0.33	0.15
	1	0.11	0.41

Notes: This table shows the proportion of firms in the whole sample which reported having persons employed participating in any kind of Internal and External CVT activities. Data from CVTS3, CVT4 and CVT5 surveys. Weighting factors were used in order to calculate proportions for each wave.

3.2.2 Fact 1b: Larger Firms Offer Greater Variety in Learning

We now compare how these learning opportunities differ across firms of different sizes.[9] In Table 3.2, we show that 32%, 46%, and 71% of firms with 5–19, 20–99, and 100+, workers respectively offer both External and Internal learning activities to their workers, and thus that the share of large firms offering both sources of learning is much larger than that of smaller firms. Interestingly, there much smaller differences in the proportion of firms offering only one source of learning by firm size, suggesting that large firms favor learning environments with both sources of learning (and thus learning variety), rather than ones with only one source of learning.[10]

Table 3.2: Share of firms Providing Internal & External Learning Activities by Firm Size

		External Learning 0	External Learning 1			External Learning 0	External Learning 1			External Learning 0	External Learning 1
Internal Learning	0	0.41	0.16	Internal Learning	0	0.28	0.15	Internal Learning	0	0.13	0.09
	1	0.11	0.32		1	0.11	0.46		1	0.07	0.71
Small firms, 5-19 Workers				Medium firms, 20-99 Workers				Large firms, 100+ Workers			

Notes: These tables show the proportion of firms of different size categories in the whole sample which reported having persons employed participating in internal and external learning activities. Data from CVTS3, CVT4 and CVT5 surveys. Weighting factors were used in order to calculate proportions for each wave. We consider firms with 5 or more employees only since smaller firms encompass a very small part of the sample, and may thus be highly selected.

[9] In what follows we consider firms with 5 or more employees since smaller firms encompass a very small part of the sample, and may thus be highly selected.

[10] In Table B.5, we show that the positive correlation between firm size and learning opportunities is robust to controlling for industry, socioeconomic, and country-year fixed effects. In addition, in Table B.6 we show that workers in larger firms spend significantly more hours engaging in both sources of learning. In particular, workers in firms with 100 or more workers spend approximately double the number of hours in both internal and external CVT courses. This positive correlation between learning hours and firm size is robust to controlling for industry, socioeconomic, and country-year fixed effects (see Table B.7), and is

These facts suggest that larger (and thus more productive) firms provide better learning environments by offering greater variety in learning options. This finding is consistent with evidence found by Engbom (2017) and Arellano-Bover (2020) showing that workers in more productive firms exhibit faster rates of skill acquisition, and with evidence found by Gregory (2019), showing that having different forms of training available is important for firms' learning environments.

3.3 Fact 2: Changes in the Sources of Learning across Workers' Lifecycles

Using the German BIBB and US NHES data, we document the importance of different sources of skill acquisition for workers, and how this importance changes across the lifecycle. We rely on both human capital accumulation variables and potential work experience variables to conduct our analysis. First, we construct a measure of "internal learning" that captures workers who have recently engaged in learning from colleagues or superiors. In the German data, "internal learning" indicates whether an individual has acquired the skills or knowledge necessary to complete the tasks in their current job through colleagues or superiors. In the US data, "internal learning" captures workers who reported receiving instruction or training from a co-worker or supervisor in their last work experience program, which is defined as a job with learning attributes.[11] Second, we construct a measure of "external learning" that captures workers who have recently engaged in learning from companies or instructor outside the current firm. In the German data, "external learning" indicates whether an individual received external on-the-job training in the previous 2–5 years, or acquired the skills/knowledge necessary to complete the tasks in their current job through external training or external firm knowledge. In the US data, "external learning" captures workers who reported taking classes or training from a company, association, union, or private instructor in their last work experience program, or ever earned a training certificate from an employment-related training program.[12]

also confirmed when we show the histograms of the share of working hours spent in each type of learning source for firms of each size (see Figure B.1).

[11] In particular, a work-experience program is defined as a job with learning attributes such as an internship, co-op, practicum, clerkship, externship, residency, clinical experience, apprenticeship, or other learning components. About 25% of the surveyed sample in the US reported having been part of such a program. In Figure C.7 we show that our results are robust to limiting only to individuals reporting participating in a work-experience program, and to decomposing across learning components that involve a "work-experience" program and those that do not.

[12] In Appendix A.2 and Appendix A.3 we provide further details on the questions and answers used for each one of the variables in the German and US surveys.

Based on these definitions, external learning focuses on the explicit role of knowledge pools outside the firm (such as teachers, instructors, specialist literature, etc) in the formation of human capital.[13] In contrast, internal learning captures learning that draws from knowledge internal to the firm. In addition, it is important to note that both of these learning variables generally capture flows and not stocks of learning investments, since they refer to skill acquisition in the current job or work-experience program (which is subject to change every few years as workers climb the career ladder or switch employers), or training incurred in the last few years.[14]

We construct our potential work experience variables using age, educational level and experience variables. Specifically, for both Germany and the US we construct potential years of experience as $Potential\ Experience = Age - Years\ of\ Schooling - 6$. We limit our sample to individuals who are currently employed, and have potential experience between 1 and 45 years, given that the number of observations outside this range is very limited. Finally, summary statistics, graphs and regressions are weighted using observation weights provided in the surveys. In Table A.1 we display some key summary statistics for the individuals in our samples. These statistics show that in both Germany and the US the proportions of individuals reporting each source of learning are sizeable, thus suggesting that both sources of learning are important to workers. 31% and 68% of workers report internal and external learning respectively in Germany, while these numbers are 23% and 44% respectively in the US.

We now document differential patterns in the prevalence of each source of worker learning along the lifecycle. In Figure 3.1, we plot how the prevalence of workers reporting engaging in internal and external learning changes with workers' potential experience in Germany and the US. We show that (1) the prevalence of internal learning decreases with worker potential experience; and (2) the prevalence of external learning has an inverted U-shape in

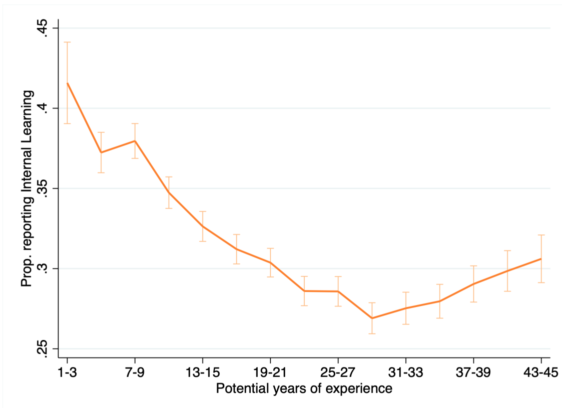
[13]Please note that formal schooling also fits this characterization of external learning, since it also draws from knowledge outside the firm. However, less than 10% of adult education corresponds to formal schooling in the EU, while over 90% corresponds to on-the-job learning (Ma et al. (2020)), making it much less important to understand adult human capital accumulation.

[14]The exception to this is the part of the external learning definition in the US which involves ever earning a training certificate from an employment-related training program, since this captures a stock of learning. Unfortunately, a comparable flow variable does not exist in this data. Nevertheless, it is worth noting that this variable is still informative of workers' flow of external learning, since a steeper increase in this variable along a specific portion of the lifecycle denotes larger positive flows at different ages. In addition, in Panel (b) of Figure C.7 we show that result showing an inverted U-shape of external learning in the lifecycle holds when we limit the external learning definition to exclude this part (and thus rely only on workers who reported taking classes from an external instructor or company in their last work experience program), though they are significantly noisier.

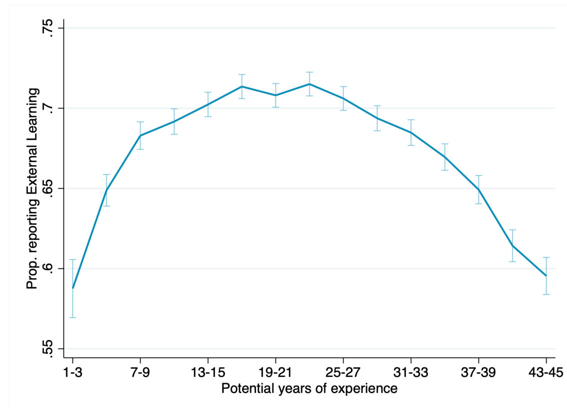
Figure 3.1: Prevalence of different sources of learning throughout workers' lifecycles in Germany and the US

Germany

(a) Internal Learning

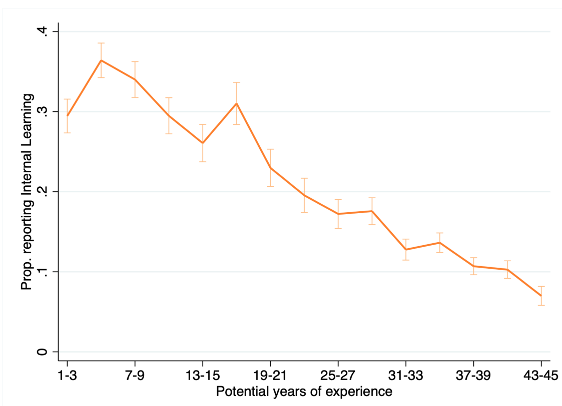


(b) External Learning

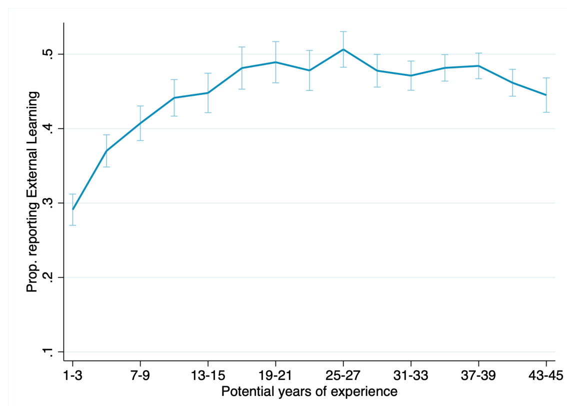


United States

(c) Internal Learning



(d) External Learning



Notes: 95% confidence intervals plotted.

worker potential experience. These patterns are robust to decomposing the data along many dimensions such as wave, cohort, gender, educational level, firm size among others.[15] In Table C.1 we show these lifecycle patterns are statistically significant even after controlling for several demographic and firm-level variables along with occupation and industry fixed effects in both settings.[16] Moreover, in Appendix C.4 we show that these patterns hold when

[15]Specifically, the patterns are robust to decomposing by one-year experience bins (Figure C.1), gender (Figure C.2), educational level (Figure C.3), survey wave in Germany (Figure C.4), cohort in Germany(Figure C.5), and firm size in Germany (Figure C.6). It is worth noting that similar to our findings in Section 3.2, workers in larger firms report a higher prevalence of both internal and external learning at all levels of potential experience.

[16]We also examine the correlation between internal and external sources of learning across workers in both

we use data from a developing country: Chile. This suggests that the lifecycle patterns we find are robust across income levels, and not unique to developed nations.[17]

In addition, we also show the results are robust to considering alternate working experience variables, namely age (see Figure C.8), and tenure (see Figure C.9). The first of these results is not surprising given the strong correlation between potential experience and age. The second result, however, suggests that these patterns are not solely a consequence of the aging process, but that working experience and human capital matter. This is further confirmed when we formally explore the correlations between current firm tenure and the sources of learning in Table C.2, given that the patterns of interest hold even when we include age fixed effects.

These lifecycle patterns hint at the differential nature of skill acquisition fostered by internal and external learning. One such dimension, which we explore in this paper, is the progressive decline in eligible coworkers that occurs with seniority and consequent increase in experience and human capital. In particular, as workers age and acquire human capital, the potential learning they can derive from their coworkers declines as the mass of colleagues who has more human capital than them, and thus that they can learn from, shrinks. In this sense, the patterns observed indicate an intuitive pattern where younger workers tend to acquire skills by observing their coworkers, while older workers, who know more than most of their coworkers, focus on learning from experts or simply working on their job.

This intuition and our findings are consistent with the evidence found by several papers suggesting that younger workers are more sensitive to peer learning than older workers (Azoulay et al. (2010), Nix (2017), Jarosch et al. (2019)). For example, Nix (2017) finds that learning spillovers within coworkers are largest for younger workers, with no impact for workers who are older than 40. Relatedly, Jarosch et al. (2019) find that the positive effects of peers' wages on future worker wages are substantially stronger for young workers. While focusing on academic networks, Azoulay et al. (2010) find that academics who had collaborators who died unexpectedly experience a decline in their quality-adjusted publication rates, and that this decline is relatively larger for young academics.

settings, and find a significant negative correlation in Germany, and a positive and significant correlation in the US. We present and discuss these results in detail in Appendix C.3.

[17] We do not include the results from Chile in the main text since the definitions of internal and external learning are much narrower from those in the US and German data. Please see Appendix C.4.1 for details on the Chilean data and variable construction.

4 Benchmark Model

In this section we develop a model that sheds light on the lifecycle results presented above by examining how the incentives to accumulate human capital through each source of learning evolve as workers age. With this, we also provide economic intuition on the mechanisms driving the results in the quantitative model of Section 5. The model features an overlapping generations structure a la Blanchard-Yaari, in which people have a finite but uncertain lifetime, and where expected remaining lifetime for any individual is independent of age. Agents stay on the labor market until death and have one unit of time which they can use to work or learn. There are two sectors in the economy: a final goods production sector, and a training (or external learning) sector providing training services. Workers produce final goods or training services by working in each of these sectors respectively, and accumulate human capital through internal or external learning. Both internal and external learning contribute to the same form of learning, and therefore capture two different modes of "general training". This means these learning investments are not specific to the current tasks performed by the worker, and can be used in other firms.[18]

4.1 Households

The model economy is populated by a unit mass of heterogeneous workers with human capital $h \in H$. Workers have a probability δ of dying each period, with $0 < \delta < 1$. Let X denote the time of death. For any period t occurring S periods after the current period, we have:

$$P(X > t) = (1 - \delta)^S.$$

The mortality rate parameter δ is assumed to be independent of age for simplicity, which

[18] An alternative to this would be to distinguish between general and firm-specific human capital, and potentially allow our two sources of learning to contribute differently to them. We focus on general human capital because as documented by Altonji and Shakotko (1987), Lazear (2009), and Kambourov and Manovskii (2009), among others, the truly firm-specific components of human capital are much less important for wage growth than the general component. In addition, we focus on the differences between learning sources arising from the pool of knowledge each of them taps into, rather than differences in the "transferrability" of this learning for two reasons. First, the organizational literature on workplace learning (see Manuti et al. (2015) for a review) suggests that both internal and external learning (often labeled formal and informal learning) are important dimensions of workplace learning, and both can contribute to forming new and transferrable skills/competences for workers. Second, although other differences between the two forms of learning may exist, our theory matches the empirical findings on the lifecycle of learning along with other key features of the data (see Section 4.5).

implies that the expected remaining lifetime is also independent of age. As such, age *per se* is not relevant for production and human capital accumulation decisions, but rather the level of human capital is. Each period, a mass δ of new workers is born. Newborns are homogeneous and have a human capital level of h_1. Workers aim to maximize their discounted lifetime income. Expected remaining lifetime utility in period τ for an individual with human capital level h is given by

$$U_\tau(h) = \sum_{t=\tau}^{\infty} \beta^{t-\tau}(1-\delta)^{t-\tau} E_t(I_t(h_t)),$$

where $I_t(h_t)$ represents the payout received by a worker with human capital h in period t.[19] Workers supply one unit of labor inelastically to the market in each period. Workers' human capital level h determines their labor productivity. This stock of human capital evolves throughout workers' lives in a logarithmic scale ladder fashion:

$$h \in \{h_1, h_2, ..., h_M\} \text{ with } \log(h_{m+1}) - \log(h_m) = \text{constant } \forall m.$$

Human capital accumulation, and climbing of the ladder is fostered by internal or external learning. Workers can work on the production sector, or the training sector. If they work in the former, they can choose to spend their time either working, in internal learning, or in external learning. If they instead work in the training sector, they spend all their time working and cannot go back to the production sector.

4.1.1 Learning and Skill Acquisition

Both internal and external learning yield stochastic movements along the human capital ladder. The success probabilities of internal and external learning are distinct, and characterized by $p_c(h)$ and $p_s(h)$, which depend on the workers' human capital level, and the matching probability with colleagues and trainers respectively. We assume that external learners (or trainees) match randomly with trainers in a one-to-one fashion.[20] We assume that in equilibrium the number of external learners N_s must equal the number of trainers N_n, and thus the matching probability of external learning is equal to one. In addition, we assume that individuals who internally learn simply observe their colleagues as they work, and learn in the process. This captures the idea that for example, workers who are learning

[19] Please see Appendix D.1 for details on this expected lifetime utility.

[20] This assumption allows us to get simpler analytical results. We relax it in our quantitative model of Section 5, where we let trainers teach several individuals simultaneously.

are brought into projects with more senior colleagues so they can observe and learn. As such, production workers face no cost of having colleagues learn from them, and can have several of these doing so at the same time. This implies the matching probability between colleagues and workers is also equal to one. We denote the size of production workers and individuals engaging in internal learning as N_l and N_c respectively.

We further assume that workers learn only from coworkers and trainers with a human capital level higher than their own. This matches up with the findings of Herkenhoff et al. (2018), who show that workers learn from more knowledgeable coworkers and not less knowledgeable ones, along with the findings of Jarosch et al. (2019) who document that peers higher up in the team wage distribution matter far more for workers' future wage outcomes than the peers below. Therefore, the probability a worker with human capital level h_i climbs the human capital ladder depends on $(1 - F_l(h_i))$ when learning internally and on $1 - F_n(h_i))$ when learning externally, where F_l and F_n denote the cumulative distributions of workers actively producing in the firm and trainers in the training sector across the human capital ladder, respectively. Similarly, we denote the cumulative distributions of workers learning internally and externally across the human capital ladder as F_c and F_s, respectively.

We also introduce an exogenous probability ϵ of climbing the human capital ladder when learning, engaging in production work or working as a trainer, which resembles other forms of learning such as learning-by-doing. This probability ensures that regardless of the starting distribution of human capital, including ones where all workers are concentrated at the lower levels of the ladder, some workers will eventually reach higher levels of human capital, thus ensuring learning possibilities and a transition to the stationary equilibrium. We assume ϵ is very small, however, to focus on other forms of learning.[21] Therefore, the probability a worker with human capital level h_i climbs the human capital ladder is $p_c(h_i) = (1 - F_l(h_i)) + \epsilon F_l(h_i)$ when learning internally, $p_s(h_i) = (1 - F_n(h_i)) + \epsilon F_n(h_i)$ when learning externally, and $p_l(h_i) = p_t(h_i) = \epsilon$ when engaging in production work or working as a trainer.

When learning internally or externally, the worker faces foregone production, but in the latter, the worker must pay a price q for the purchase of training services. This price is only paid if external learning is successful in generating an increase in human capital, however. As such, the payment to a trainer with human capital of h is stochastic. We assume further that there is no cost to the colleague, resembling the fact that the worker observes and learns

[21] In the quantitative model we allow this ϵ to be higher, and calibrate it to capture human capital accumulation and wage growth stemming from other sources of learning.

from its colleagues while they produce to no added cost to them.

4.1.2 Workers' Expected Present Value of Earnings

The expected present value of earnings for an agent with human capital h_i in any given period is given by[22]

$$EV(h_i) = \max_{l_i, s_i, c_i, n_i} EV_i,$$

where

$$EV_i = \begin{cases} w(h_i) + \beta(1-\delta)\left[(1-\epsilon)EV(h_m) + \epsilon EV(h_{m+1})\right] & \text{if } l = 1 \\ 0 + \beta(1-\delta)\left[p_c(h_i)EV(h_{m+1}) + (1-p_c(h_i))EV(h_m)\right] & \text{if } c = 1 \\ -p_s(h_i)q + \beta(1-\delta)\left[p_s(h_i)EV(h_{m+1}) + (1-p_s(h_i))EV(h_m)\right] & \text{if } s = 1 \\ E(w_n(h_i)) + \beta(1-\delta)\left[(1-\epsilon)EV(h_m) + \epsilon EV(h_{m+1})\right] & \text{if } n = 1. \end{cases}$$

In every period, workers choose whether to work, learn internally or externally, or to become trainers. l denotes the decision to engage in work in the production sector, c denotes the decision to engage in internal learning, s denotes the decision to engage in external learning, and n denotes the decision to work in the training sector. $w(h_i)$ represents the production sector wage paid to workers with human capital level h_i, and $E(w_n(h_i))$ represents the expected training income received by trainers with human capital level h_i.

4.2 Production

There are two sectors in this economy: a production sector, which produces final consumption goods, and a training (or external learning) sector, which produces training services.

4.2.1 Production Sector

There is a large number of identical production firms, which use labor from workers to produce output. Firms choose the vector of effective human capital of workers they employ, which is denoted by H^d. Let $W(H^d)$ be the total wage bill of a firm that hires the vector of workers H^d.

A firm chooses the set of workers H^d to maximize profit:

[22] Appendix D.2 contains a full description of the worker's problem.

$$\pi = \max_{H^d} y(H^d) - W(H^d),$$

where $y(.)$ is a production function that transforms H^d into output. Workers of each human capital level are paid their marginal products, so that:

$$w(h_m) = \frac{dy}{dN_{l,m}},$$

where $N_{l,m}$ is the mass of workers of type m who are actively working and producing in the firm. As such, $H_{l,m} = h_m N_{l,m}$ is the human capital input of each type.

Note here that firms are trivial. The firm does not participate or care about workers' learning decisions, since only workers who are actively producing are paid, and learners pay for their own learning expenses in full. We abstract from firm decisions in order to focus on the tradeoff between different sources of learning. In the quantitative model of Section 5 we let firms and workers jointly decide and pay for skill acquisition. In that section, we also show that firms and workers agree on the division of learning between the two sources. This further motivates the simplification here, since it indicates the tradeoff between internal and external sources of learning can be fully captured by worker decisions only.

4.2.2 Training Sector

We assume that there is a training technology producing training services that can be operated by any of the workers. These training services are provided to production workers engaging in external learning. Workers who decide to engage in external learning randomly meet trainers, and after observing the trainer's human capital decide to engage in training or not. As such, the expected payout to trainers will depend on the price of training services, which is equal to q, and the probability they get hired, which depends on their human capital and is given by $p_n(h)$:

$$E(w_n(h_i)) = p_n(h_i)q.$$

We assume that trainers can only effectively train the external learner, and thus get hired, if their human capital exceeds theirs. Therefore, we have that $p_n(h_m) = F_s(h_{m-1})$.

4.3 Characterizing the Equilibrium: Working and Climbing the Human Capital Ladder

In Appendix D.3 we define the stationary equilibrium in this model. In Appendix D.4, we provide one set of conditions that are sufficient to give rise to results consistent with the empirical facts presented. Assumption 1 imposes some structure on the production function, and implies that the wage at each human capital step tends to infinity as the amount of labor goes to zero. Assumption 2 and Assumption 3 impose some structure on the exogenous probability of learning and the human capital ladder, respectively.

There are two important issues to note here. First, there are two types of equilibria possible in this model: equilibria without external learning, and equilibria with external learning. The first type of equilibrium comes about because if we assume that there are no external learners at any human capital level, there will be no incentive for individuals to work as trainers since the payoff of this will always be zero. This in turn confirms the fact that there are no external learners, since the returns from external learning will be null. The second type of equilibrium is one where there are both trainers and external learners. In what follows, we focus in the second type of equilibrium, which is captured in Assumption 4. Second, there exist a multiplicity of equilibria within the external learning equilibrium. This arises because the dimensions across which external learning is characterized encompass both the number of trainers and external learners, and the location of these across the human capital ladder. Since both of these will depend on an unique object, the price of training q, multiple equilibria arise.[23] As such, in order to fully characterize external learning in this model, we must make further assumptions about the location of trainers in the human-capital ladder. This is captured in Assumption 5, which indicates that trainers locate at all levels of human capital beyond a certain step.[24,25] These conditions are sufficient to deliver the following results, which characterize the learning decisions and lifecycle of workers in equilibrium.[26]

[23]For example, we could have trainers located only at the mid-point of the human capital ladder, $h_{\frac{M}{2}}$, thus confining external learners to the first half of the human capital ladder; or we could have trainers located only at the last step of the human capital ladder, h_M, thus allowing external learners to exist throughout the human capital ladder.

[24]We test this formally in Section 4.5.1.

[25]Another issue to note is that the equilibrium of this model is inefficient because workers underinvest in human capital relative to the social optimum. This arises because workers do not internalize the fact that by increasing their own human capital, they also increase the opportunities for other workers by raising the probability they can learn. This creates a scope for policies to incentivize human capital accumulation. We study the consequences of such policies within the context of our quantitative model in Appendix G.3.

[26]In Appendix D.5, we present additional results that characterize the equilibrium in this model.

Proposition 1. ***Learning and Working Decisions across the Human Capital State Space***

1. *There is a unique threshold in human capital, h_{m*} below which individuals learn internally, and above which individuals learn externally: $\forall h_m \in [h_1, h_{m*-1}],\ f_c(h_m) > 0$ and $f_s(h_m) = 0$; and $\forall h_m \in [h_{m*}, h_{M-1}],\ f_c(h_m) = 0$ and $f_s(h_m) > 0$.*

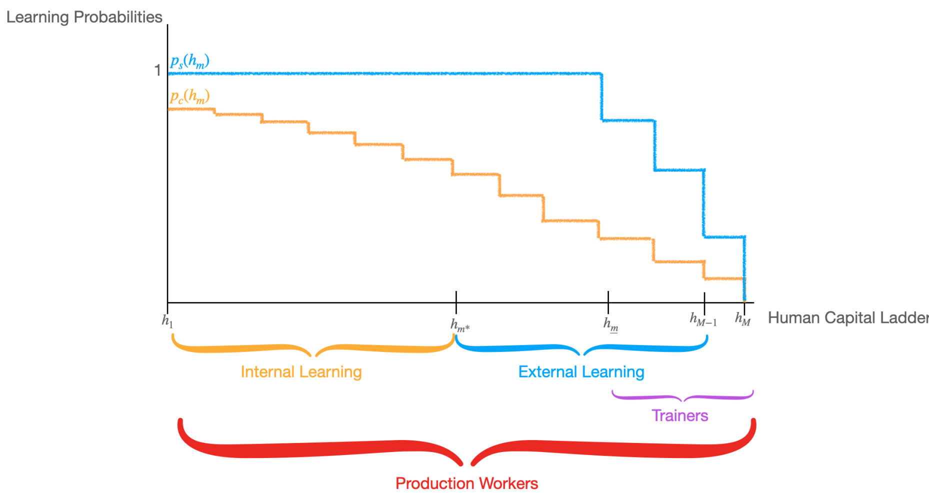
2. *There is a unique threshold in human capital, $h_{\underline{m}} > h_{m*}$ above which a positive mass of individuals work as trainers: $\forall h_m \in [h_{\underline{M}}, h_M],\ f_t(h_m) > 0$.*

Proof: See Appendix D.6.8.

This proposition characterizes the learning and working decisions across the human capital distribution. The existence of the threshold h_{m*} determining the point at which workers switch to external learning from internal learning highlights the tradeoff between the probability and cost of learning in our setup. In particular, since external learning carries an explicit cost while internal learning does not, for low human capital workers the cost of external learning is too high relative to the learning probability gain it provides. Therefore, low human capital workers acquire skills via internal learning. However, the probability of experiencing a boost in human capital from internal learning declines faster with human capital than that of external learning since trainers tend to have higher average human capital levels than production workers. This is captured by the second part of this proposition on the existence of $h_{\underline{m}} > h_{m*}$ determining the lower bar of trainer human capital, and arising due to the lack of external learners and thus training production opportunities at low levels of human capital. This threshold implies that from h_1 to h_{m*}, the probability external learning remains at 1, while the probability of internal learning declines progressively as a positive mass of production workers places at every human capital level. This makes external learning progressively more appealing than internal learning as human capital accumulates, because even though the former carries a larger cost, the probability the latter will result an increase in human capital decreases faster with human capital.

These dynamics are depicted in Figure 4.1. Individuals with the lowest level of human capital engage in either production work or internal learning. This pattern continues as we climb the human capital ladder until we reach h_{m*}, the point at which the accumulated mass of production workers is high enough so that the probability of internal learning $p_c(h_m)$ dips low enough to make it relatively more profitable to pay the cost q and learn externally with a probability $p_s(h_m)$ of 1. As we further climb the human capital ladder $p_c(h_m)$ continues to dip as we keep accumulating more production workers in each human capital step, while $p_s(h_m)$ remains at one, so that still external learning is more profitable than internal learning. Eventually, the accumulated mass of external learners is large enough so that the expected payout for trainers equals the production wage. From this point on and up to $M-1$, we will have a positive mass of trainers, external learners and production workers. Then, at the final human capital level M we have only production workers and trainers.

Figure 4.1: Learning and Working Cycle across the Human Capital Ladder



We now present an additional result that further helps characterize the evolution of the mass of workers engaging in production work throughout the human capital state-space.

Corollary 1. *The portion of production workers within each human capital level rises from* h_1 *to* h_{m*-1}.

Proof: See Appendix D.6.9.

This result indicates that the portion of production workers rises with human capital in the initial portion of the human capital ladder where internal learning occurs. This highlights the tradeoff between the probability of learning and its opportunity cost, and stems from the fact that in our framework effective units of human capital rise sufficiently fast with each step in the human capital ladder in order to support both a rise in wages and the portion of production workers through the human capital state space. In particular, this fast increase in the effective units of human capital disproportionately raises the value of production work, making the opportunity cost of learning higher at each level. Consequently, this result implies that the portion of individuals who learn internally declines as human capital rises, which will be an important feature to match the empirical facts documented before, and particularly the fact that the portion of workers learning internally declines with potential experience.

4.4 Lifecycle of Working and Learning

Armed with these results, we can now fully characterize the lifecycle of working and learning in this economy. The Blanchard-Yaari "perpetual youth" structure implies that the only force driving the work and learning decisions of individuals is the human capital level. This therefore implies that the distribution of learning and working decisions across workers of each age follow directly and solely from their corresponding distribution across the human capital state-space. In addition, given that there is no depreciation of human capital in this economy, the average human capital level of workers rises with age. As such, the evolution of work and learning decisions across the lifecycle follows the same forces as climbing the human capital ladder. In the following result, we formalize the lifecycle evolution of internal and external learning in this model, which matches our empirical results.

Proposition 2. ***Lifecycle of Internal and External Learning***

1. *The portion of production workers engaging in internal learning declines with age.*
2. *The portion of production workers engaging in external learning first rises, and then declines with age.*

Proof: See Appendix D.6.10.

The first part of this result follows from Proposition 1 and Corollary 1. At lower levels of human capital and thus at younger ages, workers' learning mode of choice is internal learning. This follows from Proposition 1, and the fact that when human capital is low the mass of coworkers with a higher human capital than the own, and therefore the probability of climbing the human capital ladder, is relatively high. However, as workers continue to age and accumulate human capital two things begin to happen. First, the portion of workers engaging in production rises given that the opportunity cost of learning rises as captured in Corollary 1. Second, the portion of workers engaging in external learning increases, as progressively more workers reach human capital level h_{m*}, where the probability of internal learning is sufficiently low to make it relatively more profitable to pay the training cost and learn externally. These two forces drive the portion of production workers engaging in internal learning to decline with age.

The second part of this result follows from Proposition 1. As workers age and human capital accumulates, the portion of external learners begins to rise as progressively more workers reach human capital level h_{m*}, where the probability of internal learning is sufficiently low to make it relatively more profitable to pay the training cost and learn through external learning. However, as human capital continues to accumulate when workers age, the portion of external learners begins to decline as progressively more workers reach the last human capital level h_M, where there is no learning.[27]

This result paints a clear picture of lifecycle learning. Initially, when workers are young and have a low level of human capital, they join the production sector and face (1) a large contingent of coworkers with larger human capital than the own, making the probability of internal learning high; and (2) a low opportunity cost of working, since productivity is low. These factors lead a large portion of young workers to engage in internal learning, and a small remaining portion to engage in production work. As workers start to age and average human capital rises, however, the average opportunity cost of learning rises, leading to a rise in the portion of workers engaging in work, and consequently a decline in the portion of workers engaging in internal learning. As human capital continues to rise with aging, the contingent of coworkers with larger human capital than the own shrinks, reducing the probability individuals can learn from coworkers. This leads a progressively larger portion of workers that engage in external learning, which incurs a cost, but involves matching with a better pool of workers, thus increasing the probability of climbing the human capital ladder. Eventually, however, this rise in the portion of external learners reverses as workers progressively reach the highest level of human capital, and thus engage solely in production work and training work.

4.5 Evidence on Testable Predictions

Our benchmark model yields a series of testable predictions we can examine in the data. In particular, there are four key testable predictions we show support for in this section. The first one is a structural prediction, matching important theoretical elements of our framework. The second and third correspond to learning predictions, matching the key learning results of the benchmark model. The last one corresponds to a direct implication of our theory.

[27] Note that the portion of workers engaging in external learning could decline even prior to workers reaching human capital level h_M if the portion of production workers and trainers rises with human capital in between h_{m*} and h_M. However, this is not necessary for our result since a larger contingent of individuals reaching human capital level h_M will yield a decline in the portion of external learners.

4.5.1 Distribution of Trainers across Human Capital Ladder is Skewed Left

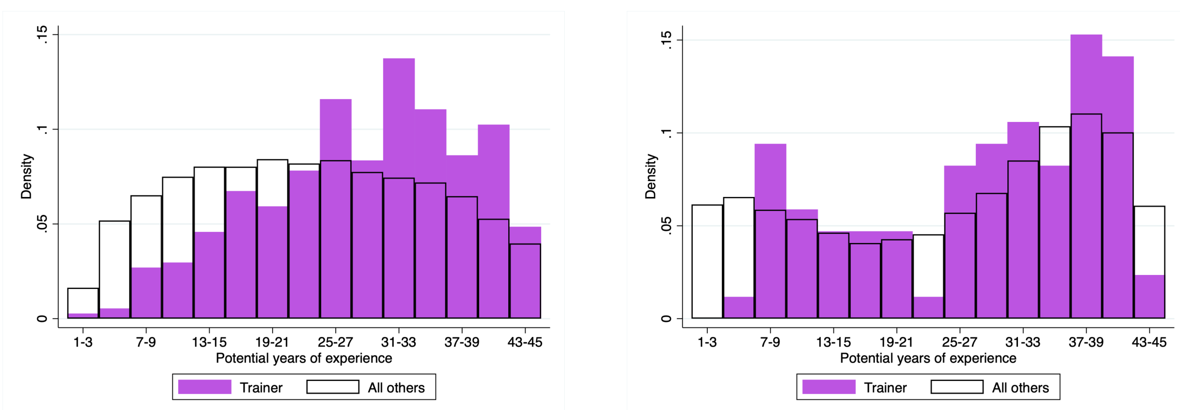
Our benchmark model relies on two key results to characterize the distribution of trainers across the human capital state space. First, our theory implies that the human capital distribution of trainers has a a higher starting point, median and mean relative to the distribution of production workers.[28] Second, we assume in Assumption 5 that trainers locate at all levels of human capital beyond this starting point. We provide support for these modeling elements in Figure 4.2, where we plot the histograms of trainers and production workers in both Germany and the US by potential experience.[29]

The plots show that the distributions of trainers in both Germany and the US span across all human capital levels, but heavily concentrate among higher levels of potential experience relative to other workers.[30] In Table E.1 we present the results of quantile regressions at the first, second and third quartiles of potential years of experience on the trainer variable (where the omitted category is production worker). The results from these regressions indicate that the 25th, 50th and 75th percentiles of potential years of experience for trainers are generally

Figure 4.2: Histograms of potential experience for trainers and production workers

(a) Germany

(b) United States



[28] This is captured formally in Lemma 7.

[29] We define trainers as workers who report an occupation that involves training, teaching or instruction activities outside of school and university education. Production workers on the other hand are captured by all other workers outside of trainers, though the results are analogous if we solely focus on workers with professional and technical occupations outside of trainers (see Figure E.2 and Table E.3). In Appendix E.1.1 we provide further details on the construction of the trainer and production worker variables in the German and US surveys.

[30] In Figure E.1 and Table E.2, we compare the distribution of trainers to the distribution of external learners across these variables, showing similar patterns.

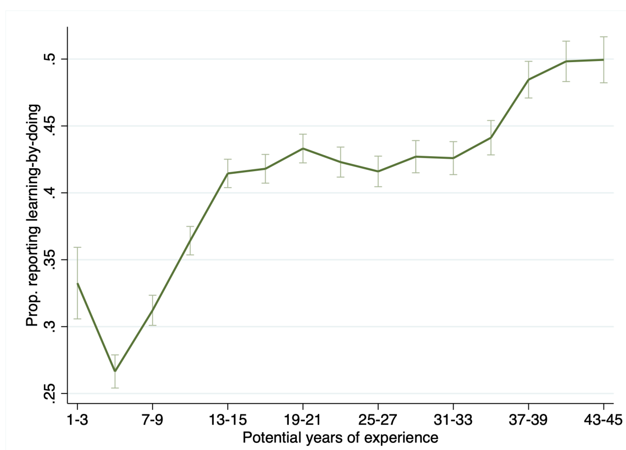
larger (though not always statistically significant) than that of production workers in both settings, even after controls.

4.5.2 Portion of Workers who do not Engage in Explicit Learning Rises with Human Capital

One key prediction of our benchmark model is that the portion of workers who do not learn from either of the two sources rises with human capital. We provide evidence for this prediction using our German data. We construct a measure of "Learning-by-Doing" which captures individuals who did not invest in explicit forms of learning to acquire skills for their job, but rather acquired the necessary professional skills by doing the job itself.[31] In Figure 4.3 we plot how the prevalence of workers reporting learning-by-doing changes with potential experience. We find that the prevalence of this generally increases with workers' potential experience throughout the whole lifecycle.[32] In Table E.4 we show that the positive correlation between experience and learning-by-doing is statistically significant even after controlling for several demographic and firm-level variables.

Figure 4.3: Prevalence of Learning-by-Doing throughout workers' lifecycles in Germany

Germany



Notes: 95% confidence intervals plotted.

[31]Please see Appendix E.2.1 to see details on the contruction of this variable.

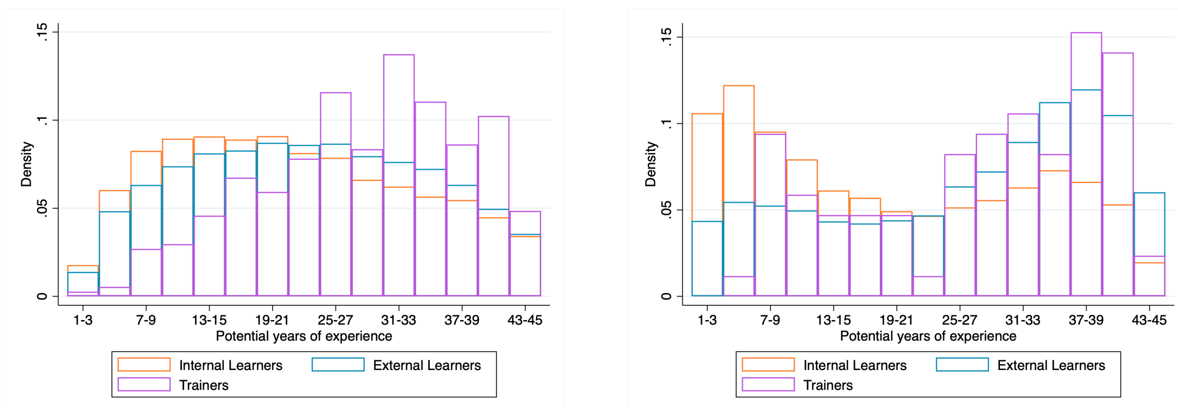
[32]This result is robust to decomposing the data by one-year experience bins (Figure E.3a), gender (Figure E.3b), educational level (Figure E.3c), survey wave (Figure E.3d), and firm size (Figure E.3e); and considering the age (Figure E.3f), or number of years with current employer as an alternate working experience variable (Figure E.3g). For the latter, however, the portion of employees with no explicit learning investments only increases sharply towards the end of the lifecycle.

4.5.3 Human Capital Ranking across Different Types of Workers

Our benchmark model predicts that different types of workers locate in different areas of the human capital ladder. As suggested by Proposition 1 and depicted in Figure 4.1, individuals who engage in internal learning concentrate in the lower part of the human capital distribution, while external learners and trainers concentrate in the middle and higher parts, respectively. We provide evidence for this prediction using our German and US data. We plot the histograms of individuals who report engaging in internal or external learning, or being trainers in both Germany and the US by potential experience in Figure 4.4.[33]

Figure 4.4: Histograms of potential experience for each type of worker

(a) Germany (b) USA



The plots show that the distributions of trainers in both Germany and the US heavily concentrate among higher levels of potential experience relative to both external and internal learners. Among these, the distribution of internal learners is particularly heavily concentrated among lower levels of potential experience, while the distribution of external learners is more evenly distributed. In Table E.5 we formally test these differences through quantile regressions at the first, second and third quartiles of potential years of experience on the external learning and trainer variables (where the omitted category is internal learning). The results from these regressions indicate that the 25th and 50th percentiles of potential years of experience for trainers and external learners are generally larger than that of internal learners in both settings, particularly for trainers. However, external learners appear to have lower 75th percentile levels than internal learners in Germany.[34]

[33] We define individuals who report internal and external learning as in Section 3; see Appendix A.2 and Appendix A.3 for details. In addition, we define trainers as in Section 4.5.1; see Appendix E.1.1 for details.

[34] In Appendix E.3.1 we provide an additional test of human capital differences between internal and external

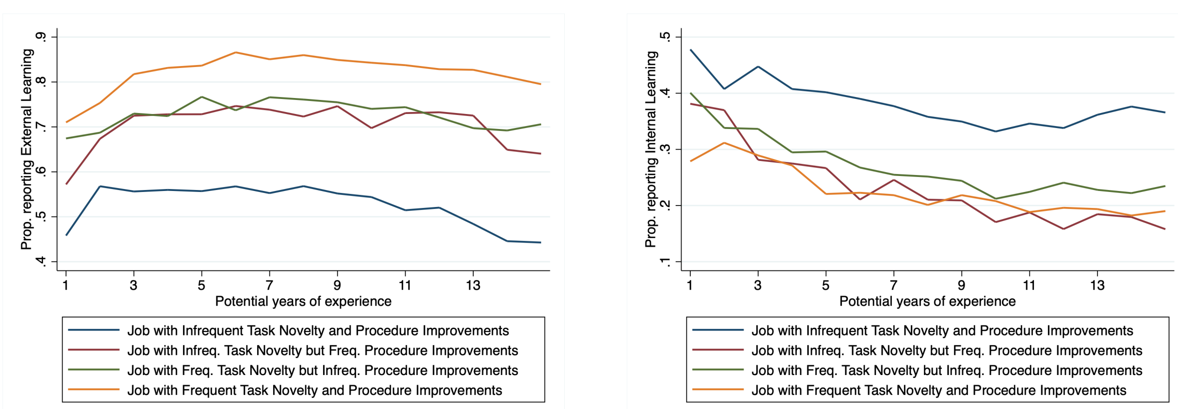
4.5.4 Workers who Use Innovative Techniques Rely More on External Learning

We now present evidence supporting a natural implication of our theory: workers whose jobs require the use of innovative techniques or tasks are better served by learning from external sources that tap into knowledge that is not currently available in the firm. We provide evidence for these predictions using our German data. We rely on information available on all seven waves of the survey regarding the frequency with which workers have to adapt to new situations and try new procedures at their jobs. Using this data, we build two measures of "work-related novelty", which capture respectively whether a worker reports always or frequently being (1) faced with new tasks she has to familiarize herself (Job with Frequent Task Novelty); or (2) having to improve previous procedures or try something new (Job with Frequent Procedure Improvements).[35]

In Panel (a) of Figure 4.5 we show that workers who report having jobs with both frequent task novelty and procedure improvements have the highest rates of external learning, while workers who report infrequent task novelty and procedure improvements have the lowest rates of external learning. In Panel (b) we show that the opposite is true for internal learning. We show the existence of this correlation more formally by regressing the external and internal learning variables on the two measures of "work-related novelty". We document

Figure 4.5: Prevalence of Internal and External Learning by "Work-related Novelty"

(a) External Learning | (b) Internal Learning



learners by exploring differences on the skill-content of tasks performed and tools used for these tasks by each of these two groups of workers. We find that internals learners exhibit lower levels of task complexity than external learners; and that external learners are more likely to use "white-collar" tools than internal learners, while the opposite is true for "blue-collar" tools.

[35]Please see Appendix E.4.1 for details on the construction of these measures.

our findings in Table E.8, and show that workers with jobs with either task novelty or procedure improvements are more likely to report engaging in external learning, and less likely to report internal learning.

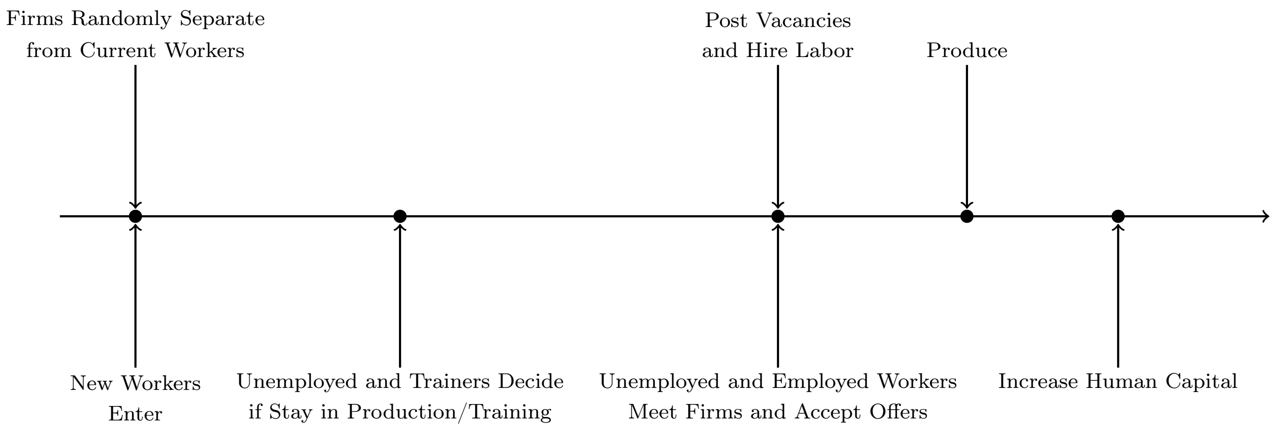
5 Quantitative Analysis

In order to quantify the importance of internal and external learning to lifecycle human capital accumulation and wage growth, we now embed our two-source learning mechanism in a quantitative framework. We consider a search structure that explicitly considers the rich interactions between firms and workers in learning investments. With this, we account for the key roles played by both firms and workers on on-the-job learning, and consider the incentives faced by each. We calibrate the model to the US economy. We examine the equilibrium properties of our model, and perform counterfactual exercises that measure the importance of internal and external learning on lifecycle human capital accumulation, wage growth, and wage dispersion.

5.1 Quantitative Model Setup

We embed our analytical model into a Burdett-Mortensen framework with labor market frictions and joint learning decision-making by workers and firms. This framework allows us to consider the different incentives and rich interactions of firms and workers in learning investments, and is motivated by research showing that employers and employees jointly choose on-the-job learning investments, and that labor market frictions are a key driver of these choices (Acemoglu (1997), Acemoglu and Pischke (1999), Moen and Rosén (2004)). As in the analytical model, we assume that workers are endowed with one unit of time per period, have a probability δ of dying each period, and aim to maximize their discounted lifetime income. Each period, a mass of new workers is born to replace the workers who died. Newborns are homogeneous, and have a human capital level of h_1. Individuals accumulate human capital throughout their lives by climbing a human capital ladder, and can engage either in the production sector or the training sector in every period. We assume that the training sector is frictionless, while the production sector is characterized by labor market frictions and firm heterogeneity a la Burdett and Mortensen (1998). Firms in the production sector post vacancies and wage rates per efficiency unit of labor to attract both unemployed individuals and workers from other firms, and meet their matches by random search. After matching, workers and firms in the production sector jointly decide and pay for internal and external learning investments. We assume that workers can divide their time between production, and learning from each source in every period, and that the probability of climbing the human capital ladder depends on the time spent on each source of learning, and the likelihood of finding a colleague or trainer with higher human capital than the own. Workers' human capital can also depreciate, driving workers to descend the human capital ladder. The timing of the model is presented in Figure 5.1 and will be described in more detail in the following sections.

Figure 5.1: Timing of Events in Each Period



Before turning our attention to the quantitative model's details, it is worth highlighting a few key assumptions that enable us to keep the model tractable. First, we assume that each firm posts a unique wage per efficiency unit. In models where differences in human capital matter only linearly for production, posting single piece-rate wages is optimal for firms because they are indifferent between hiring two workers with low human capital, or one worker with high human capital. In our setup, however, because workers with higher human capital change the learning landscape, firms would like to post different wage rates to attract workers of different human capital levels. However, this problem is computationally intractable since it requires directed search in the labor market. Thus, in order to keep the model tractable, we assume that firms post a unique wage per efficiency unit when posting a vacancy. In addition, we assume that the value of unemployment is low enough at all levels of human capital such that all unemployed will take any job offer.

In addition, we assume that when posting this wage rate firms do not reveal their workers' human capital distribution. This precludes the possibility that firms with the same productivity level choose different wage and skill distribution combinations to attract workers.

Moreover, we assume that there are no endogenous mergers and acquisitions (further restricting the possibility that existing firms create new ones to follow different strategies), and that workers cannot pay for training by taking a lower wage after matching with a firm. Finally, we focus on the model's stationary equilibrium where firms' human capital distributions are stationary. Taken jointly, these assumptions imply that firms with the same level of productivity have the same wage rate and human capital distribution of workers. Thus, due to the supermodularity of the production function, and as explained in more detail below, firms with higher levels of productivity unequivocally have both higher wage rates and more skilled workers, implying that the wage rate posted is a sufficient statistic for workers when evaluating job offers.

5.1.1 Production Sector and Frictional Labor Market

As explained before, we consider that the production sector is characterized by frictional labor markets and heterogenous firms that post vacancies and wage rates per efficiency unit of labor to attract workers who they meet via random search (Burdett and Mortensen, 1998). We have a measure $\bar{M}$ of production firms which produce a homogeneous good and differ in their productivity level z, which follows a Frechet distribution: $z \sim H(z) = \exp(-z^{-\kappa})$. Firms post vacancies $v(z)$ at the start of each period, with a contract stipulating the wage rate per efficiency unit of labor $w(z)$. The vacancy cost is given by $\frac{c_v v(z)^{1+\gamma_v}}{1+\gamma_v}$ and is assumed to be convex in v (i.e. $\gamma_v > 0$) to ensure that firms with different productivity levels coexist. The total number of vacancies is then $V = \bar{M} \int v(z) dH(z)$, and the wage offer distribution is described by $F(w(z)) = \int_{z_{min}}^{z} v(z') dH(z')/V$.[36]

At the beginning of each period, workers' contracts are destroyed exogenously with a probability δ_{job}, and new workers are born. These exogenously laid-off workers and newly born workers both enter the unemployment pool. Before job search happens, these unemployed individuals choose whether to look for a job in the production sector, or to switch to the training sector. Similarly, trainers choose whether to continue working in the training sector or switch back to the production sector and look for a job jointly with the other unemployed. Moreover, a portion η of employed production workers search for new jobs while on the job, and switch firms if they match with a new firm that offers a wage rate exceeding their current one.[37] We denote the total number of job searchers as $\tilde{U}$, which includes the unemployed and

[36] This wage rate offer distribution uses the result that $w(z)$ is increasing in productivity z as shown in Appendix F.1 and in Burdett and Mortensen (1998).

[37] Please note that on-the-job searchers and unemployed workers do not explicitly compare firms' learning

on-the-job searchers. The matching function between vacancies and searchers is $c_M V^{1-\phi}\tilde{U}^{\phi}$. The market tightness is defined as $\theta = \frac{V}{\tilde{U}}$, with $q(\theta) = \frac{c_M V^{1-\phi}\tilde{U}^{\phi}}{V}$ denoting the contact rate for firms and $\theta q(\theta)$ capturing the contact rate for searchers.

Once workers and firms are matched, worker i's production in firm j is given by

$$y_{ji} = z_j h_i.$$

Thus, the production function is supermodular, a firm with higher productivity generates more revenue per unit of labor, and human capital and firm productivity are complements as in Acemoglu and Pischke (1998) and Bagger et al. (2014). Vacancies and wage rates are determined by firms' first-order conditions that trade off the benefits (lower leaving rates of workers and higher chances of poaching on-the-job searchers) and costs (lower profits per efficiency unit of remaining labor) of high wage rates, combined with the min-mean wage ratio b (boundary condition).[38] Please see Appendix F.1 for details.

5.1.2 Training Sector

We assume the training sector is frictionless, and thus that unemployed workers can freely choose to switch to this sector and operate the training technology. We consider that the amount of training services provided by a trainer is proportional to her human capital level, given that high-skill individuals can typically teach several students simultaneously and their returns in the production sector are also proportional to human capital levels. The expected payout of a trainer with human capital h_i is thus $h_i p_n(h_i) q$, where q is the price of training services and $p_n(h_i)$ denotes the probability of matching with an external learner with lower human capital than the own. This probability is given by the cumulative distribution of external learners at h_{i-1}: $p_n(h_i) = F_s(h_{i-1})$.

environments when accepting offers and only consider wage rate differentials since higher wage rate (or more productive) firms also offer larger learning opportunities in the stationary equilibrium as shown in Section 5.3.2. Thus, the wage rate posted is a sufficient statistic for workers when evaluating job offers.

[38] As shown by Hornstein et al. (2011), search and matching models with reasonable unemployment benefits have difficulty in generating the amount of frictional wage dispersion present in the data. Thus, because our focus is on learning decisions, we choose to match the frictional wage dispersion by assuming the lowest wage rate to be $w_{\min} = b\bar{w}$, where $\bar{w}$ denotes the average wage rate and b is a constant. We assume that the unemployed will take any job offer, which can be rationalized by low, often negative, values of unemployment benefits. This assumption matches empirical findings of the offer acceptance rate being close to one (van den Berg (1990)). Because under these assumptions unemployment benefits do not affect any other equilibrium outcomes, we abstract from unemployment benefits in the model.

5.1.3 Joint Decision of Learning

Firms and workers in the production sector jointly decide learning investments after matching. Since workers typically engage in both learning and production at their jobs, we consider that the time allocation is divisible, and thus workers can spend time on both modes of learning and production in each period. Specifically, we assume that the worker and the firm jointly choose the overall learning time g and the portion of the learning time spent on internal and external learning, g_c and $1 - g_c$. The per-period probability by which a worker of human capital h_i in firm z moves up the human capital ladder is given by

$$p_{learn}(h_i, z) = \min\left(\left[(A_c p_c(h_i, z) g_c)^{\frac{\sigma-1}{\sigma}} + (A_s p_s(h_i)(1-g_c))^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} g^{\gamma} + \epsilon, 1\right). \qquad (1)$$

$p_c(h_i, z)$ and $p_s(h_i)$ denote the probabilities of matching with a higher-human-capital worker or trainer, respectively, and are thus given by $p_c(h_i, z) = 1 - F_l(h_i, z)$ where $F_l(., z)$ captures the cdf of workers at firm z, and $p_s(h_i) = 1 - F_n(h_i)$, where $F_n(.)$ captures the cdf of trainers in the training sector. A_c and A_s capture workers' cognitive ability to learn internally and externally, respectively. $\sigma > 1$ is the elasticity of substitution between the two modes of learning, which allows for imperfect substitutability.[39] $0 < \gamma < 1$ captures the degree of diminishing returns of learning time, which ensures workers spend time on both production and learning. ϵ is the exogenous probability of climbing the human capital ladder, akin to learning-by-doing. We consider that a worker with human capital level h_i has a probability $\frac{\delta_h(h_i - h_1)}{h_i - h_{i-1}}$ of descending the human capital ladder by one step, where δ_h is the depreciation rate of human capital accumulated from learning. This depreciation function implies that human capital cannot descend beyond h_1, and that the probability of descending the human capital ladder increases with human capital.[40] h_1 captures the basic physical and cognitive abilities inherent to workers and which do not depreciate, and similar Blundell et al. (2021), we assume that the human capital that depreciates is that acquired via learning.

Firms pay a fraction μ of total learning costs, while workers pay the rest.[41] The costs of internal learning correspond solely to foregone production, with no cost to the colleague the worker learns from. δ_s denotes the the decrease in the time spent in production when workers spend one unit of time learning. External learning also faces foregone production, but

[39] This specification nests our analytical model as a special case with $\sigma \to \infty$.

[40] This is analogous to having a constant rate of depreciation, where at higher levels of human capital workers need to spend more time learning to fully make-up for human capital lost to depreciation.

[41] We assume that this cost shares are fixed, and workers cannot pay for training by renegotiating with the firm to take a lower wage rate.

requires an additional payment q to the trainer for a successful match, which is endogenously determined by equalling supply and demand of training services.

We assume that if $g^W(h_i)$ and $g^F(h_i)$ are the optimal overall learning times from the worker's and the firm's perspectives, respectively, the overall learning time g will be given by $g(h_i) = min\{g^W(h_i), g^F(h_i)\}$. This assumption implies that the overall time spent on learning is determined by the party with lower affordability. For instance, if firms bear all of the training costs, workers may desire large training levels, yet firms would not like to pay for them. We denote the proportion of learning time spent on internal learning from worker's and the firm's perspectives as $g_c^W(h_i)$ and $g_c^F(h_i)$, respectively.

We now solve for the overall learning time and the optimal proportion spent on internal learning separately for workers and firms. For a worker i with human capital level h_i in firm j with productivity z, the worker's value function is given by

$$\begin{aligned} V^W(h_i, z) = \max_{g^W, g_c^W} \; & \underbrace{w(z)}_{\text{wage rate}} \underbrace{h_i}_{\text{human capital}} - \underbrace{(1-\mu)\left[\delta_s z h_i g_c^W + (\delta_s z h_i + q p_s(h_i))(1-g_c^W)\right] g^W}_{\text{learning costs borne by the worker}} \\ & + \underbrace{\beta(1-\delta)(1-\delta_{job})(1-\eta\theta q(\theta)\bar{F}(w))\mathbb{E}\left[p_{learn}(h_i,z)V^W(h_{i+1},z) + (1-p_{learn}(h_i,z))V^W(h_i,z)\right]}_{\text{the worker's future value if stays at current firm}} \\ & + \underbrace{\beta(1-\delta)(1-\delta_{job})\eta\theta q(\theta)\int_{w(z')>w} \mathbb{E}\left[p_{learn}(h_i,z)V^W(h_{i+1},z') + (1-p_{learn}(h_i,z))V^W(h_i,z')\right] dF(w(z'))}_{\text{the worker's future value for job-to-job transitions}} \\ & + \underbrace{\beta(1-\delta)\delta_{job}\mathbb{E}\left[p_{learn}(h_i,z)\max\{V^U(h_{i+1}), V^{TR}(h_{i+1})\} + (1-p_{learn}(h_i,z))\max\{V^U(h_i), V^{TR}(h_i)\}\right]}_{\text{the worker's future value for job separations}}, \end{aligned} \tag{2}$$

where $\bar{F}(w(z)) = 1 - F(w(z))$, and the expectation is taken with regard to uncertainty about realizations of human capital depreciation. The value functions of unemployment and becoming a trainer are given by

$$V^U(h_i) = \theta q(\theta) \int V(h_i, z) dF(w(z)) + (1-\theta q(\theta))\beta(1-\delta)\mathbb{E}\max\{V^U(h_i), V^{TR}(h_i)\}$$

and

$$V^{TR}(h_i) = w_n(h_i) + \beta(1-\delta)\mathbb{E}\max\{V^U(h_i), V^{TR}(h_i)\}.$$

On the other hand, firm j's value function from matching with worker i can be recursively written as

$$\begin{aligned} V^F(h_i, z) = \max_{g^F, g_c^F} & \underbrace{(z - w(z))h_i}_{\text{labor revenue}} - \underbrace{\mu\left[\delta_s z h_i g_c^F + (\delta_s z h_i + q p_s(h_i))(1-g_c^F)\right] g^F}_{\text{learning costs borne by the firm}} + \underbrace{\int_{i'\in j, i'\neq i} \Delta V^F(h_{i'}, z) di'}_{\text{benefits to other workers}} \\ & + \underbrace{\beta(1-\delta)(1-\delta_{job})(1-\eta\theta q(\theta)\bar{F}(w))\mathbb{E}\left[p_{learn}(h_i, z)V^F(h_{i+1}, z) + (1-p_{learn}(h_i, z))V^F(h_i, z)\right]}_{\text{the firm's future value if the worker stays}}, \end{aligned} \tag{3}$$

We use the term $\int_{i'\in j, i'\neq i} \Delta V^F(h_{i'}, z) di$ to capture the fact that having worker i in firm j potentially benefits the other workers in this firm since it changes the pool of colleagues they can learn from. $\Delta V^F(h_{i'}, z)$ captures the change in $V^F(h_{i'}, z)$ coworkers experience from having worker i, who has human capital h_i, in the firm. Thus, when choosing the human capital investments for worker i, the firm will take into account that this may also benefit other workers in the firm.[42]

From these value equations, and given the wage rate $w(z)$, we can solve for the total time spent on learning, and the proportion of this spent on internal learning that maximize workers' and firms' value functions respectively.[43] The proportion of learning time spent on internal learning is given by:

$$\frac{g_c^W}{1-g_c^W} = \frac{g_c^F}{1-g_c^F} = \frac{(\delta_s z h_i)^{-\sigma}(A_c p_c(h_i, z))^{\sigma-1}}{(\delta_s z h_i + q p_s(h_i))^{-\sigma}(A_s p_s(h_i))^{\sigma-1}}. \tag{4}$$

The optimal share of learning time spent on internal and external learning is identical between the firm and the worker, since they face the same incentives in the division of time. A larger likelihood of internal learning $A_c p_c$, increases the proportion of learning time spent on internal learning. Moreover, a lower cost of internal learning relative to external learning, $\frac{\delta_s z h}{\delta_s z h + q p_s(h_i)}$, which prevails when workers are low in the human capital ladder, also contributes to a higher proportion of learning time spent on internal learning.

[42]In principle, the increment in colleagues' human capital due to worker i's increment can also in turn affect worker i's human capital. Due to computational intractability, we abstract from these higher-order effects.

[43]The equations describing wage rate $w(z)$ are presented in Appendix F.1.

The total time spent on learning that maximize workers' and firms' value functions, g^W and g^F are given by

$$g^W = \left\{ \frac{(1-\delta_{job})(1-\eta\theta q(\theta)\bar{F}(w))\mathbb{E}\frac{\partial V^W(h_i,z)}{\partial h_i} + (1-\delta_{job})\eta\theta q(\theta)\int_{w(z')>w}\mathbb{E}\frac{\partial V^W(h_i,z')}{\partial h_i}dF(w(z')) + \delta_{job}\mathbb{E}\frac{\partial \max\{V^U(h_i),V^{TR}(h_i)\}}{\partial h_i}}{(1-\mu)\left[(\delta_s z h_i/A_c p_c(h_i,z))^{1-\sigma} + ((\delta_s z h_i + q p_s(h_i))/A_s p_s(h_i))^{1-\sigma}\right]^{\frac{1}{1-\sigma}}/\gamma\beta(1-\delta)} \right\}^{1/(1-\gamma)},$$

and

$$g^F = \left\{ \frac{(1-\delta_{job})(1-\eta\theta q(\theta)\bar{F}(w))\mathbb{E}\frac{\partial V^F(h_i,z)}{\partial h_i}}{\mu\left[(\delta_s z h_i/A_c p_c(h_i,z))^{1-\sigma} + ((\delta_s z h_i + q p_s(h_i))/A_s p_s(h_i))^{1-\sigma}\right]^{\frac{1}{1-\sigma}}/\gamma\beta(1-\delta)} \right\}^{1/(1-\gamma)}.$$

With some abuse of notation, we use the expressions $\frac{\partial V^W(h_i,z)}{\partial h_i}$ and $\frac{\partial V^F(h_i,z)}{\partial h_i}$ to capture the increment in the worker's and the firm's values of climbing one more step in the human capital ladder, respectively.[44] For each party, the desired total learning time depends on the relative benefits and costs of human capital accumulation. In particular, the firm prefers a lower learning time relative to the worker, since it cannot glean the benefits from the worker's human capital increment after she leaves the firm.

Thus, since training is determined by the party with lower affordability, learning investments are generally determined by the firm. This leads to two key inefficiencies that generate an underinvestment in learning. First, firms do not internalize workers' gains from learning and the gains of future employers from a better pool of hires. This type of inefficiency reflects a hold-up problem and has been discussed in several papers such as Acemoglu (1997), Acemoglu and Pischke (1998), Moen and Rosén (2004), among others. Second, firms do not internalize that learning investments improve the economy's learning environment by changing the skill composition of coworkers and trainers. This suggests a positive externality of firms' investments in workers' skills which is novel and has been underexplored by the previous literature.[45], In Appendix G.3 we assess the role of subsidies that pay for a portion of firms' overall learning costs to correct these inefficiencies. We find that subsidizing learning can

[44] Particularly, the term $\left[(\delta_s z h_i/A_c p_c(h_i,z))^{1-\sigma} + ((\delta_s z h_i + q p_s(h_i))/A_s p_s(h_i))^{1-\sigma}\right]^{\frac{1}{1-\sigma}}$, can be viewed as the unit cost of learning, which is a CES aggregator of the unit cost of internal learning ($\frac{\delta_s z h_i}{A_c p_c(h_i,z)}$) and the unit cost of external learning ($\frac{(\delta_s z h_i + q p_s(h_i))}{A_s p_s(h_i,z)}$).

[45] There are two other potential inefficiencies in our economy. First, workers also fail to internalize the positive externalities of learning investments. However, since we assume time spent on learning is determined by the party with lower affordability (i.e, the party who would like a lower level of learning), firms decisions on learning are predominant. Second, the pattern of positive assortative matching documented above can result in low human capital individuals having too few chances to learn from more knowledgeable coworkers. Recent papers that focus on coworker learning have documented the importance of this inefficiency (Herkenhoff et al. (2018), Jarosch et al. (2019)). We focus our attention on inefficiencies arising from firms underinvestments in skills instead, and the potential for subsidies to learning to correct them.

generate sizeable increases to human capital and aggregate output, and that the impact of jointly subsidizing both sources of learning is much larger than the effects of subsidizing each source individually.

5.2 Calibration

We calibrate the above framework to the United States. There are two sources for parameter values: exogenously calibrated parameters using the literature and data, and internally calibrated parameters that match targeted moments.

5.2.1 Exogenously Calibrated Parameters

Table 5.1 presents the set of exogenously calibrated parameters. A period in the model is one quarter. We calibrate the discount rate to be $\beta = 0.99$. The death rate is set to $\delta = \frac{1}{160}$ to correspond to 40 years (160 quarters) of working life on average. Without loss of generality, we let the step size of the human capital ladder be $\gamma_h = 0.05$ such that climbing a step implies a 5% increase in human capital. We normalize the lower bound of human capital ladder h_1 to 1. We set the number of firms to be $\bar{M} = 0.05$ (relative to the number of workers) such that the average size of the firm is around 20 according to the US Economic Census in 2007. γ captures the degree of diminishing returns of human capital investments (in terms of effective hours) in producing new human capital. Imai and Keane (2004) find this parameter to be 0.22, while Manuelli and Seshadri (2014) estimate it to be 0.48. We set $\gamma = 0.35$ following the average of these estimates. We calibrate the elasticity of matches with regard to the number of searchers, $\phi = 0.7$, according to Shimer (2005). We obtain the min-mean wage ratio $b = 0.5$ from Hornstein et al. (2011).

Table 5.1: Exogenously Calibrated Parameters

Label	Description	Value
β	discount rate	0.99
δ	death rate	0.006
γ_h	step size of human capital ladder	0.05
h_1	lower bound of human capital ladder	1
$\bar{M}$	measure of firms	0.05
γ	degree of diminishing returns in learning	0.35
ϕ	elasticity of matches with regard to searchers	0.7
b	min-mean wage ratio	0.5
δ_s	time cost of learning	0.7
μ	share of learning costs borne by firms	0.8
γ_v	curvature of vacancy cost	1

Due to the lack of US data, we calibrate some parameters using data from other countries. We set the time cost of learning $\delta_s = 0.7$ and the share of learning costs borne by firms $\mu = 0.8$ from the EU Adult Education Survey (EU-AES).[46] This large share of learning costs borne by firms suggests that firms play a key role in the formation of on-the-job human capital. Finally, we obtain the curvature of vacancy costs $\gamma_v = 1$ from Dix-Carneiro et al. (2019)'s estimate on Brazilian firms.

5.2.2 Internally Calibrated Parameters

We are left with 10 parameters to estimate: the cognitive abilities in learning from internal and external sources $\{A_c, A_s\}$, the elasticity of substitution between the two modes of learning σ, the exogenous probability of moving up the human capital ladder ϵ, the depreciation rate of human capital δ_h, the constant in the matching function c_m, the constant in vacancy costs c_v, the shape parameter of the firm productivity distribution κ, the exogenous separation rate of workers δ_{job}, and on-the-job search intensity η.

Table 5.2: Moments in the Data and the Model

Description	Model	Data
Share of total time spent on external learning	0.006	0.006
Share of total time spent on internal learning	0.013	0.013
Ratio of new to all workers' average time spent on external learning	1.52	1.51
Average wage growth (per quarter) within 0–40 years of experience	0.005	0.005
Average wage growth (per quarter) within 25–40 years of experience	0.001	0.001
Unemployment rate	0.06	0.06
Labor market tightness (#vacancies/#unemployed)	0.55	0.55
Shape parameter of firm employment distribution	1.12	1.10
Share of workers that remain employed in next quarter	0.97	0.97
Share of workers that remain in the same firm in next quarter	0.94	0.94

Notes: The relative shares of time spent on external and internal learning are drawn from Ma et al. (2020), which we match to the time spent on formal and informal training. The ratio of new (with 1 year of experience) to all workers' average time spent on external learning is computed using the NHES data. Average wage growth per quarter is drawn from Lagakos et al. (2018). Unemployment rate and labor market tightness are averaged over 1994–2007, using the data from FRED. Shape parameter of firm employment distribution is from Axtell (2001). The share of workers that remain employed in the next quarter and the share of workers that remained in the same firm in the next quarter are from Donovan et al. (2020).

We estimate these parameters using the method of moments to minimize the squared differences between model and data moments. We target the following 10 data moments: the average share of time spent on internal and external learning, the ratio of new to all workers'

[46] The EU-AES collects information on participation in education and learning activities including job-related training, among others, and covers around 666,000 adults ages 25–64. This data was collected during 2007, 2011, and 2017 in 26, 27, and 28 EU member states, respectively. See Ma et al. (2020) for details on this data and calibration of this parameter.

average time spent on external learning, the average quarterly returns to experience within 0–40 and 25–40 years of experience, the unemployment rate, the labor market tightness, the tail shape parameter of the firm employment distribution, the share of workers that remain employed in the next quarter, and the share of workers that remain in the same firm in the next quarter. The sources of these moments are listed in Table 5.2.

Table 5.3 reports the values of the internally calibrated parameters. Overall, the parameter values are reasonable and in line with other work. We find the the elasticity of substitution between the two modes of learning to be $\sigma = 2.18$, suggesting moderate substitutability. Each person has a 3% chance to climb the human capital ladder exogenously each period. Our calibrated quarterly depreciation rate of human capital from learning $\delta_h = 0.014$ is similar to the annual depreciation rate of 0.06–0.08 of training returns estimated by Blundell et al. (2021) using British labor surveys. The on-the-job search intensity parameter is calibrated to $\eta = 0.2$, similar to around 0.3 found in Faberman et al. (2017). With these parameter values, our model is able to match the targeted data moments quite well, as suggested in Table 5.2.

Table 5.3: Internally Calibrated Parameters

Label	Description	Value
A_c	cognitive ability of internal learning	0.78
A_s	cognitive ability of external learning	0.46
σ	elasticity of substitution between two modes of learning	2.18
ϵ	exogenous human capital gain	0.03
δ_h	depreciation rate of human capital	0.014
c_m	constant in matching function	0.59
c_v	constant in vacancy costs	7.76
κ	shape parameter of firm productivity distribution	9.71
δ_{job}	exogenous separation rate	0.03
η	on-the-job search intensity	0.20

5.3 Properties of Equilibrium

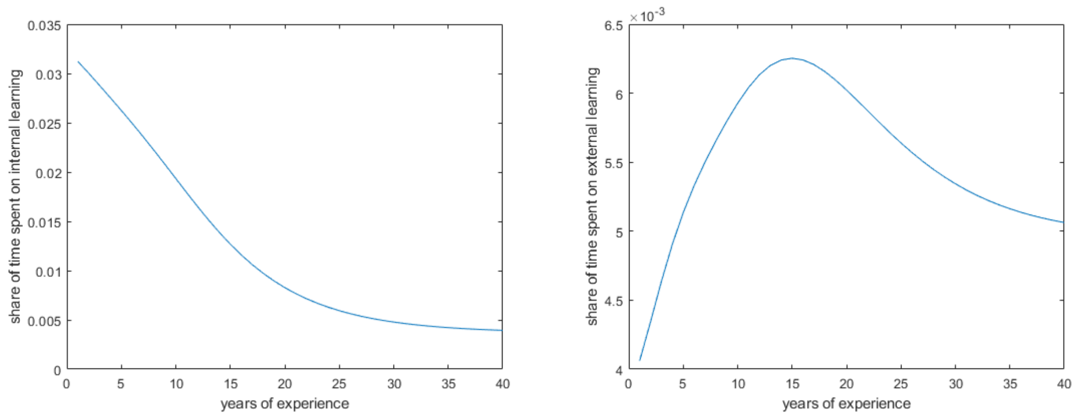
We now examine some key properties of our stationary equilibrium, and how they match up with both our empirical evidence, and findings in the literature. We explore properties relating to the lifecycle of learning, the relationship between productivity and the firm's learning environment, sorting, and the importance of coworkers for human capital and wage dynamics.

5.3.1 The Lifecycle of Learning and Human Capital Distribution

Figure 5.2 shows that our quantitative model yields lifecycle results analogous to those of the analytical model in Section 4 and thus matches the empirical findings in Section 3. In particular, the model generates that as experience increases, the total time spent on internal learning declines, whereas the time spent on external learning first increases and then declines.

Figure 5.2: Lifecycle Patterns of Internal and External Learning

(a) Internal Learning (b) External Learning



Note: These figures depict total time spent in each source of learning, namely $g_c \times g$, and $(1 - g_c) \times g$.

This pattern stems from the same forces as in the analytical model. When workers are young and have a low level of human capital, they join the production sector and face a large contingent of coworkers with higher human capital than the own. Moreover, their opportunity costs of learning are relatively low compared with the returns of human capital accumulation. This incentivizes both firms and workers to invest a considerable amount of time in learning, particularly via internal sources. As workers become more experienced and their human capital rises, the contingent of coworkers with higher human capital than the own shrinks, reducing the probability individuals can learn from coworkers. This leads firms and workers to invest a larger portion of time in external learning, which allows the worker to match with a better pool of individuals to learn from. Eventually, however, this rise in the portion of external learners is reversed as workers become more productive (and reach the highest human capital step) and thus the relative benefit of learning decreases. [47]

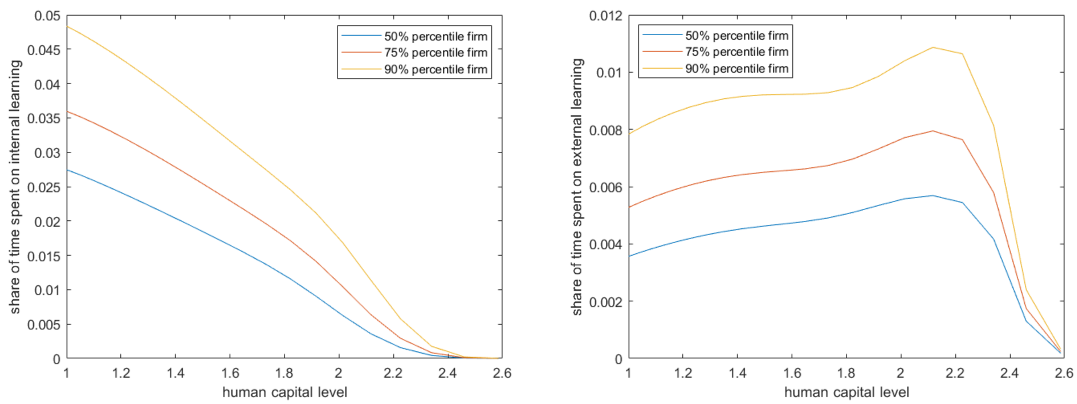
[47] In Figure G.1 we explore how the distribution of human capital changes as workers age by plotting the distribution of human capital levels for a given cohort of workers observed at different ages. We find that workers' human capital grows rapidly during the first few years after entering the labor force, and slows

5.3.2 Firm Productivity and Learning Environment

We now consider how learning environments vary across firms. Specifically, we consider how the provision of internal and external sources of learning varies with intrinsic firm productivity, and how these two source shape firms' learning environments. In Figure 5.3, we plot the share of time spent on internal and external learning at each human capital level for firms at the 50th, 75th, and 90th percentile in the firm productivity distribution.

Figure 5.3: Time Spent on Each Learning Activity by Firm Productivity

(a) Internal Learning (b) External Learning



We find that more productive firms provide better learning environments by offering greater variety in learning options. Although all firms follow the same lifecycle patterns of learning documented empirically and described above in Section 5.3.1, at all levels of human capital workers in more productive firms spend more time in both internal and external learning, and thus climb the human capital ladder faster.[48] This stems from the fact that more productive firms exhibit both larger returns to skill acquisition (due to supermodularity of the production function), and a better pool of coworkers to learn due positive assortative matching. The equilibrium pattern of assortative matching in the model is described in

down in later years, consistent with the evidence on the lifecycle returns to experience (Rubinstein and Weiss (2006)). In our model, this slowdown stems not only from the depreciation of human capital, but also from the reduction in the scope of learning that occurs as workers climb up the human capital ladder and have fewer colleagues and trainers to learn from.

[48] To further illustrate how workers' human capital is affected by their relative positions in the human capital ladder and their employers' characteristics, we plot the lifecycle human capital and learning patterns for two workers in Figure G.2, along with their employers' productivity levels. When these workers are young they learn mostly from internal sources, but substitute away from these and towards external sources as they climb the human capital ladder. When matched with more productive firms, these workers spend more time on both internal and external learning, and thus enjoy a faster rise in human capital.

Appendix G.1.1.[49]

This finding matches our empirical evidence showing that workers in larger European firms spend significantly more hours engaging in both sources of learning, and is also consistent with evidence found by Engbom (2017) and Arellano-Bover (2020) showing that workers in more productive firms exhibit faster rates of skill acquisition. In addition, this finding is consistent with the evidence found by Gregory (2019), showing that having different forms of training available is important for firms' learning environments.

5.4 Counterfactual Analysis: Role of Each Source of Learning

To assess the role of each source of learning in explaining aggregate human capital accumulation, wage growth, and wage dispersion, we now perform counterfactual exercises which subsequently shut down each of the two sources of learning, and examine how the stationary equilibrium changes. First, we set the productivity of external learning A_s to zero, so that the amount of time spent on learning from external sources has no bearing on the probability of climbing the human capital ladder. Then, we set the productivity of learning from internal sources A_c to zero, so that the amount of time spent on learning from internal sources has no bearing on the probability of climbing the human capital ladder.

5.4.1 Role of Each Source of Learning in Aggregate Human Capital

Table 5.4 summarizes the shares of time spent on each source of learning and workers' average human capital levels in the stationary equilibrium of our baseline model, and the stationary equilibrium of the model without internal and external sources of learning. We find that both learning from internal and external sources contribute largely and roughly equally to workers' human capital: without learning from external sources, workers' human capital decreases by 30%, whereas without learning from internal sources, workers' human capital decreases by 29%. Without the two sources of learning, our model still predicts positive human capital gains due to the exogenous learning-by-doing probability.

Moreover, the results suggest that even though these two sources of learning are substitutable

[49]To further validate our model and examine the importance of coworkers for human capital and wage dynamics, in Appendix G.1.2 we compare our quantitative results with Herkenhoff et al. (2018) who use data from the US to show that a worker's future wage is affected by the average wages of its coworkers in the current firm. Our model generates a positive correlation between the future wage realizations of workers and the wage of their coworkers that is similar in magnitude to that found by Herkenhoff et al. (2018). Similar to their results, we also find that these results are particularly marked for workers who are paid less than their coworkers.

for each individual worker, they are highly complementary in the aggregate. In particular, shutting down external learning leads to a sharp decline in the time spent on learning from internal sources, and similarly, shutting down internal learning leads to a sharp decline in the time spent on learning from outside sources. This aggregate complementarity stems from the fact that the existence of each source of learning improves the potential pool and probability of the other source. The existence of external learning increases employees' human capital within the firm, and raises the scope of internal learning. The existence of internal learning raises the human capital of workers who eventually become trainers and thus provides a better potential pool of trainers.

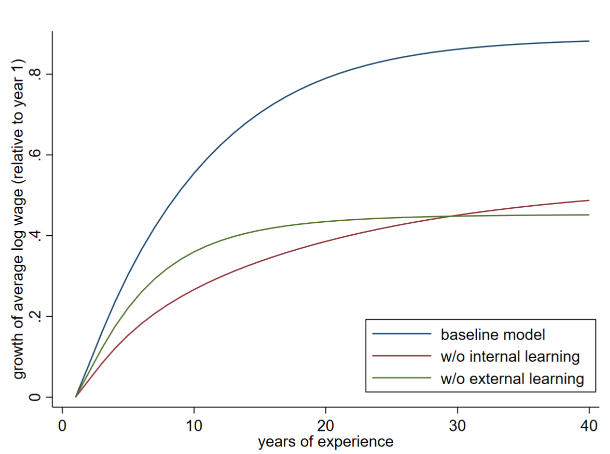
Table 5.4: Counterfactual Exercises

	Workers' Share of Time Spent on Learning		
	External Learning	Internal Learning	Avg Human Capital
Calibrated Economy	0.59%	1.32%	1.87
W/o External Learning	0	0.85%	1.31
W/o Internal Learning	0.47%	0	1.32
W/o Both Sources of Learning	0	0	1.10

5.4.2 Role of Each Source of Learning in Wage Growth

Figure 5.4 plots the lifecycle wage growth in the baseline equilibrium and the counterfactual scenarios. We find that lifetime wage growth is considerably lower in the absence of either internal or external learning compared to the baseline model. In addition, we find that although internal learning is more critical for young workers' wage growth, external learning is more critical for overall lifetime wage growth. Given that internal learning is cheaper than

Figure 5.4: Lifetime Wage Growth in Baseline and Counterfactual Exercises



external learning, young workers' wages grow considerably slower in the scenario without internal learning than that without external learning. However, since the returns to internal learning are depleted more quickly since trainers tend to have higher human capital levels than production workers, older workers' wages grow considerably slower in the scenario without external learning than that without internal learning. This latter effect is quite strong, and causes overall lifetime wage growth to be lower in the scenario without external learning than that without internal learning.

5.4.3 Role of Each Source of Learning in Wage Dispersion Over the Lifecycle

We now consider how wage dispersion changes over the lifecycle in our benchmark equilibrium and counterfactual scenarios. To understand the drivers behind wage dispersion dynamics, we decompose wage dispersion into different components. Note that earnings/wages can be expressed as:

$$\begin{aligned}\log(\text{wage}) &= \log\left(\underbrace{w(z)}_{\text{wage rate}}\underbrace{h}_{\text{human capital}} - \underbrace{(1-\mu)\left[\delta_s z h g_c + (\delta_s z h + q p_s(h_i))(1-g_c)\right] g}_{\text{learning costs borne by the worker}}\right) \\ &= \log w(z) + \log h + \log\left(\frac{w(z)h - (1-\mu)\left[\delta_s z h g_c + (\delta_s z h + q p_s(h_i))(1-g_c)\right] g}{w(z)h}\right).\end{aligned}$$

In Figure 5.5, we plot the variance of log wages across workers in the baseline and counterfactual scenarios, in addition to the different components driving it: the variance of the log wage rate per efficiency unit of labor, the variance of log human capital, and the residual, which captures the variance of the learning costs borne by the workers along with the covariance between different components. We have three main findings. First, we find that wage dispersion increases over the lifecycle, consistent with the literature.[50] Second, we find that in both the baseline and counterfactual cases, the variance of the log wage rate per efficiency unit of labor decreases with experience as workers who initially match with low productivity firms offering low wage rates climb the job ladder.

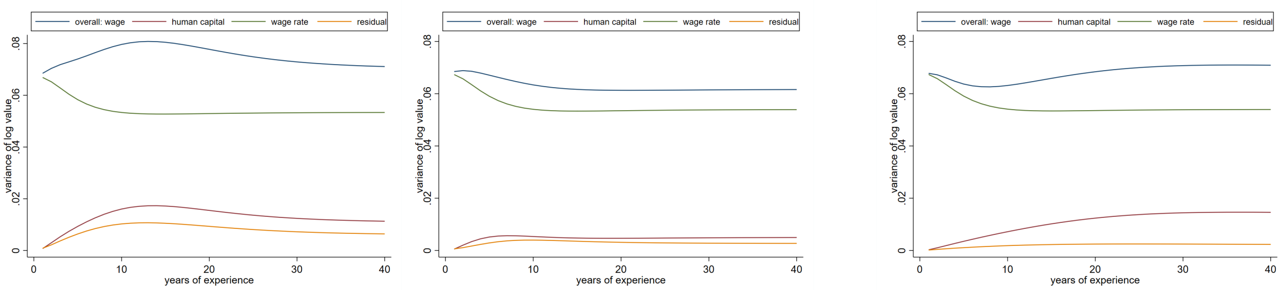
Finally, and most importantly, we find that wage dispersion is driven by external learning in the model. Since all new workers are born with the same human capital level, the dis-

[50]The wage dispersion predicted by our model is relatively small compared with the data as we abstract from individual heterogeneity in innate abilities. The slight hump-shape observed is also consistent with the literature (Lagakos et al. (2018)).

persion of human capital is zero at the beginning of workers' careers in both the baseline and counterfactual scenarios. The dispersion of human capital emerges as a consequence of learning being stochastic, and particularly the fact that some workers are more lucky than others and able to climb the human capital ladder faster. Without external learning, human capital dispersion remains low throughout the lifecycle as workers learn from and catch up to colleagues who have no external opportunities to further learn. Without internal learning, skill acquisition is more expensive, and thus only a small number of workers climb the human capital ladder. This causes human capital dispersion to rise throughout the lifecycle, as learning opportunities are depleted slowly since trainers are positively selected.[51]

Figure 5.5: Components Driving Wage Dispersion Over the Lifecycle

(a) Baseline Model (b) Without External Learning (c) Without Internal Learning



6 Conclusions

On-the-job human capital accumulation is key in accounting for lifecycle earnings dynamics and dispersion. However, the interactions and importance of different sources in shaping workers' lifecycle learning have been underexplored. In this paper, we explore both empirically and theoretically how different sources of worker-level skill acquisition shape workers' lifecycle human capital accumulation. We document two novel facts that speak to the importance of both sources of learning from both firm and worker perspectives. First, using enterprise survey data from Europe, we show that both internal and external sources of learning are widely provided by firms to their workers, and that larger firms provide better learning environments by offering greater variety in learning options. Second, we use detailed worker qualification data from Germany and the United States to show that both sources of

[51] This is further illustrated in Figure G.4, which plots the distribution of human capital across workers in our baseline equilibrium, and the counterfactual scenarios, and indicates that the distribution of human capital is more dispersed when we shut down internal learning than when we shut down external learning.

learning are important to workers, and have markedly different lifecycle patterns: the prevalence of internal learning decreases with worker experience; and the prevalence of external learning has an inverted U-shape in worker experience. We build a benchmark model where the incentives to engage in each source of skill acquisition evolve throughout the lifecycle due to shifts in the relative position of the worker in the human capital distribution to shed light on the mechanisms giving rise to these facts. Then, we embed this mechanism in a search model with firm heterogeneity that considers the rich interactions of firms and workers in learning investments. We use this quantitative framework to assess the importance of each source in shaping firms' learning environments, and in explaining aggregate human capital acquisition and lifecycle wage growth and dispersion.

Our results have several implications. First, our results suggest that policies aimed at encouraging apprenticeships or other internal learning practices are especially important for younger workers, but may have important spillover effects as these workers age and are able to teach younger workers. Second, our results suggest that internal and external sources of learning are highly complementary in the aggregate, so policies aimed towards increasing firms' external (internal) training investments may have large effects on internal (external) learning that amplify the human capital accumulation and growth effects of these policies. Finally, our results suggest that other sources of human capital accumulation, such as schooling or nutrition may also be important to incentivize on-the-job human capital accumulation since they increase the pool of knowledge in the economy, and thus increase the returns from learning. Studying the interactions between these different sources of learning occurring at different points in the lifecycle would be an interesting avenue for future research.

References

Abowd, J. M., McKinney, K. L., and Zhao, N. L. (2018). Earnings Inequality and Mobility Trends in the United States: Nationally Representative Estimates from Longitudinally Linked Employer-Employee Data. *Journal of Labor Economics*, 36. 106

Acemoglu, D. (1997). Training and Innovation in an Imperfect Labour Market. *The Review of Economic Studies*, 64(3):445. ISBN: 00346527. 2, 6, 30, 37

Acemoglu, D. and Pischke, J.-S. (1998). Why Do Firms Train? *The Quarterly Journal of Economics*, 113(1):79–119. 2, 6, 33, 37

Acemoglu, D. and Pischke, J.-S. (1999). Beyond Becker: Training in Imperfect Labour Markets. *The Economic Journal*, 109(453):112–142. ISBN: 1468-0297. 7, 30

Altonji, J. G. and Shakotko, R. A. (1987). Do Wages Rise with Job Seniority? *Review of Economic Studies*, 54(3):437–459. 16

Arellano-Bover, J. (2020). Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size. 5, 7, 12, 43

Autor, D. H. (2001). Why do temporary help firms provide free general skills training? *Quarterly Journal of Economics*, 116(4):1409–1448. ISBN: 9780470343890. 7

Axtell, R. L. (2001). Zipf Distribution of U.S. Firm Sizes. *Science*, 293(5536):1818–1820. 39

Azoulay, P., Wang, J., and Zivin, J. S. G. (2010). Superstar Extinction. *The Quarterly Journal of Economics*, 125(2):549–589. 6, 15

Bagger, J., Fontaine, F., Postel-Vinay, F., and Robin, J. M. (2014). Tenure, experience, human capital, and wages: A tractable equilibrium search model of wage dynamics. *American Economic Review*, 104(6):1551–1596. 2, 7, 33

Bagger, J. and Lentz, R. (2019). An Empirical Model of Wage Dispersion with Sorting. *Review of Economic Studies*, 86(1):153–190. 7

Barlevy, G. (2008). Identification of search models using record statistics. *The Review of Economic Studies*, 75(1):29–64. 7

Barth, E., Bryson, A., Davis, J. C., and Freeman, R. (2016). It's Where You Work: Increases in the Dispersion of Earnings across Establishments and Individuals in the United States. *Journal of Labor Economics*, 34. 106

Becker, G. S. (1962). Investment in Human Capital: A Theoretical Analysis. *Journal of Political Economy*, 70(5, Part 2):9–49. 2

Becker, G. S. (1964). *Human Capital*. University of Chicago Press, Chicago. 6

Ben-Porath, Y. (1967). The Production of Human Capital Over the Life Cycle. *Journal of Political Economy*, 75(4-1):352–365. 2, 7

Benhabib, J., Perla, J., and Tonetti, C. (2021). Reconciling Models of Diffusion and Innovation: A Theory of the Productivity Distribution and Technology Frontier. *Econometrica*, 89(5):2261–2301. 6

Blundell, R. W., Costa Dias, M., Goll, D., and Meghir, C. (2021). Wages, Experience and Training of Women. *Journal of Labor Economics*, 39(1):275–315. 34, 40

Bowlus, A. J. and Liu, H. (2013). The contributions of search and human capital to earnings growth over the life cycle. *European Economic Review*, 64:305–331. 7

Bunzel, H., Christensen, B. J., Kiefer, N. M., and Korsholm, L. (1999). *Equilibrium search with human capital accumulation*, volume 99. Centre for Labour Market and Social Research. 7

Burdett, K., Carrillo-Tudela, C., and Coles, M. G. (2011). Human capital accumulation and labor market equilibrium. *International Economic Review*, 52(3):657–677. 7

Burdett, K. and Mortensen, D. T. (1998). Wage Differentials , Employer Size , and Unemployment. *International Economic Review*, 39(2):257–273. 30, 32, 104

Burstein, A. and Monge-Naranjo, A. (2007). Foreign Know-How, Firm Control, and the Income of Developing Countries. *NBER Working Paper Series*. arXiv: 1011.1669v3 ISBN: 9788578110796. 2

Caicedo, S., Lucas, R. E., and Rossi-Hansberg, E. (2019). Learning, career paths, and the distribution of wages. *American Economic Journal: Macroeconomics*, 11(1):49–88. 2, 6

Card, D., Kluve, J., and Weber, A. (2018). What works? A meta analysis of recent active labor market program evaluations. *Journal of the European Economic Association*, 16(3):894–931. 2, 8

de la Croix, D., Doepke, M., and Mokyr, J. (2016). Clans, Guilds, and Markets: Apprenticeship Institutions and Growth in the Pre-Industrial Economy. Technical report. 6

De Melo, R. L. (2018). Firm wage differentials and labor market sorting: Reconciling theory and evidence. *Journal of Political Economy*, 126(1):313–346. 7

DiNardo, J. E. and Pischke, J. S. (1997). The returns to computer use revisited: Have pencils changed the wage structure too? *Quarterly Journal of Economics*, 112(1):291–303. 97

Dix-Carneiro, R., Goldberg, P., Meghir, C., and Ulyssea, G. (2019). Trade and Informality in the Presence of Labor Market Frictions and Regulations . 39

Donovan, K., Lu, J., and Schoellman, T. (2020). Labor Market Flows and Development. 39

Eeckhout, J. and Kircher, P. (2011). Identifying sorting-In theory. *Review of Economic Studies*, 78(3):872–906. 7

Engbom, N. (2017). Worker Flows and Wage Growth over the Life-Cycle: A Cross-Country Analysis. Publication Title: Mimeo. 5, 12, 43

Engbom, N. (2021). Labor market fluidity and human capital accumulation. 7

Engbom, N., Moser, C., and Sauermann, J. (2022). Firm Pay Dynamics. *NBER Working Paper Series*. 7

Faberman, R. J., Mueller, A. I., Sahin, A., and Topa, G. (2017). Job Search Behavior among the Employed and Non-Employed. *NBER Working Paper Series*. 40

Fitzenberger, B. and Völter, R. (2007). Long-run effects of training programs for the unemployed in East Germany. *Labour Economics*, 14(4):730–755. 8

Friedrich, B., Laun, L., Meghir, C., and Pistaferri, L. (2021). Earnings Dynamics and Firm-Level Shocks. *NBER Working Paper Series*. 7

Garicano, L. (2000). Hierarchies and the Organization of Knowledge in Production. 108(5):874–904. 2, 6

Garicano, L. and Rossi-Hansberg, E. (2004). Inequality and the Organization of Knowledge. *The American Economic Review*, 94(2):197–202. ISBN: 0002-8282. 2, 6

Garicano, L. and Rossi-Hansberg, E. (2006). Organization and Inequality in a Knowledge Economy. *Quarterly Journal of Economics*, (November):1383–1435. 2, 6

Gregory, V. (2019). Firms as Learning Environments: Implications for Earnings Dynamics and Job Search. pages 1–57. 2, 5, 7, 12, 43

Hagedorn, M., Law, T. H., and Manovskii, I. (2017). Identifying Equilibrium Models of Labor Market Sorting. *Econometrica*, 85(1):29–65. 7

Heckman, J. J. (1976). A Life-Cycle Model of Earnings, Learning, and Consumption. *Journal of Political Economy*, 84(4):S11–S44. 7

Heckman, J. J., Lalonde, R. J., and Smith, J. A. (1999). The economics and econometrics of active labor market programs. *Handbook of Labor Economics*, 3(1):1865–2097. 2, 8

Herkenhoff, K., Lise, J., Menzio, G., and Phillips, G. M. (2018). Production and Learning in Teams. Technical report. Publication Title: NBER Woking Paper Series. 2, 5, 6, 18, 37, 43, 107

Hornstein, A., Krusell, P., and Violante, G. L. (2011). Frictional wage dispersion in search models: A quantitative assessment. *American Economic Review*, 101(7):2873–98. 33, 38

Huggett, M., Ventura, G., and Yaron, A. (2011). Sources of lifetime inequality. *American Economic Review*, 101(7):2923–2954. ISBN: 00028282. 7

Imai, S. and Keane, M. (2004). Intertemporal labor supply and human capital accumulation. *International Economic Review*, 45(2):601–641. 38

Jarosch, G. (2022). Searching for Job Security and the Consequences of Job Loss. *NBER Working Paper Series*. 7

Jarosch, G., Oberfield, E., and Rossi-hansberg, E. (2019). Learning from Coworkers. Technical report. Publication Title: NBER Woking Paper Series. 2, 6, 15, 18, 37

Jovanovic, B. (2014). Misallocation and growth. *American Economic Review*, 104(4):1149–1171. 6

Kambourov, G. and Manovskii, L. (2009). Occupational Specificity of Human Capital. *International Economic Review*, 50(1):63–115. 16

Karahan, F., Ozkan, S., and Song, J. (2022). Anatomy of Lifetime Earnings Inequality: Heterogeneity in Job Ladder Risk vs. Human Capital. *FRB St. Louis Working Paper Series*. 7

Kluve, J. (2010). The effectiveness of European active labor market programs. *Labour Economics*, 17(6):904–918. Publisher: Elsevier B.V. 2, 8

Lagakos, D., Moll, B., Porzio, T., Qian, N., and Schoellman, T. (2018). Life Cycle Wage Growth across Countries. *Journal of Political Economy*, 126(2). 7, 39, 45

Lazear, E. P. (2009). Firm-Specific Human Capital: A Skill-Weights Approach. *Journal of Political Economy*, 117(5):699–726. 16

Lise, J., Meghir, C., and Robin, J. M. (2016). Matching, sorting and wages. *Review of Economic Dynamics*, 19(November):63–87. Publisher: Elsevier Inc. 7

Lucas, R. E. (2009). Ideas and growth. *Economica*, 76(301):1–19. 7

Lucas, R. E. and Moll, B. (2014). Knowledge Growth and the Allocation of Time. *Journal of Political Economy*, 122(1):1–51. arXiv: 1011.1669v3 ISBN: 0022-3808. 6

Luttmer, E. G. J. (2014). An Assignment Model of Knowledge Diffusion and Income Inequality. 6

Ma, X., Nakab, A., and Vidart, D. (2020). Human Capital Investment and Development: The Role of On-the-job Training. 2, 7, 13, 39

Manuelli, R. E. and Seshadri, A. (2014). Human capital and the wealth of nations. *American Economic Review*, 104(9):2736–2762. 2, 7, 38

Manuti, A., Pastore, S., Scardigno, A. F., Giancaspro, M. L., and Morciano, D. (2015). Formal and informal learning in the workplace: a research review: Formal and informal learning in the workplace. *International Journal of Training and Development*, 19(1):1–17. 16

McKenzie, D. (2017). How effective are active labor market policies in developing countries? A critical review of recent evidence. *World Bank Research Observer*, 32(2):127–154. 2, 8

Moen, E. R. and Rosén, (2004). Does poaching distort training? *Review of Economic Studies*, 71(4):1143–1162. 2, 7, 30, 37

Nix, E. (2017). Learning Spillovers in the Firm. 2, 6, 15

Perla, J. and Tonetti, C. (2014). Equilibrium Imitation and Growth. *Journal of Political Economy*, 122(1):1–21. ISBN: 00223808. 6

Rosen, S. (1976). A Theory of Life Earnings. *Journal of Political Economy*, 84(4):S45–S67. 7

Rubinstein, Y. and Weiss, Y. (2006). Chapter 1 Post Schooling Wage Growth: Investment, Search and Learning. *Handbook of the Economics of Education*, 1:1–67. ISBN: 9780444513991. 2, 7, 42

Shimer, R. (2005). The Cyclical Behavior of Equilibrium Unemployment and Vacancies. *American Economic Review*, 95(1):25–49. 38

Shimer, R. and Smith, L. (2000). Assortative Matching and Search. 68(2):343–369. 7

Song, J., Price, D., Guvenen, F., Bloom, N., and Wachter, T. V. (2019). Firming Up Inequality. *Quarterly Journal of Economics*. 106

Teulings, C. and Gautier, P. (2004). The Right Man for the Job. *Review of Economic Studies*, 71:553–580. 7

van den Berg, G. J. (1990). Search Behaviour, Transitions to Non-participation and the Duration of Unemployment. *The Economic Journal*, 100(402):842–865. 33

What Works - Centre for Local Economic Growth (2016). Evidence Review 1: Employment Training. Technical report, London School of Economics. 2, 8

Yamaguchi, S. (2010). Job search, bargaining, and wage dynamics. *Journal of Labor Economics*, 28(3):595–631. 7

A Appendix: Data Construction

A.1 European CVT Data

The European Union's Continuing Vocational Training (EU-CVT) enterprise survey collects information from enterprises across the European Union, and focuses on enterprises' investments in continuing vocational training of their staff, and provides information on the types, content and volume of continuing training, enterprises' own training resources and use of external training providers, costs of continuing training, and initial vocational training. CVT surveys have been carried out for the reference years 1993, 1999, 2005, 2010 and 2015. However, due to data availability, we rely on the three in 2005, 2010, and 2015, labeled as CVT3, CVT4 and CVT5. These three surveys provide a sample of 78,000, 101,000 and 111,000 enterprises, respectively, from across all EU member states and Norway.

A.1.1 Firms' Investments in Skills

To build our measures of internal and external learning opportunities offered by firms, we combine measures of External and Internal CVT Courses and Other Types of CVT Activities from the EU-CVT survey manuals. In particular, firms are considered to offer internal learning if they offer either internal continuing vocational training (formal internal learning) or "other forms of CVT" that draw on the internal knowledge pool (informal internal learning). Similarly, firms are considered to offer external learning if they offer either external continuing vocational training (formal external learning) or "other forms of CVT" that draw on external knowledge (informal external learning). We know explore each of these categories closely.

- Continuing Vocational Training refers to education or training activities that are planned in advance, organized, or supported with the specific goal of learning and financed at least partially by the enterprise. These activities aim to generate "the acquisition of new competences or the development and improvement of existing ones" for firms' employees. These courses are typically separated from the active workplace (for example, they take place in a classroom or at a training institution), show a high degree of organization, and the content is designed for a group of learners (e.g., a curriculum exists). These courses can be organized and taken within the firm, or outside the firm, corresponding to internal and external learning respectively.
- Other Forms of CVT are geared towards learning and are typically connected to the

active work and the active workplace, but they can also include participation (instruction) in conferences, trade fairs, etc. These are often characterized by self-organization by the individual learner or by a group of learners and are typically tailored to the workers' needs. We use the following types of "other forms of CVT" specifically:[52]

– Guided-on-the job training: "It is characterized by planned periods of training, instruction or practical experience in the workplace using the normal tools of work, either at the immediate place of work or in the work situation. The training is organized (or initiated) by the employer. A tutor or instructor is present. It is an individual-based activity, i.e. it takes place in small groups only (up to five participants)." This is categorized as internal learning.

– Job rotation, exchanges, secondments, or study visits: "Job rotation within the enterprise and exchanges with other enterprises as well as secondments and study visits are other forms of CVT only if these measures are planned in advance with the primary intention of developing the skills of the workers involved. Transfers of workers from one job to another which are not part of a planned developmental programme should be excluded." This is categorized as internal learning.

– Learning or quality circles: "Learning circles are groups of persons employed who come together on a regular basis with the primary aim of learning more about the requirements of the work organisation, work procedures and work- places. Quality circles are working groups, having the objective of solving pro- duction and workplace-based problems through discussion. They are counted as other forms of CVT only if the primary aim of the persons employed who participate is learning." This is categorized as internal learning.

– Participation in conferences, workshops, trade fairs, and lectures: "Participation (instruction received) in conferences, workshops, trade fairs and lectures are considered as training actions only when they are planned in advance and if the primary intention of the person employed fo r participating is training/learning." This is categorized as external learning.

[52] Among the "Other types of CVT Activities", the survey also contemplates Self-directed learning, which is more akin to learning-by-doing and thus not considered here.

A.2 German BIBB Data

The BIBB/IAB/BAuA surveys provide comprehensive data to analyze both the cross-sectional and temporal evolution of the qualifications and working conditions of the German workforce. However, this data has two important limitations. First, there is variation on the questions asked across survey waves. This partially compromises the comparability of our skill acquisition measures across waves, and thus the longitudinal nature of the data. However, these changes do not matter for the aggregate lifecycle patterns of on-the-job learning if the age distribution of respondents is constant across waves. More importantly, Table C.1 indicates that the results are robust to controlling for survey wave dummies, while Figure C.4 shows the results are robust to considering each wave separately. Second, the response rate to the survey is low, reaching levels as low as 44%. To address this issue, we adjust all of our results using the weighting schemes provided by BIBB to adjust for both selection probabilities of households and target persons caused by the sample design and the selective failures due to refusals.

A.2.1 Skill Acquisition and Potential Experience

The BIBB questions regarding human capital accumulation changed considerably throughout the past 6 surveys. Therefore, to construct the variables that capture whether the worker engages in "internal learning" or "external learning", different questions (and variables) had to be used as indicators. For the constructed variables, the following guidelines were used.

- "Internal Learning": is a binary variable that indicates whether an individual has acquired the skills/knowledge necessary to complete the tasks required in their current job through colleagues or superiors. This question remains relatively stable throughout the surveys, except for (1) the 1979 survey, which does not distinguish between learning-by-doing and internal learning (and is thus excluded); (2) the 2006 survey, which asks about having received professional development through coaching from superiors, and (3) the 2011/2012 and 2017/2018 surveys, when no related question was asked. It is also important to note that the skill acquisition questions in the 1979–1999 surveys had a slight change after the 1986 survey. In the 1979 and 1986 surveys, individuals could list all the sources through which they acquired the skills needed for their current jobs, whereas in 1992 and 1999 they only listed the two main ones.

- "External Learning": is a binary variable that indicates whether an individual received external on-the-job training in the previous 2–5 years, or acquired the skills/knowledge necessary to complete the tasks in their current job through external training or external firm knowledge.

  - 1979, 1985/1986: For these two waves, external learning corresponds to (1) reporting that the sources of professional knowledge/skills for the job include on-the-job training or continued training; and/or (2) attending any courses with the purpose of training in the 5 years that preceded the survey.
  - 1991/1992, 1998/1999: For these two waves, external learning corresponds to (1) reporting that the two main sources of professional knowledge/skills for the job are on-the-job training or continued training; and/or (2) attending any courses with the purpose of training in the 5 years that preceded the survey, specifically: visiting trade fairs, congresses, or technical lectures; instruction by external agents, or reading circles at the workplace; and reading of trade journals, or specialist literature.
  - 2005/2006: For this wave external learning corresponds to (1) attending any courses with the purpose of training in the 2 years that preceded the survey, specifically: visiting trade fairs, congresses, or or technical lectures; instruction by external agents, or reading circles at the workplace; reading of trade journals, or specialist literature; and learning from computer-based or internet sources; and/or (2) claiming it is important to attend seminars or courses to perform one's occupational activity.
  - 2011/2012, 2017/2018: For these waves, external learning corresponds (1) attending any courses with the purpose of training in the 2 years that preceded the survey (no specific types); and/or (2) claiming it is important to attend seminars or courses to perform one's occupational activity.

- Potential experience was constructed as: $Age - Years\ of\ Schooling - 6$
  - $Years\ of\ Schooling$ was constructed using: $Year\ of\ Graduation - Birth\ Year - 6$.
- The number of years with current employer variable was constructed directly from the corresponding variable in the survey for years 1979, 1985/1986 and 1991/1992, and as $Current\ Year - Year\ Start\ with\ Current\ Employer$ for 1998/1999, 2005/2006, 2011/2012 and 2017/2018. Promotions within the same company are not considered

employer switches. For self-employed workers or business owners, this variable captures the years since the start of running this business or occupation.

A.2.2 Other Variables

Hourly wages are constructed using the monthly wage and regular hours data. In the first few surveys, monthly wage would be answered in an ordinal fashion, with interviewee's picking among different wage ranges. In more recent surveys, the answer is given in exact amounts. Thus, individual wages in early waves are imputed by the mid-point of the reported wage range. Wages are deflated using the German CPI with base 2015 and currency adjusted to account for change to Euro.

A.3 US NHES Data

The National Household Education Survey was first deployed in 1991, and repeated on 1993, 1995, 1996, 1999, 2001, 2003 2005, 2007, 2012 and 2016. The adult education module was not included in every survey, however, and limited to 1991, 1995, 1999, 2001, 2003 2005, and 2016. Moreover, information on internal learning was only first included in the 2016 wave. The data was collected via telephone surveys, which are representative of the US population at large. The adult education module, in particular, focused on non-institutionalized individuals 16 years of age and older.

A.3.1 Skill Acquisition and Potential Experience

In this section, we provide further information on the construction of our key skill acquisition variables of interest, and also of potential experience in the United States. To construct the variables that capture whether the worker experiences "internal learning" and "external learning", we rely on the following questions and guidelines.

- "Internal Learning": is a binary variable that takes a value of one for workers who reported receiving instruction or training from a co-worker or supervisor in their last work experience program, and a value of zero for all other workers surveyed. As such, both workers who reported participating in a work-experience program but did not receive instruction from coworkers or supervisors, and workers who do not report having recently participated in a work experience program are assumed to not have this source of learning. This follows from the definition of work-experience program, which is defined as a job with learning attributes, such as an internship, co-op, practicum, clerkship, externship, residency, clinical experience, apprenticeship, or other learning components.[53]

- "External Learning": is a binary variable that takes a value of one for workers who either reported taking classes or training from a company, association, union, or private instructor in their last work experience program; or ever earned a training certificate from an employment-related training program. The variable takes a value of zero for all other workers. Therefore, workers who reported participating in a work-experience program but did not receive training, and workers who fail to report both having recently participated in a work experience program and receiving an employment-related training certificate are assumed to not have this source of learning. This again follows from the definition of work-experience program, which is defined as a job with learning attributes, such as an internship, co-op, practicum, clerkship, externship, residency, clinical experience, apprenticeship, or other learning components.[54]

- Potential experience was constructed as: $Age - Years\ of\ Schooling - 6$
  - $Years\ of\ Schooling$ was constructed by mapping the educational attainment to the corresponding years of schooling. We omit workers with an educational attainment of less than secondary, since we can't directly map this into years of schooling.

A.3.2 Other Variables

- Hourly wages are constructed using the yearly work earnings, weeks worked, and regular hours data. Yearly work earnings and weeks worked are answered in an ordinal fashion. Thus, yearly wage earnings and weeks worked are imputed by the mid-point of the reported range.

[53]In Panel (a) of Figure C.7, we show that the results are robust to limiting the internal learning variable only to individuals reporting participating in a work-experience program.

[54]In Panels (b) and (c) of Figure C.7, we show that the results are robust to decomposing across the two learning components mentioned before.

A.4 Summary Statistics of Learning Data in Germany and the US

Table A.1: Summary Statistics in Germany and the US

	(1) Mean	(2) Std. Dev.	(3) Minimum	(4) Maximum	(5) # Obs.
Germany					
Reports internal learning	0.31	0.46	0	1	109478
Reports external learning	0.68	0.47	0	1	173391
Woman	0.42	0.49	0	1	174647
Age	40.15	11.17	15	74	174647
Years of Education	10.75	2.67	0	25	174647
Potential Years of Experience	23.4	11.59	1	45	174647
Years with Current Employer	11.34	9.91	0	70	166964
Hourly Wage (Euros of 2015)	8.96	9.07	0	207.15	117293
Firm size 1-9	0.23	0.42	0	1	165770
Firm size 10-99	0.37	0.48	0	1	165770
Firm size 100+	0.4	0.49	0	1	165770
USA					
Reports internal learning	0.23	0.42	0	1	29399
Reports external learning	0.44	0.5	0	1	29399
Woman	0.52	0.5	0	1	29399
Age	41.03	12.48	16	66	29399
Years of Education	14.58	2.12	12	20	29399
Potential Years of Experience	20.72	12.49	1	45	29217
Hourly Wage (Dollars of 2016)	27.34	37.93	1.23	2307.69	27767

B Appendix: Robustness of Fact 1

B.1 Robustness of Fact 1a

Table B.1: Share of Firms Providing CVT Courses and Other Types of CVT Activities

Country	CVT Courses		Other Types of CVT activities			
	Internal CVT Courses	External CVT Courses	Conferences, workshops or lectures	Guided on the Job Training	Job Rotation	Learning and quality circles
Germany	0.436	0.532	0.438	0.524	0.085	0.151
France	0.329	0.666	0.193	0.253	0.094	0.082
United Kingdom	0.418	0.532	0.399	0.655	0.204	0.218
Italy	0.263	0.403	0.263	0.260	0.100	0.035
Spain	0.186	0.564	0.198	0.354	0.108	0.127
Poland	0.134	0.219	0.123	0.189	0.054	0.022
Romania	0.117	0.157	0.084	0.139	0.069	0.051
Belgium	0.457	0.605	0.364	0.437	0.159	0.161
Portugal	0.212	0.360	0.216	0.389	0.064	0.099
Czech Rep.	0.391	0.611	0.243	0.367	0.048	0.088
Hungary	0.171	0.307	0.211	0.198	0.034	0.057
Sweden	0.600	0.724	0.499	0.611	0.355	0.103
Bulgaria	0.163	0.177	0.130	0.227	0.057	0.087
Denmark	0.485	0.640	0.513	0.452	0.157	0.189
Slovak Rep.	0.352	0.531	0.427	0.326	0.091	0.188
Finland	0.352	0.671	0.346	0.391	0.119	0.119
Norway	0.676	0.694	0.481	0.704	0.327	0.221
Latvia	0.124	0.271	0.145	0.475	0.059	0.056
Estonia	0.326	0.552	0.279	0.439	0.163	0.103
Cyprus	0.185	0.447	0.266	0.328	0.089	0.166
Luxembourg	0.473	0.580	0.390	0.444	0.170	0.197
Malta	0.284	0.322	0.320	0.435	0.142	0.136
Total	0.275	0.433	0.242	0.328	0.101	0.091

Notes: This table shows the proportion of firms in which workers participate in CVT Courses and other types of CVT activities for each country. Results are simple averages of respective proportions from three different CVT survey waves: CVTS3, CVTS4 and CVTS5. Weighting factors were used in order to calculate proportions for each wave. Last row "Total" is an average for all waves and all countries sampled.

Table B.2: Share of Workers' Hours spent in CVT Courses

Country	Share of working hours spent learning		
	CVT Courses	Internal CVT Courses	External CVT Courses
Germany	0.006	0.003	0.003
France	0.007	0.002	0.005
United Kingdom	0.006	0.003	0.003
Italy	0.006	0.003	0.003
Spain	0.007	0.003	0.004
Poland	0.005	0.002	0.003
Romania	0.007	0.005	0.003
Belgium	0.012	0.007	0.006
Portugal	0.010	0.005	0.006
Czech Rep.	0.007	0.003	0.004
Hungary	0.005	0.001	0.004
Sweden	0.007	0.004	0.003
Bulgaria	0.008	0.005	0.003
Denmark	0.009	0.004	0.005
Slovak Rep.	0.008	0.003	0.005
Finland	0.006	0.003	0.004
Norway	0.009	0.006	0.004
Latvia	0.005	0.002	0.003
Estonia	0.007	0.003	0.004
Cyprus	0.006	0.003	0.003
Luxembourg	0.012	0.007	0.005
Malta	.	.	.
Total	0.007	0.003	0.004

Notes: This table shows the share of working hours in the last calendar year in which workers participate in CVT courses for each country. Results are simple averages of respective proportions from three different CVT survey waves: CVTS3, CVTS4 and CVTS5. Proportions of hours are conditional on the firm having persons employed participating in CVT courses. Weighting factors were used in order to calculate proportions for each wave. Last row "Total" is an average for all waves and all countries sampled. Please note that information on hours spent learning is only available for CVT courses (both Internal and External), and not Other Types of CVT activities.

Table B.3: Share of Workers Taking Part in Other CVT Activities

Country	Conferences, workshops or lectures	Guided on the Job Training	Job Rotation	Learning and quality circles
Germany	0.208	0.385	0.084	0.186
France	0.145	0.329	0.226	0.197
United Kingdom	0.182	0.397	0.151	0.209
Italy	0.136	0.259	0.193	0.192
Spain	0.217	0.418	0.210	0.203
Poland	0.170	0.388	0.112	0.093
Romania	0.141	0.382	0.219	0.153
Belgium	0.253	0.327	0.132	0.151
Portugal	0.144	0.325	0.178	0.273
Czech Republic	0.210	0.518	0.140	0.156
Hungary	0.127	0.342	0.118	0.163
Sweden	0.338	0.398	0.243	0.177
Bulgaria	0.135	0.488	0.165	0.241
Denmark	0.236	0.327	0.126	0.281
Slovak Republic	0.168	0.427	0.103	0.393
Finland	0.179	0.230	0.106	0.126
Norway	0.240	0.322	0.200	0.275
Latvia	0.262	0.362	0.104	0.187
Estonia	0.158	0.319	0.140	0.203
Cyprus	0.232	0.313	0.157	0.213
Luxembourg	0.275	0.348	0.143	0.247
Total	0.184	0.378	0.171	0.198

Notes: This table shows the proportion of Workers taking part in other CVT activities for each country. Results are simple averages of respective proportions from two different CVT survey waves: CVTS3 and CVTS4. CVTS5 was excluded since it only provides 3 percentage ranges of persons employed participating in these activities, so comparability is impossible with former two surveys. Proportions of Workers are conditional on the firm having persons employed participating in other CVT activities. Weighting factors were used in order to calculate proportions for each wave. Last row "Total" is an average for all waves and all countries sampled.

Table B.4: Share of firms Providing Internal & External Learning Activities of Different Kinds

		External CVT Courses 0	External CVT Courses 1			External CVT Courses 0	External CVT Courses 1			External Other 0	External Other 1
Internal CVT Courses	0	0.43	0.25	Internal Other	0	0.37	0.19	Internal Other	0	0.48	0.08
Internal CVT Courses	1	0.07	0.25	Internal Other	1	0.13	0.31	Internal Other	1	0.22	0.22

Notes: These tables show the proportion of firms in the whole sample which reported having persons employed participating in different kinds of Internal and External CVT activities. We include: (a) Internal and External CVT Courses Only, (b) Other Types of Internal CVT Activities and External CVT Courses, and (c) Other Types of Internal and External CVT activities. Data from CVTS3, CVT4 and CVT5 surveys. Weighting factors were used in order to calculate proportions for each wave.

B.2 Robustness of Fact 1b

Table B.5: Correlation between Different Types of Learning Provision and Firm Size

Dep. Variables	CVT			Internal CVT			External CVT		
Firm size: 20–49	0.116***	0.117***	0.117***	0.110***	0.109***	0.109***	0.118***	0.119***	0.121***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Firm size: 50–99	0.200***	0.199***	0.199***	0.205***	0.200***	0.200***	0.227***	0.227***	0.228***
	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Firm size: 100–250	0.282***	0.280***	0.280***	0.312***	0.304***	0.304***	0.324***	0.324***	0.325***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Firm size: 251+	0.297***	0.293***	0.293***	0.381***	0.372***	0.372***	0.352***	0.349***	0.350***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Observations	286,321	286,321	286,321	286,321	286,321	286,321	286,321	286,321	286,321
R-squared	0.171	0.179	0.179	0.154	0.160	0.160	0.157	0.168	0.170
Country-Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry FE		Y	Y		Y	Y		Y	Y
Socioec. controls			Y			Y			Y

This table shows the coefficients from regressing a variable indicating whether firms report any kind of CVT, Internal CVT or External CVT activity on different firm size categories, where the omitted category encompasses firms with 5–19 Workers. We consider firms with 5 or more employees only since smaller firms encompass a very small part of the sample, and may thus be highly selected. Regressions are performed using observation weights provided in the surveys. Year fixed effects correspond to year of CVT survey fixed effects. Industry categories at the 1-digit level (NACE). Socioeconomic controls encompass log of per-capita GDP. Robust standard errors in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.6: Average Number and Share of Hours spent in CVT Courses by Firm Size, For All Firms

	CVT Courses	Internal CVT Courses	External CVT Courses
	Panel A: Average Share of Working Hours		
Small firms, 5–19 Workers	0.0027	0.0011	0.0016
Medium firms, 20–99 Workers	0.0037	0.0015	0.0022
Large firms, 100+ Workers	0.0058	0.0028	0.0030
Total	0.0040	0.0018	0.0023
	Panel B: Average Number of Hours per Worker		
Small firms, 5–19 Workers	4.861	2.006	2.855
Medium firms, 20–99 Workers	6.077	2.451	3.626
Large firms, 100+ Workers	9.491	4.505	4.987
Total	6.550	2.834	3.716

Notes: This table shows the average share and number of working hours per firm in which workers participated in CVT courses in the last calendar year. Shares and Total hours are calculated for all firms in each size category, and are presented in Panel A and Panel B, respectively. Results are simple averages of respective calculations from three different CVT survey waves: CVT3, CVT4 and CVT5. Weighting factors were used in order to properly calculate proportions for each wave. We consider firms with 5 or more employees only since smaller firms encompass a very small part of the sample, and may thus be highly selected.

Table B.7: Correlation between Hours per Worker spent on Types of Learning and Firm Size

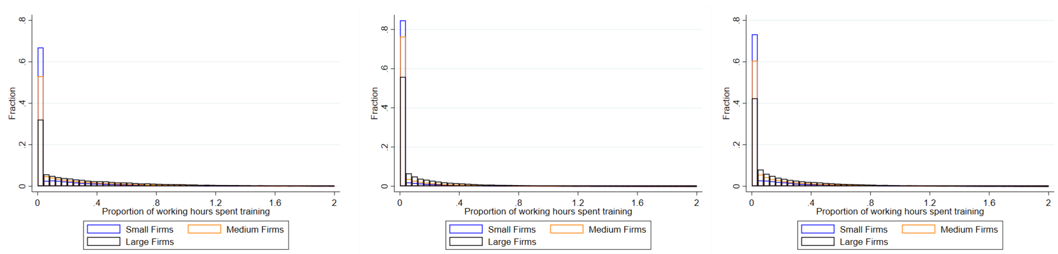
Dep. Variables	CVT			Internal CVT			External CVT		
Firm size: 20–49	0.703**	0.737**	0.759**	0.432*	0.424*	0.429*	0.270*	0.313*	0.330**
	(0.314)	(0.316)	(0.317)	(0.247)	(0.247)	(0.249)	(0.164)	(0.165)	(0.165)
Firm size: 50–99	1.476***	1.432***	1.450***	0.773***	0.710***	0.714***	0.703***	0.723***	0.736***
	(0.280)	(0.284)	(0.284)	(0.192)	(0.193)	(0.193)	(0.169)	(0.173)	(0.172)
Firm size: 100–250	3.111***	3.034***	3.047***	1.704***	1.605***	1.608***	1.407***	1.429***	1.439***
	(0.319)	(0.321)	(0.321)	(0.204)	(0.204)	(0.204)	(0.200)	(0.202)	(0.202)
Firm size: 251+	3.615***	3.427***	3.441***	2.567***	2.413***	2.417***	1.048***	1.013***	1.024***
	(0.266)	(0.285)	(0.283)	(0.176)	(0.182)	(0.181)	(0.149)	(0.163)	(0.162)
Observations	273,870	273,870	273,870	273,870	273,870	273,870	273,870	273,870	273,870
R-squared	0.015	0.021	0.022	0.008	0.010	0.010	0.016	0.023	0.023
Country-Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry Controls		Y	Y		Y	Y		Y	Y
Socioec. controls			Y			Y			Y

This table shows the coefficients from regressing yearly number of hours per worker spent on any kind of CVT, Internal CVT or External CVT activity on different firm size categories, where the omitted category encompasses firms with 5–19 Workers. We consider firms with 5 or more employees only since smaller firms encompass a very small part of the sample, and may thus be highly selected. Dependent variables were constructed for all firms in each size category, including those reporting no learning activities and thus zero hours of learning. Regressions are performed using observation weights provided in the surveys. Year fixed effects correspond to year of CVT survey fixed effects. Industry categories at the 1-digit level (NACE). Socioeconomic controls encompass log of per-capita GDP. Robust standard errors in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure B.1: Histograms of Share of Hours spent in CVT Courses by Firm Size

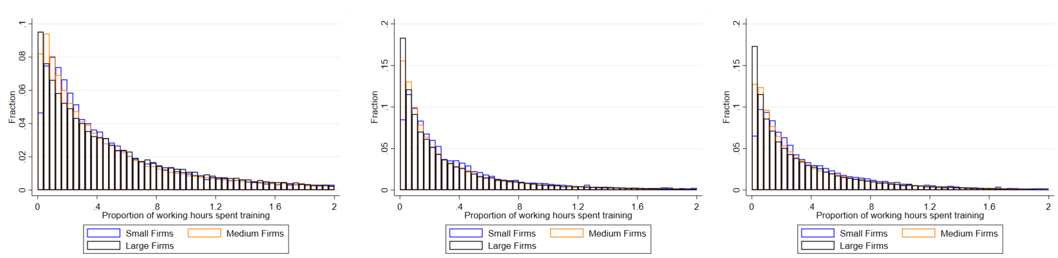
Panel A: For all Firms

(a) CVT Courses (b) Internal CVT courses. (c) External CVT courses.



Panel B: For all Firms, with zero share of learning hours not plotted)

(d) CVT Courses (e) Internal CVT courses. (f) External CVT courses.



Notes: These figures show histograms of the share of working hours per firm in which workers participate in CVT courses for the entire distribution (Panel A), and with zero learning hours not plotted (Panel B). Data from three different CVT survey waves: CVT3, CVT4 and CVT5. Weighting factors were used in order to properly calculate proportions for each wave. We consider firms with 5 or more employees only since smaller firms encompass a very small part of the sample, and may thus be highly selected.

C Appendix: Robustness of Fact 2

Table C.1: Correlations between different sources of learning and potential experience

Dep. Variables	Internal Learning			External Learning		
Germany						
Potential Yrs. Experience	-0.0104***	-0.00630***	-0.00312***	0.0105***	0.00574***	0.00216***
	(0.000672)	(0.000781)	(0.000798)	(0.000552)	(0.000443)	(0.000520)
Potential Yrs. Experience^2	0.000167***	8.37e-05***	4.34e-05***	-0.000247***	-0.000124***	-6.66e-05***
	(1.37e-05)	(1.61e-05)	(1.69e-05)	(1.14e-05)	(9.08e-06)	(1.07e-05)
Constant	0.439***	0.337***	0.340***	0.602***	0.697***	0.702***
	(0.00734)	(0.0141)	(0.0141)	(0.00592)	(0.00820)	(0.00997)
Observations	109,478	69,495	36,813	173,391	126,129	85,280
R-squared	0.006	0.129	0.071	0.005	0.181	0.191
Year FE		Y	Y		Y	Y
Demographic Controls		Y	Y		Y	Y
Worker type FE		Y	Y		Y	Y
Industry FE		Y	Y		Y	Y
Occupation FE		Y	Y		Y	Y
Firm size FE		Y	Y		Y	Y
Wage Controls			Y			Y
USA						
Potential Yrs. Experience	-0.00489***	-0.00943***	-0.00999***	0.0153***	0.0110***	0.0114***
	(0.00120)	(0.00103)	(0.00105)	(0.00141)	(0.00139)	(0.00142)
Potential Yrs. Experience^2	-4.13e-05*	9.76e-05***	0.000110***	-0.000277***	-0.000180***	-0.000188***
	(2.49e-05)	(2.14e-05)	(2.17e-05)	(3.05e-05)	(3.02e-05)	(3.08e-05)
Constant	0.352***	0.270***	0.262***	0.289***	0.278***	0.276***
	(0.0119)	(0.0141)	(0.0146)	(0.0129)	(0.0176)	(0.0180)
Observations	29,217	29,217	27,585	29,217	29,217	27,585
R-squared	0.040	0.228	0.224	0.013	0.073	0.075
Demographic Controls		Y	Y		Y	Y
Worker type FE		Y	Y		Y	Y
Industry FE		Y	Y		Y	Y
Occupation FE		Y	Y		Y	Y
Wage Controls			Y			Y

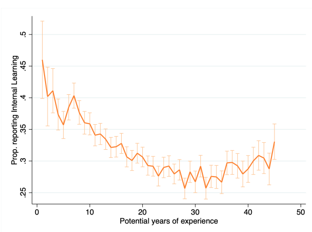
Internal learning, external learning and potential years of experience described in text for both countries. All regressions weighted using observation weights provided in the surveys. *Germany:* Year fixed effects correspond to year of survey fixed effects. Demographic controls include educational attainment level and gender. Worker type categories include laborer, private employee, government employee, self-employed, freelancer, or family caregiver. Industry categories at the 1-digit level. Occupation categories at the 2-digit level (ISCO 88). Firm size is a categorical variable indicating whether the firm where the worker works at has less than 4 workers, 5–9 workers, 10–49 workers, 50–99 workers, 100–499 workers, 500–999 workers and 1000 or more workers. Wage controls include the current hourly wage for the worker. *USA:* Demographic controls include educational attainment level, race, census region, and gender. Worker type categories include private employee, government employee, self-employed, or working without pay. Industry and Occupation categories at the 2-digit level (ACS 2015). Wage controls include the current hourly wage for the worker. Robust standard errors in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

C.1 Decomposition[55]

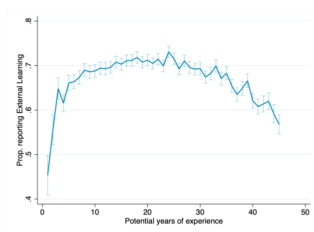
Figure C.1: Sources of learning throughout workers' lifecycles by one-year experience bins

Germany **USA**

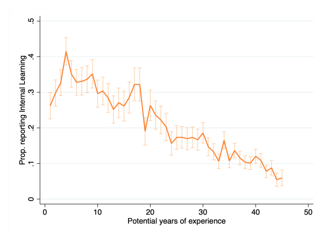
(a) Internal Learning



(b) External Learning



(c) Internal Learning



(d) External Learning

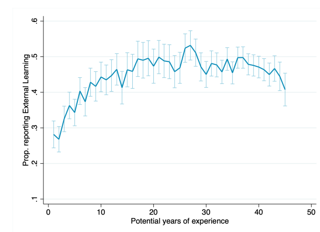
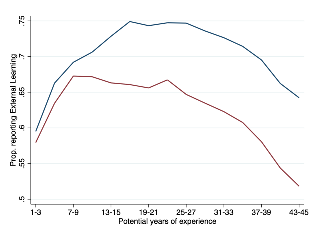


Figure C.2: Sources of learning throughout workers' lifecycles by gender

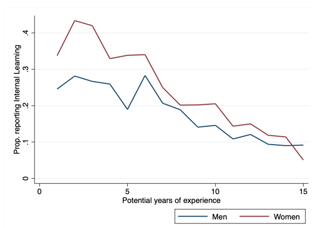
Germany **USA**

(a) Internal Learning

(b) External Learning



(c) Internal Learning



(d) External Learning

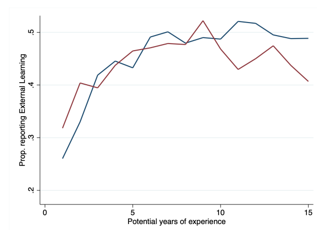
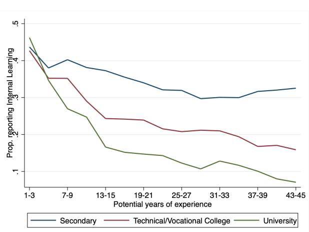


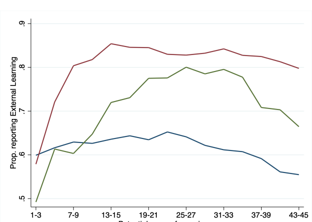
Figure C.3: Sources of learning throughout workers' lifecycles by educational level

Germany **USA**

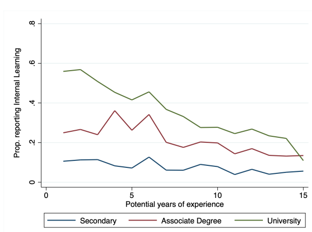
(a) Internal Learning



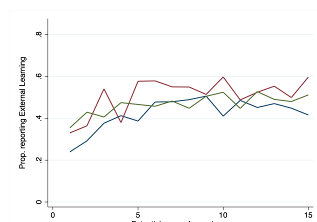
(b) External Learning



(c) Internal Learning



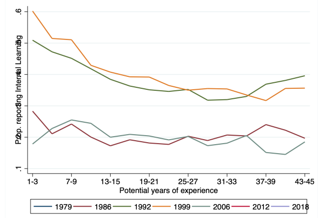
(d) External Learning



[55] 95% confidence intervals included for some plots, but omitted in plots that consider several groups for clarity.

Figure C.4: Sources of learning throughout workers' lifecycles by wave in Germany

(a) Internal Learning



(b) External Learning

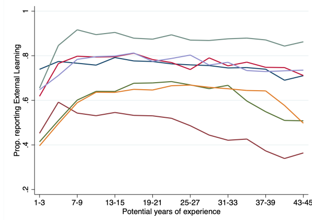
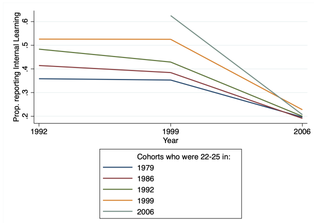
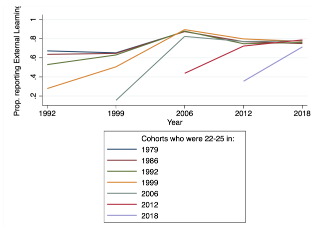


Figure C.5: Sources of learning throughout workers' lifecycles by cohort in Germany

(a) Internal Learning



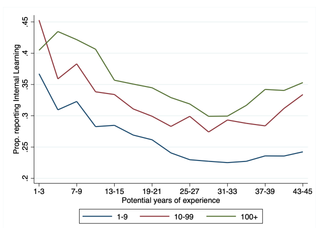
(b) External Learning



Notes: Due to comparability and data availability issues across waves, these figures rely on learning data from 1992 to 2006 (internal learning), and 1992 to 2018 (external learning). Please see Appendix A.2 for details on the learning measures in Germany, and comparability issues across waves.

Figure C.6: Sources of learning throughout workers' lifecycles by firm size in Germany

(a) Internal Learning



(b) External Learning

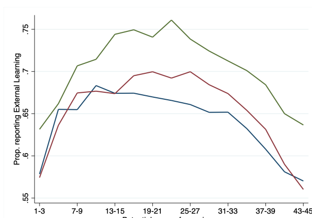
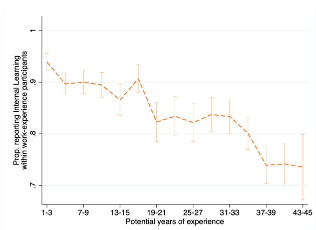
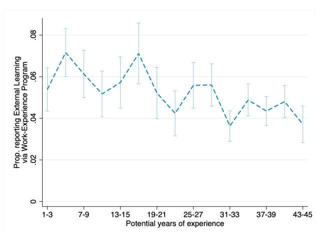


Figure C.7: Decomposing different types of internal and external learning in the US

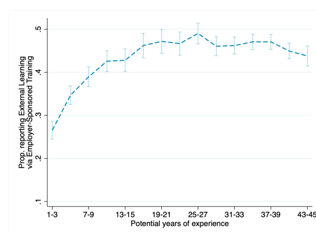
(a) Prevalence of internal learning within work-experience program participants



(b) External Learning via Participation in Work-Experience Program



(c) External Learning via Certificate in Employer-Sponsored Training

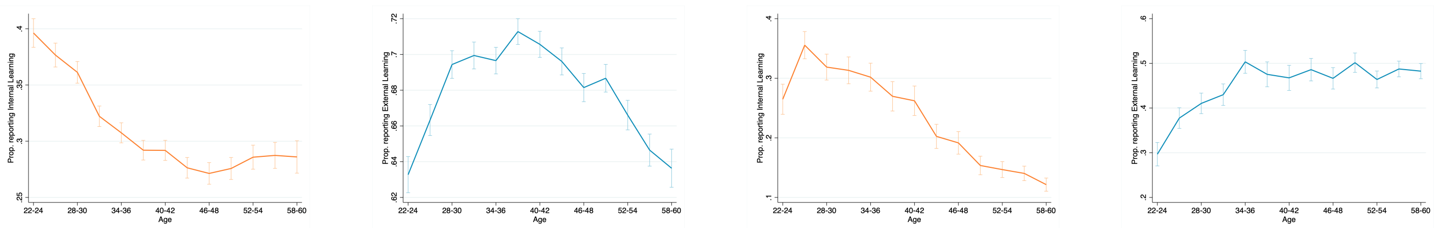


C.2 Alternate Experience Variables

Figure C.8: Sources of learning by age

Germany **USA**

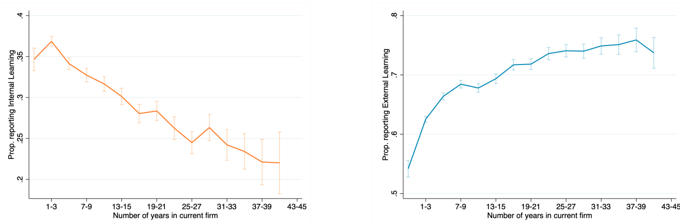
(a) Internal Learning (b) External Learning (c) Internal Learning (d) External Learning



Notes: We exclude individuals who below 21 or over 60, since there are very few such observations. 95% confidence intervals plotted.

Figure C.9: Sources of learning by tenure in the current firm in Germany

(a) Internal Learning (b) External Learning



Notes: We exclude individuals with more than 42 years of tenure, since there are very few such observations (we choose 42 specifically since it is one 3-year bin below our 45 years cut for potential experience). 95% confidence intervals plotted.

Table C.2: Correlations between different sources of learning and current tenure

Dep. Variables	Internal Learning				External Learning			
Germany								
Yrs. w/ Current Emp.	-0.00591***	-0.00422***	-0.00311***	-0.00210**	0.00891***	0.00474***	0.00442***	0.00650***
	(0.000560)	(0.000669)	(0.000901)	(0.000968)	(0.000473)	(0.000514)	(0.000629)	(0.000487)
Yrs. w/ Current Emp.^{2}	5.63e-05***	4.53e-05**	6.77e-05**	4.98e-05*	-0.000145***	-9.65e-05***	-0.000103***	-0.000106***
	(1.67e-05)	(1.99e-05)	(2.74e-05)	(2.91e-05)	(1.39e-05)	(1.48e-05)	(1.82e-05)	(0.0000138)
Constant	0.372***	0.283***	0.342***	1.102***	0.609***	0.691***	0.671***	0.7932**
	(0.00354)	(0.0125)	(0.0162)	(0.0198)	(0.00304)	(0.00963)	(0.0128)	(0.3919)
Observations	102,761	68,877	36,703	36,703	165,275	124,372	84,222	84,222
R-squared	0.007	0.127	0.077	0.079	0.009	0.193	0.206	0.1957
Year FE		Y	Y	Y		Y	Y	Y
Demographic Controls		Y	Y	Y		Y	Y	Y
Worker type FE		Y	Y	Y		Y	Y	Y
Industry FE		Y	Y	Y		Y	Y	Y
Occupation FE		Y	Y	Y		Y	Y	Y
Firm size FE		Y	Y	Y		Y	Y	Y
Wage Controls			Y	Y			Y	Y
Age FE				Y				Y

Internal learning, external learning and years with current employer described in text. All regressions weighted using observation weights provided in the surveys. *Germany:* Year fixed effects correspond to year of survey fixed effects. Demographic controls include educational attainment level and gender. Worker type categories include laborer, private employee, government employee, self-employed, freelancer, or family caregiver. Industry categories at the 1-digit level. Occupation categories at the 2-digit level (ISCO 88). Firm size is a categorical variable indicating whether the firm where the worker works at has less than 4 workers, 5–9 workers, 10–49 workers, 50–99 workers, 100–499 workers, 500–999 workers and 1000 or more workers. Wage controls include the current hourly wage for the worker. Robust standard errors in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

C.3 Correlation between Sources of Learning

We examine the correlation between internal and external sources of learning across workers in both settings. We do this, because a strong positive correlation between the two sources of learning studied could potentially indicate a selection mechanism, where certain workers are more likely to engage in learning overall potentially independent of experience.

Table C.3: Correlation between different sources of learning

Dep. Variables	Internal Learning		
Germany			
External Learning	-0.203***	-0.154***	-0.149***
	(0.00354)	(0.00456)	(0.00606)
Constant	0.438***	1.176***	1.227***
	(0.00301)	(0.0159)	(0.0200)
Observations	109,478	69,495	36,813
R-squared	0.045	0.150	0.101
Year FE		Y	Y
Demographic Controls		Y	Y
Age FE		Y	Y
Worker type FE		Y	Y
Industry FE		Y	Y
Occupation FE		Y	Y
Firm size FE		Y	Y
Wage Controls			Y
USA			
External Learning	0.0949***	0.0903***	0.0936***
	(0.00773)	(0.00697)	(0.00710)
Constant	0.185***	0.0924***	0.0807***
	(0.00474)	(0.0105)	(0.0114)
Observations	29,399	29,398	27,766
R-squared	0.013	0.240	0.237
Demographic Controls		Y	Y
Age FE		Y	Y
Worker type FE		Y	Y
Industry FE		Y	Y
Occupation FE		Y	Y
Wage Controls			Y

Internal learning and external learning described in text for both countries. All regressions weighted using observation weights provided in the surveys. *Germany:* Year fixed effects correspond to year of survey fixed effects. Demographic controls include educational attainment level and gender. Worker type categories include laborer, private employee, government employee, self-employed, freelancer, or family caregiver. Industry categories at the 1-digit level. Occupation categories at the 2-digit level (ISCO 88). Firm size is a categorical variable indicating whether the firm where the worker works at has less than 4 workers, 5–9 workers, 10–49 workers, 50–99 workers, 100–499 workers, 500–999 workers and 1000 or more workers. Wage controls include the current hourly wage for the worker. *USA:* Demographic controls include educational attainment level, race, census region, and gender. Worker type categories include private employee, government employee, self-employed, or working without pay. Industry and Occupation categories at the 2-digit level (ACS 2015). Wage controls include the current hourly wage for the worker. Robust standard errors in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

To this end, we study the correlation between the by regressing the former on the latter, and including a battery of controls and fixed effects. The results are summarized in Table C.3. We find a significant negative correlation external and internal learning which is robust to controlling for several demographic and firm-level variables in Germany. For the US, we find a positive and significant correlation. However, this positive correlation likely corresponds to the fact that this learning occurs in the context of a "work experience program", namely a job with learning attributes, such as an internship, co-op, practicum, clerkship, externship, residency, clinical experience, apprenticeship, or similar. As such, several sources of learning are potentially more likely to coexist.

C.4 Evidence from Chile

We now consider the robustness of the lifecycle patterns of learning documented when we consider data from a developing country: Chile. We use data from 4 waves of the Social Protection Survey (Encuesta de Proteccion Social, or EPS) conducted in 2002, 2004, 2006 and 2009. This data features a longitudinal structure, and has a sample of about 16,000 subjects. The data includes information on work histories, income and wealth, education, health, social security, marital history, household information and on-the-job learning.

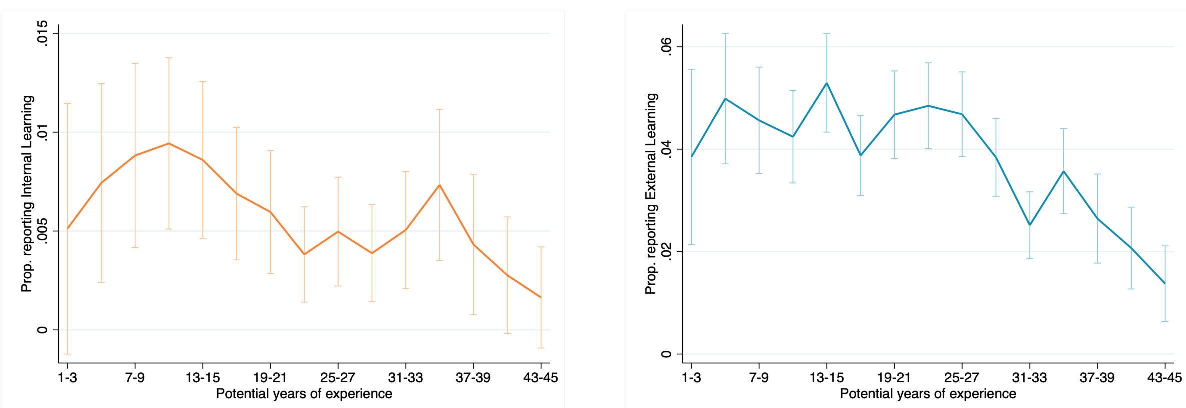
Using this data, we construct measures of internal and external learning which capture individuals who reported attending a course that was imparted by the employer or an external agent (such as a training firm, private teacher, a nonprofit organization, etc), respectively during the last year. Please note that this data asks individuals to consider only the three most important formal courses attended in the last few years, and thus leaves out both informal learning activities and courses taken outside of the three most important ones. This contrasts with the learning variables constructed for the US and Germany, which instead ask whether individuals received any form of instruction, or learned their skills used for their jobs from colleagues or external agents. We also construct measures of Potential Experience in this Chilean data using age and educational attainment: $Potential\ Experience = Age - Years\ of\ Schooling - 6$. As before, we limit our sample to individuals who are currently employed, and have potential experience between 1 and 45 years. In addition, we limit our analysis to individuals who appear in all four waves considered, since the questions in the survey vary between panel and non-panel individuals. Summary statistics, graphs and regressions are weighted using observation weights provided in the surveys. In Appendix C.4.2 we display some key summary statistics for the individuals in this data. For further details on this data and the construction of our variables of interest please see Appendix C.4.1.

In Figure C.10, we plot how the prevalence of workers reporting attending courses imparted by the employer or external actors changes with workers' potential experience in Chile. Consistent with the results found for the US and Germany, we find that (1) the prevalence of internal learning decreases with worker potential experience; and (2) the prevalence of external learning has an inverted U-shape in worker potential experience. These results are slightly noisier than those in the baseline settings, however, since the proportion of individuals reporting learning activities is much lower likely due to the exclusion of both informal learning activities and less-important courses. In Table C.5, we show these correlations follow the expected signs, but are generally not statistically significant at conventional levels, likely due to the narrow view of internal and external learning in this data.

Figure C.10: Prevalence of different sources of learning throughout workers' lifecycles in Chile

Chile

(a) Internal Learning (b) External Learning



Notes: 95% confidence intervals plotted.

C.4.1 Data and Variable Construction

The EPS survey was first deployed in 2002, and repeated on 2004, 2006, 2009, 2012, 2015 and 2020. The on-the-job training module has been included in every wave, and asks about the main on-the-job training courses attended in the last few years, along with questions on how long these courses lasted, who paid for them, and the usefulness of the concepts learned. Several issues prevent us using the data in the latter three waves, however. First, the 2012 wave has a high degree of error, and is not considered suitable for statistical analysis. Second, the on-the-job training module changed substantially in the 2015 and 2020 waves. Previous waves ask individuals to consider the three most important formal courses received and collects information on course duration, while the latter two waves asks only about the most important course and does not collect course duration. The data was collected via in-person visits to subjects' homes, and focused on individuals 18 years of age and older. The data is representative of the 18 and older Chilean population.

Skill Acquisition and Potential Experience

In this section, we provide further information on the construction of our key skill acquisition variables of interest, and also of potential experience in Chile. To construct the variables that capture whether the worker experiences "internal learning" and "external learning", we rely on the following questions and guidelines.

- "Internal Learning": is a binary variable that indicates whether an individual reported attending a course during the last year that was (1) imparted by the employer or firm's parent company; and (2) very or somewhat useful for their work; among the three main on-the-job courses attended in the last few years reported by the worker.
- "External Learning": is a binary variable that indicates whether an individual reported attending a course during the last year that was (1) imparted by a training firm, an equipment manufacturer, a nonprofit organization, the municipality, a private teacher, or some other institution; and (2) very or somewhat useful for their work; among the three main on-the-job courses attended in the last few years reported by the worker.
- Potential experience was constructed as: $Age - Years\ of\ Schooling - 6$
  - $Years\ of\ Schooling$ was constructed by mapping the educational attainment to the corresponding years of schooling.

Other Variables

Hourly wages are constructed using the monthly wage and weekly regular hours data. Wages are deflated using the Chilean CPI with base 2015.

C.4.2 Summary Statistics and Additional Results

Table C.4: Summary Statistics in Chile

	(1) Mean	(2) Std. Dev.	(3) Minimum	(4) Maximum	(5) # Obs.
Reports internal learning	0	0.06	0	1	26929
Reports external learning	0.03	0.17	0	1	26929
Woman	0.39	0.49	0	1	26929
Age	39.58	10.56	16	70	26929
Years of Education	11.82	3.17	0	19	26929
Potential Years of Experience	21.76	11.43	1	45	26929
Hourly Wage (Chilean Pesos of 2015)†	1402.57	8655.09	0	334642.78	25191
Firm size 1-9	0.39	0.49	0	1	26383
Firm size 10-99	0.24	0.43	0	1	26383
Firm size 100+	0.37	0.48	0	1	26383

†: Hourly wage winsorized to exclude those above the 99% percentile.

Table C.5: Correlations between different sources of learning and potential experience in Chile

Dep. Variables	Internal Learning				External Learning			
Potential Yrs. Experience	-0.0001	-0.0002	-0.0002	-0.0007	0.0005	0.001	0.001	0.0009
	(0.0001)	(0.0001)	(0.0002)	(0.0008)	(0.0006)	(0.0007)	(0.0007)	(0.002)
Potential Yrs. Experience^2	7.97e-07	4.00e-06	5.22e-06*	2.30e-05*	-2.12e-05*	-1.70e-05	-1.68e-05	3.30e-05
	(2.80e-06)	(2.97e-06)	(3.17e-06)	(1.30e-05)	(1.19e-05)	(1.24e-05)	(1.31e-05)	(2.72e-05)
Constant	0.006***	0.006**	0.006**	-0.0002	0.031***	-0.004	-0.007	-0.027
	(0.002)	(0.003)	(0.003)	(0.015)	(0.007)	(0.009)	(0.01)	(0.054)
Observations	26,929	26,072	24,445	22,611	26,929	26,072	24,445	22,611
R-squared	0.000	0.006	0.007	0.341	0.001	0.024	0.025	0.369
Year FE		Y	Y	Y		Y	Y	Y
Demographic Controls		Y	Y	Y		Y	Y	Y
Worker type FE		Y	Y	Y		Y	Y	Y
Industry FE		Y	Y	Y		Y	Y	Y
Firm size FE		Y	Y	Y		Y	Y	Y
Wage Controls			Y	Y			Y	Y
Individual FE				Y				Y

Internal learning, external learning and potential years of experience described in text. All regressions weighted using observation weights provided in the survey. *Chile:* Year fixed effects correspond to year of survey fixed effects. Demographic controls include educational attainment level and gender. Worker type categories include employer, own-account, private employee, government employee, live-in domestic worker, live-out domestic worker, unpaid family worker or military employee. Industry categories at the 1-digit level (note, occupation codes not available in all survey waves). Firm size is a categorical variable indicating whether the firm where the worker works at has 1 worker, 2–9 workers, 10–19 workers, 20–49 workers, 50–99 workers, 100–199 workers, 200–499 workers and 500 or more workers. Wage controls include the current hourly wage for the worker, winsorized to exclude those above the 99% percentile. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

D Appendix: Benchmark Model

D.1 Worker's Expected Utility

Expected remaining lifetime utility in period τ for an individual with human capital level h is given by:

$$U_\tau(h) = E_\tau \sum_{t=\tau}^{\infty} \beta^{t-\tau} I_t(h_t)$$

Where $I_t(h_t)$ represents the payout received by a worker with human capital h in period t. Given $t > \tau$, let Z_t denote a stochastic variable with two different possible outcomes:

$$Z_t = \begin{cases} E_\tau(I_t(h_t)) \text{ if } X > t \text{ (person is still alive at } t) \\ 0 \text{ if } X \leq t \text{ (person is dead at } t) \end{cases}$$

Then:

$$U_\tau(h) = \sum_{t=\tau}^{\infty} \beta^{t-\tau} E_t(Z_t) = \sum_{t=\tau}^{\infty} \beta^{t-\tau} P(X > t) E_t(I_t(h_t)) + P(X < t) \times 0 = \sum_{t=\tau}^{\infty} \beta^{t-\tau} (1-\delta)^{t-\tau} E_t(I_t(h_t))$$

We assume that the worker is making all time-t choices conditional on time t information, and thus $I_t(h_t)$ is conditional on time-t information, namely h_t and choices made that period.

Notice also that the expected lifetime is $E(X) = \sum_{t=1}^{\infty} \delta t(1-\delta)^{t-1} = \delta \sum_{t=0}^{\infty} t(1-\delta)^t = \frac{1}{\delta}$. Although there is no upper bound on possible lifetime, the probability of becoming very old is extremely small for values of δ consistent with a realistic life expectancy.

D.2 Worker's Lifetime Problem

We can write the lifetime problem solved by each worker i born in period τ as:

$$U = \max_{\{l_{i,t}, s_{i,t}, c_{i,t}, t_{i,t}, h_{i,t+1}} \sum_{t=\tau}^{\infty} \left(\beta(1-\delta)\right)^{t-\tau} E_t(I(h_{i,t}))$$

Subject to:

$$1 = \begin{cases} \mathbb{1}_{l_i=1} + \mathbb{1}_{s_i=1} + \mathbb{1}_{c_i=1} \text{ if production sector} \\ \mathbb{1}_{n_i=1} \text{ if training sector} \end{cases}$$

$$E_t(I(h_{i,t})) = \begin{cases} w(h_{i,t}) \text{ if } l_{i,t} = 1 \\ 0 \text{ if } c_{i,t} = 1 \\ -(1 - F_n(h_{i,t}))q \text{ if } s_{i,t} = 1 \\ E(w_n(h_{i,t})) \text{ if } n_{i,t} = 1 \end{cases}$$

$$h_{i,j+1} = h_{i,j} + \begin{cases} X_{l,i,j} & \text{if work in production} \\ X_{c,i,j} & \text{if learn through colleagues} \\ X_{s,i,j} & \text{if train} \\ X_{n,i,j} & \text{if work as trainer} \end{cases}$$

Where X_l, X_c and X_s are stochastic binary Bernoulli variables with probability of success of ϵ, $p_c(h_{i,j})$, $p_s(h_{i,j})$, and ϵ respectively.

A worker with human capital h_m therefore has the following Bellman equation in every period:

$$EV(h_m) = \underset{l,s,c,n}{max} EV_m$$

Where:

$$EV_m = \begin{cases} w(h_m) + \beta(1-\delta)\left[(1-\epsilon)EV(h_m) + \epsilon EV(h_{m+1}))\right] \text{ if } l = 1 \\ 0 + \beta(1-\delta)\left[p_c(h_m)EV(h_{m+1}) + (1-p_c(h_m))EV(h_m)\right] \text{ if } c = 1 \\ -p_s(h_m)q + \beta(1-\delta)\left[p_s(h_m)EV(h_{m+1}) + (1-p_s(h_m))EV(h_m)\right] \text{ if } s = 1 \\ E(w_n(h_i)) + \beta(1-\delta)\left[(1-\epsilon)EV(h_m) + \epsilon EV(h_{m+1})\right] \text{ if } n = 1 \end{cases}$$

D.3 Stationary Equilibrium

Definition D.1. The model's stationary equilibrium can be defined recursively, and is characterized by:

1. Value function $V(F_l, F_c, F_s, F_n, , N_l, N_c, N_s, N_n, h)$, and policy functions $l(F_l, F_c, F_s, F_n, N_l, N_c, N_s, N_n, h)$, $s(F_l, F_c, F_s, F_n, N_l, N_c, N_s, N_n, h)$, $c(F_l, F_c, F_s, F_n, h)$, $t(F_l, F_c, F_s, F_n, N_l, N_c, N_s, N_n, h)$ for workers of each human capital level.

2. Policy function vector of labor inputs $H^d(F_l, F_c, F_s, F_n, N_l, N_c, N_s, N_n, h)$ for the production firm.

3. Wage function $w(F_l, F_c, F_s, F_n, N_l, N_c, N_s, N_n, h)$

4. Cost of external training services $q(F_l, F_c, F_s, F_n, N_l, N_c, N_s, N_n)$

5. Perceived laws of motion for the distribution of workers actively producing in the firm $F_l'(h) = \hat{F}_l(h)$, the distribution of internal learners $F_c'(h) = \hat{F}_c(h)$, the distribution of external learners $F_s'(h) = \hat{F}_s(h)$, and the distribution of trainers in the training sector $F_n'(h) = \hat{F}_n(h)$.

6. Perceived laws of motion for the mass of workers actively producing in the firm $N_l = \hat{N}_l$, the mass of internal learners $N_c = \hat{N}_c$, the mass of external learners $N_s = \hat{N}_s$, and the mass of trainers working in the training sector $N_n = \hat{N}_n$.

   Such that:

- Given wages (3), cost of external training services (4), and the perceived laws of motion (5) and (6), each worker is maximizing lifetime utility by choosing (1).

- Given wages (3), firms are maximizing profits by choosing (2) and support free entry.

- Labor market clears:

$$H^d(F_c, F_n, F_s, N_l, N_c, N_s, N_n) = \begin{pmatrix} h_1 \int_{i|h_i=h_1} l(h_1)di \\ h_2 \int_{i|h_i=h_2} l(h_2)di \\ \cdots \end{pmatrix}$$

- Training services market clears:

$$\sum_{h_m=h_1}^{h_M} \int_{i|h_i=h_m} s(h_i)di = \sum_{h_m=h_1}^{h_M} \int_{i|h_i=h_m} n(h_i)di$$

- Perceptions are correct (distributions and masses equal what workers are doing).[56]

$$\hat{F}_l(h) = \frac{\sum_{h_m=h_1}^{h_M} \int_{i|h_i=h_m} l(h_m)di}{\sum_{h_m=h_1}^{h_M} \int_{i|h_i=h_m} l(h_m)di}$$

$$\hat{F}_c(h) = \frac{\sum_{h_m=h_1}^{h_M} \int_{i|h_i=h_m} s(h_m)di}{\sum_{h_m=h_1}^{h_M} \int_{i|h_i=h_m} c(h_m)di}$$

$$\hat{F}_s(h) = \frac{\sum_{h_m=h_1}^{h_M} \int_{i|h_i=h_m} s(h_m)di}{\sum_{h_m=h_1}^{h_M} \int_{i|h_i=h_m} s(h_m)di}$$

$$\hat{F}_n(h) = \frac{\sum_{h_m=h_1}^{h_M} \int_{i|h_i=h_m} n(h_m)di}{\sum_{h_m=h_1}^{h_M} \int_{i|h_i=h_m} n(h_m)di}$$

$$\hat{N}_l = \sum_{h_m=h_1}^{h_M} \int_{i|h_i=h_m} l(h_i)di$$

$$\hat{N}_c = \sum_{h_m=h_1}^{h_M} \int_{i|h_i=h_m} c(h_i)di$$

$$\hat{N}_s = \sum_{h_m=h_1}^{h_M} \int_{i|h_i=h_m} s(h_i)di$$

$$\hat{N}_n = \sum_{h_m=h_1}^{h_M} \int_{i|h_i=h_m} n(h_i)di$$

D.4 Assumptions to Characterize Equilibrium

Assumption 1. *y is constant returns to scale, symmetric across inputs, and increasing and concave in each input.*

Assumption 2. *The exogenous probability of learning is low enough such that $\epsilon < \frac{\delta}{1-\delta}$.*

[56] These also guarantee equilibrium is stationary, as distributions and masses will not be shifting over time.

Assumption 3. *The rise in the level human capital as we climb the human capital ladder is large enough so that:*

$$\log \left[Q^{-1} \left(\frac{h_m(1-\beta(1-\delta)(1-\frac{\delta((1-\delta)\epsilon)^{M-2}}{(1-(1-\delta)(1-\epsilon))^{M-2}}))}{h_{m+1}\beta(1-\delta)\frac{\delta((1-\delta)\epsilon)^{M-2}}{(1-(1-\delta)(1-\epsilon))^{M-2}}} \right) \right] > \log(h_{m+1}) - \log(h_m)$$

Where if G denotes the first derivative of y with respect to one input: $G(x) = \frac{y(x,z,\ldots)}{dx}$, Q denotes the ratio of two of these: $= Q\left(\frac{x}{y}\right) = \frac{G(x)}{G(y)}$.

Assumption 4. *The equilibrium allocation features external learning as a mode of human capital acquisition, and thus includes a positive mass of both trainers and external learners: $\exists m_1, m_2 \mid f_s(m_1) > 0,\ f_t(m_2) > 0$.*

Assumption 5. *The mass of trainers is positive from $h_{\underline{m}}$ to h_M, where $h_{\underline{m}}$ is the lowest human capital level where trainers locate: $f_t(h_m) > 0\ \forall h_m > h_{\underline{m}}$.*

D.5 Additional Results of Benchmark Model

Lemma 1. *Within each human capital step h_m there is a positive mass of output workers ($f_l(h_m) > 0\ \forall h_m$), and for all human capital steps except the last, there is a positive mass of learners ($f_c(h_m) + f_s(h_m) > 0\ \forall h_m$).*

Proof: See Appendix D.6.1.

Lemma 2. *Within each human capital step h_m where learning occurs ($m < M$), all learners are of the same type, meaning they either learn internally or externally: $f_c(h_m) > 0 \Rightarrow f_s(h_m) = 0$, and $f_s(h_m) > 0 \Rightarrow f_c(h_m) = 0$.*

Proof: See Appendix D.6.2.

Lemma 3. *The production wage increases with human capital, such that for each h_m $w(h_m) > w(h_{m-1})$.*

Proof: See Appendix D.6.3.

Lemma 4. *If the mass of trainers is positive at a given human capital level, we must then have that there is a positive mass of external learners in the human capital level below: $f_t(h_m) > 0 \Rightarrow f_s(h_{m-1}) > 0$.*

Proof: See Appendix D.6.4.

Lemma 5.

1. *For every human capital level where the mass of external learners is positive, we must have a strictly higher human capital level with a positive mass of trainers:* $f_s(h_m) > 0 \Rightarrow \exists h_{m_1} > h_m \mid f_t(h_{m_1}) > 0$.

2. *At and beyond the maximum human capital level where the mass of trainers is positive, we must have no external learners: Let* $h_{\bar{m}} = max\{h_m \,|\, f_t(h_m) > 0\}$. *Then,* $\forall h_m \geq \; h_{\bar{m}},\ f_s(h_m) = 0$.

Proof: See Appendix D.6.5.

Lemma 6. *At the lowest level of human capital,* h_1*, individuals engage in internal learning rather than external learning:* $f_c(h_1) > 0$*. Therefore, the lowest human capital level at which individuals engage in external learning is larger than* h_1*:* $h_{m*} = min\{h_m \mid f_s(h_m) > 0\} > h_1$.

Proof: See Appendix D.6.6.

Lemma 7.

1. *The mass of internal learners is positive from* h_1 *to* h_{m*-1}*:* $f_c(h_m) > 0\ \forall h_m \in [h_1, h_{m*-1}]$.

2. *The lowest human capital level at which individuals work as trainers is larger than* h_2*:* $h_{\underline{m}} = min\{h_m \mid f_t(h_m) > 0\} > h_2$.

3. *The mass of external learners is positive from* h_{m*} *to* $h_{\underline{m}-1}$*:* $f_s(h_m) > 0\ \forall h_m \in [h_m, h_{\underline{m}-1}]$.

Proof: See Appendix D.6.7.

D.6 Proofs

D.6.1 Proof of Lemma 1

Assumption 1 ensures that the marginal product of labor, and therefore the wage tends to infinity as the amount of labor goes to zero. Therefore, there is a positive mass of output workers, $f_l(h_m) > 0\ \forall h_m$, within each level of human capital.

Our setup also implies that for all human capital steps except the last, there is a positive mass of learners ($f_c(h_m) + f_s(h_m) > 0\ \forall h_m$). To prove this we proceed by contradiction. Assume that at some human capital level h_{m_1} workers do not engage in learning. This also implies that at all levels of human capital greater than h_{m_1} workers do not engage in learning, since the incentives to learn decrease as we climb the human capital ladder.

The mass of workers at h_{m_1} is given by $N_l f_l h_{m_1}$. Notice that since no learning occurs at h_{m_1}, the total mass of individuals with human capital of h_{m_1+1} or above will be given by $N_l f_l(h_{m_1})\frac{\epsilon(1-\delta)}{\delta}$. In Assumption 2, we assume $\epsilon < \frac{\delta}{1-\delta}$, and thus $N_l f_l(h_{m_1})\frac{\epsilon(1-\delta)}{\delta} < N_l f_l(h_{m_1})$. Since in Assumption 1 we assume the production function y is concave in each input, this implies that all wages for human capital levels greater than h_{m_1} are higher than that at h_{m_1}. This further implies that the expected lifetime value of h_{m_1+1}, $EV(h_{m_1+1})$, is larger than the lifetime value of h_{m_1}, $EV(h_{m_1})$.

Notice that from the worker's Bellman equation described in Appendix D.2, if we subtract the value of working from the value of internal learning at h_{m_1} we get:

$$\beta(1-\delta)\left[p_c(h_{m_1})-\epsilon\right]\left[EV(h_{m_1+1})-EV(h_{m_1})\right]$$

Recall also that the probability of internal learning is given by $p_c(h_m) = (1 - F_l(h_m)) + F_l(h_m)\epsilon$, and as such, at h_{m_1}:

$p_c(h_m) = f_l(h_{m_1})\frac{\epsilon(1-\delta)}{\delta} + (1 - f_l(h_{m_1})\frac{\epsilon(1-\delta)}{\delta})\epsilon = \frac{\epsilon(1-\delta)}{\delta} f_l(h_{m_1})(1-\epsilon) + \epsilon > \epsilon.$

As such, since $p_c(h_m) > \epsilon$ and $EV(h_{m_1+1}) > EV(h_{m_1})$ the value of working is lower than the value of internal learning . This violates the assumption that at h_{m_1} no workers learn. Therefore, at all levels of human capital h_m workers engage in learning.

D.6.2 Proof of Lemma 2

Within each level of human capital, all learners choose either one of the two modes of learning. This is because both internal and external learning follow a linear production of human capital, meaning neither the increase in human capital, the probability of learning, nor the cost of each source of learning directly depend on the number of learners within each capital step. As such, although under some parametrizations the value of internal learning may be equal to the value of external learning for some human capital level, this won't be the case in general.

D.6.3 Proof of Lemma 3

We will proceed by contradiction. Suppose that there is a h_m such that $w(h_m) \leq w(h_{m-1})$. Then, either of two things must be true: (1) individuals do not learn at h_{m-1} since the return of moving up the ladder is at best equal (and climbing the ladder would imply giving up on wages for one period); or (2) individuals learn to reach h_m only insofar it is an intermediate step to reach a higher level of human capital which yields a higher return. The first case is a direct violation of Lemma 1. In the second case, since the incentives of placing at h_m are just to reach a higher human capital step, all workers with human capital h_m choose to learn, yielding no production workers at that level, thus also violating Lemma 1.

D.6.4 Proof of Lemma 4

We will proceed by contradiction. Assume we have that $f_t(h_m) > 0$, and $f_s(h_{m-1}) = 0$. Since $f_t(h_m) > 0$, we must have that the expected trainer wage at h_m is equal to the production wage: $w(h_m) = qp_n(h_m) = qF_s(h_{m-1})$. Now, since we don't have any external learners at $m-1$, meaning $f_s(h_{m-1}) = 0$, we have that $F_s(h_{m-1}) = F_s(h_{m-2})$. This implies, in turn, that the trainer wage at $m-1$ is equal to that at m: $qF_s(h_{m-2}) = qF_s(h_{m-1}) = w(h_m)$. However, this would then imply that there are no production workers at $m-1$, since it is more profitable to be a trainer than a production worker. This however is not possible given Lemma 1.

D.6.5 Proof of Lemma 5

The first part follows because for every human capital level h_m where the mass of external learners is positive, we must have a strictly higher human capital level with a positive mass of trainers in order for the the probability of external learning at h_m to be larger than zero, and thus for external learners at h_m to have incentives to learn through external learning. The second part follows because at the maximum human capital level for which trainers exist $\bar{m}$, there cannot be a positive mass of external learners because the probability of external learning is zero at and after this point.

D.6.6 Proof of Lemma 6

We will proceed by contradiction. Assume that at the lowest level of human capital, h_1, workers engage in external learning. Then, from the worker's Bellman equation described in Appendix D.2, we must have that the difference between the value of external learning and internal learning is positive and thus:

$$\beta(1-\delta)\left[p_s(h_1)-p_c(h_1)\right]\left[EV(h_2)-EV(h_1)\right] > qp_s(h_1)$$

Further, we must have that the value of external learning equals the value of working at h_1, and thus:

$$\beta(1-\delta)\left[p_s(h_1)-\epsilon)\right]\left[EV(h_2)-EV(h_1)\right] = qp_s(h_1)+w(h_1)$$

We can then combine these two equations, noting that $p_s(h_1)=1$ since no trainers place at h_1 and rearrange to get:

$$q(p_c(h_1)-\epsilon) < w(h_1)(1-p_c(h_1))$$

Assumption 5 implies that trainers place at h_M. Since trainers of that human capital can train all workers, at h_M the trainer wage must equal the production wage: $q = w(h_M)$. If we further ignore the terms with ϵ, since ϵ is very small, the equation before becomes:

$$w(h_M)(p_c(h_1)) < w(h_1)(1-p_c(h_1))$$

Notice that from Lemma 3, we know $w(h_M) > w(h_1)$. Further, we know that at h_1, the contingent of workers with human capital higher than h_1 is large, and thus we must have $p_c(h_1) = 1-F_l(h_1) > 1-p_c(h_1) = F_l(h_1)$. This is confirmed in Corollary 1. Thus, we must have $w(h_M)(p_c(h_1)) > w(h_1)(1-p_c(h_1))$, which contradicts the equation above.

D.6.7 Proof of Lemma 7

The first part of Lemma 7 follows from the fact that by definition, h_{m*} is the first level of human capital where individuals choose external learning. The second part follows directly from Lemma 6. In particular, given that the lowest possible human capital level external learners will place at is h_2, and that trainers must be more knowledgeable than external learners in order to effectively train them and get paid, the lowest human capital level trainers will consider placing at is h_3. The third follows from the fact that the probability of internal learning dips as we climb the human capital ladder. In order to show this, we will proceed by contradiction. In particular, assume that $\exists h_m \in [h_m*, h_{\underline{m}-1}] \mid f_s(h_m) = 0$. Without loss of generality, assume $h_m = h_{m*+1}$. Then, Lemma 2 implies that at h_{m*+1}, individuals learn internally, while at h_m* individuals learn externally.

First, we know that since there is a positive mass of workers at every human capital level, we can write:

$$V(h_m) = V_l(h_m) = w(h_m) + \beta(1-\delta)\left[(1-\epsilon)V(h_m) + \epsilon V(h_{m+1})\right] \forall h_m$$

We can ignore ϵ since its very small, as before to get: $V(h_m) = \frac{w(h_m)}{1-\beta(1-\delta)} \forall h_m$.

In addition, we have that the difference between the value of external learning and the value of internal learning is negative at $m*$, while the opposite is true at $m*+1$. Then, we can combine the worker's Bellman equations described in Appendix D.2 for h_{m*} and h_{m*+1} with the equation above for $V(h_m)$ and the fact that since the first mass of trainers occurs at $h_{\underline{m}}$, $p_s(h_{m*}) = p_s(h_{m*+1}) = 1$ to get:

$$(1 - p_c(h_{m*}))\left[w(h_{m*+1}) - w(h_{m*})\right] > (1 - p_c(h_{m*+1}))\left[w(h_{m*+2}) - w(h_{m*+1})\right] \qquad (1)$$

In addition, we know the values of working and external learning must be equal at h_{m*}, while the values of working and internal learning must be equal at h_{m*+1}. We can then combine the worker's value functions described in Appendix D.2 for h_{m*} and h_{m*+1} with the equation above for $V(h_m)$ and $p_s(h_{m*}) = p_s(h_{m*+1}) = 1$ to get:

$$w(h_{m*}) = -q + \frac{\beta(1-\delta)}{1-\beta(1-\delta)}\left[w(h_{m*+1}) - w(h_{m*})\right]$$
$$w(h_{m*+1}) = \frac{\beta(1-\delta)p_c(h_{m*+1})}{1-\beta(1-\delta)}\left[w(h_{m*+2}) - w(h_{m*+1})\right]$$

Since we know from Lemma 3 that $w(h_{m*+1}) > w(h_{m*})$, we can combine these two equations and combine further with the Bellman equation differences above to get:

$$p_c(h_{m*+1})\left[w(h_{m*+2}) - w(h_{m*+1})\right] > p_c(h_{m*})\left[w(h_{m*+1}) - w(h_{m*})\right] \qquad (2)$$

Combining Equation (2) with Equation (1) we get:

$$p_c(h_{m*+1})\left[w(h_{m*+2})-w(h_{m*+1})\right] > p_c(h_{m*})\frac{(1-p_c(h_{m*+1})}{1-p_c(h_{m*})}\left[w(h_{m*+2})-w(h_{m*+1}\right]$$

Canceling and rearranging this implies:

$$p_c(h_{m*+1}) > p_c(h_{m*})$$

This is however not true, since $p_c(h_{m*+1}) = 1 - F(h_{m*+1}) < 1 - F_l(h_{m*}) = p_c(h_{m*})$

As such, we must have that the mass of external learners is positive from h_{m*} to $h_{\underline{m}-1}$: $f_s(h_m) > 0 \; \forall h_m \in [h_m, h_{\underline{m}-1}]$.

D.6.8 Proof of Proposition 1

The first part follows directly from Lemma 6, Assumption 4, Lemma 7, Assumption 5 and Lemma 4. In particular, Lemma 6 implies that at the lowest level of human capital h_1, workers learn internally. This continues until h_{m*}, which denotes the lowest human capital level where individuals learn externally, which must exist given Assumption 4. Then, Lemma 7 guarantees that from h_{m*} to $h_{\underline{m}-1}$ (human capital level below where trainers start locating), workers continue to engage in external learning. Then, since from Assumption 5 trainers locate in all human capital levels from $h_{\underline{m}}$ to h_M, we know by Lemma 4 that external learners must locate in all human capital levels from $h_{\underline{m}-1}$ to h_{M-1}.

The second part follows from Assumption 5 and Lemma 5. In particular, Assumption 5 indicates that trainers locate at all levels of human capital beyond $h_{\underline{m}}$, while since the lowest human capital level where external learners locate is h_{m*}, Lemma 5 implies that $h_{\underline{m}} > h_{m*}$.

D.6.9 Proof of Corollary 1

We will first notice that if we let N_m represent the total mass of individuals of type m, and N_l the total mass of workers, we have that the portion of workers within each human capital level is given by: $k_l(h_m) = \frac{f_l(h_m)N_l}{N_m}$.

We want to show that $\frac{k_l(h_{m+1})}{k_l(h_m)} > 1 \; \forall h_m \in [h_1, h_{m*-2}]$. This implies showing:

$$\frac{\frac{f_l(h_{m+1})N_l}{N_{m+1}}}{\frac{f_l(h_m)N_l}{N_m}} > 1 \ \forall h_m \in [h_1, h_{m*-2}] \iff \frac{f_l(h_{m+1})}{f_l(h_m)} > \frac{N_{m+1}}{N_m} \ \forall h_m \in [h_1, h_{m*-2}]$$

In order to show this, first notice that from Proposition 1, we know that individuals with human capital levels from h_1 to h_{m*-1} only engage in production work, or internal learning. Therefore, if we let N_c represent the total mass of internal learners, we have:

$$N_m = f_l(h_m)N_l + f_c(h_m)N_c$$

From the equation before, we want to show:

$$\frac{f_l(h_{m+1})}{f_l(h_m)} > \frac{f_l(h_{m+1})N_l + f_c(h_{m+1})N_c}{f_l(h_m)N_l + f_c(h_m)N_c} \iff \frac{f_l(h_{m+1})}{f_l(h_m)} > \frac{f_c(h_{m+1})}{f_c(h_m)}$$

In other words, we want to show that the relative mass of production workers rises faster than the relative mass of internal learners. Notice that within each human capital level, the mass of individuals who engage in work or internal learning is pinned down by the net benefit of internal learning relative to work. If this net benefit increases, then the mass of individuals who learn increases and the mass of individuals who work decreases. The opposite is true if this net benefit decreases. As such, since workers with human capital levels from h_1 to h_{m*-1} only engage in production work, or internal learning, in order to show $\frac{f_l(h_{m+1})}{f_l(h_m)} > \frac{f_c(h_{m+1})}{f_c(h_m}$ it is enough to show that $\frac{f_l(h_{m+1})}{f_l(h_m)} > 1$.

To do this we will use wage expressions and the Bellman equation. In particular, from the worker's Bellman equation described in Appendix D.2, if we subtract the value of working from the value of internal learning and ignore ϵ as before we get:

$$w(h_m) = \beta(1-\delta)p_c(h_m)\left[EV(h_{m+1}) - EV(h_m)\right] \ \forall h_m \in [h_1, h_{m*-1}]$$

The left hand side of this equation represents the cost of internal learning relative to work, while the right hand side represents its relative benefit. We can now substitute $EV(h_m) = \frac{w(h_m)}{1-\beta(1-\delta)}$ as in Lemma 7 and rearrange to get:

$$\frac{w(h_{m+1})}{w(h_m)} = \frac{1-\beta(1-\delta)(1-p_c(h_m))}{\beta(1-\delta)p_c(h_m)} \tag{1}$$

Now, we can use the expression for wages from the production function denoting the link between wages and f_l to note[57]

$$\frac{w(h_{m+1})}{w(h_m)} = \frac{h_{m+1}}{h_m}\frac{G\left(h_{m+1}f_l(h_{m+1})\right)}{G\left(h_m f_l(h_m)\right)} \tag{2}$$

Where the function G denotes the first derivative of y. We denote $\frac{G(x)}{G(z)} = Q\left(\frac{x}{y}\right)$. Taking the inverse function of Q, rearranging and plugging in Equation (1) we get:

$$\frac{f_l(h_{m+1})}{f_l(h_m)} = \frac{h_m}{h_{m+1}}Q^{-1}\left(\frac{h_m(1-\beta(1-\delta)(1-p_c(h_m)))}{h_{m+1}\beta(1-\delta)p_c(h_m)}\right)$$

We thus want to show:

$$Q^{-1}\left(\frac{h_m(1-\beta(1-\delta)(1-p_c(h_m)))}{h_{m+1}\beta(1-\delta)p_c(h_m)}\right) > \frac{h_{m+1}}{h_m}$$

Notice that the right hand side will be the lowest possible whenever the term inside it is the largest, since the production function y is concave, and thus Q and Q^{-1} are decreasing functions. This occurs when $p_c(h_m)$ is as low as possible, and thus entails the maximum number possible of individuals engaging in work up to h_m (since that will imply a large accumulation mass of workers that h_m individuals cannot learn from). Notice that since the cost of learning increases with h_m this would then also imply individuals engaging in work after h_m. As such, in this extreme situation we would have no learners, and all climbing of the human capital stems from ϵ. It is straightforward to show in this case that:

$$p_c(h_m) = \epsilon(1-\delta)N_m = \frac{\delta((1-\delta)\epsilon)^m}{(1-(1-\delta)(1-\epsilon))^m}$$

Notice that this will the lowest whenever m is the lowest, which given we've assumed an external learning equilibrium, occurs at $m* = M-2$.

[57]In particular, we have that the wage is given by:

$$w(h_m) = \frac{dY}{dN_{l,m}} = \frac{dy(h_1 f_l(h_1), h_2 f_l(h_2), ..., h_M f_l(h_M))}{d\left[h_m f_l(h_m)\right]}h_m$$

The second equality follows from noting that the total amount of labor of type m is: $N_{l,m} = f_l(h_m)N_l$, and y is homogeneous of degree one.

Thus, taking logs, we want to show that the step size in the human capital ladder, $\frac{h_{m+1}}{h_m}$ is large enough such that:

$$\log\left[Q^{-1}\left(\frac{h_m(1-\beta(1-\delta)(1-\frac{\delta((1-\delta)\epsilon)^{M-2}}{(1-(1-\delta)(1-\epsilon))^{M-2}}))}{h_{m+1}\beta(1-\delta)\frac{\delta((1-\delta)\epsilon)^{M-2}}{(1-(1-\delta)(1-\epsilon))^{M-2}}}\right)\right] > \log(h_{m+1}) - \log(h_m)$$

This is assumed in Assumption 3. Therefore we have that $\frac{f_l(h_{m+1})}{f_l(h_m)} > 1 > \frac{f_c(h_{m+1})}{f_c(h_m)}\ \forall h_m \in [h_1, h_{m*-2}]$, and thus that $\frac{k_l(h_{m+1})}{k_l(h_m)} > 1\ \forall h_m \in [h_1, h_{m*-2}]$.

D.6.10 Proof of Proposition 2

The Blanchard-Yaari structure implies that the only force driving the work and learning decisions of individuals is the human capital level. This therefore implies that the distribution of learning and working decisions across workers of each age follow directly and solely from their corresponding distribution across the human capital state-space. Thus, we will first characterize the distribution across the human capital state-space of workers of each age.

Working recursively, and ignoring the portion of workers ϵ who learn while engaging in production work since ϵ is small, we can show that the portion of individuals of each human capital level h_m at each age j, $\pi_j(h_m)$, can be described by:

$$\pi_j(h_m) = \begin{cases} [k_l(h_1) + g_c(h_1)(1-p_c(h_1))]^{m-1} \text{ if } m = 1 \\ \left[\prod_{k=0}^{m-1} p_c(h_k)g_c(h_k)\right]\left[\sum_{j-m\geq k_{m-1}\geq k_{m-2}\geq \ldots \geq k_1 \geq 0}\left(x_m^{j-m-k_{m-1}} x_1^{k_1} \prod_{n=2}^{m-1} x_n^{k_n - k_{n-1}}\right)\right] \text{ if } 1 < m < j \\ \prod_{k=0}^{m-1} p_c(h_k)g_c(h_k) \text{ if } m = j \\ 0 \text{ if } m > j \end{cases}$$

Where: $x(h_m) = k_l(h_m) + g_t(h_m) + g_c(h_m)(1 - P_{learn*(h_m)}(h_m))$, and:

$$P_{learn*(h_m)} = \begin{cases} p_c(h_m) \text{ if } h_m < h_{m*} \\ p_s(h_m) \text{ if } h_m \geq h_{m*} \end{cases}$$

From this, we can see that given that there is no depreciation of human capital in this economy, the distribution of human capital shifts right as workers age. Formally, the sum of portions of workers in each human capital bucket right of h_1 increases with age:

$$\sum_{m=2}^{M} \pi_j(h_m) > \sum_{m=2}^{M} \pi_{j-1}(h_m)$$

The first part of Proposition 2, namely the decline in the portion of internal learners with age follows from Proposition 1 and Corollary 1. Between ages 1 and $m*-1$, workers have human capital levels between h_1 and h_{m*-1}, which progressively feature a smaller proportion of learners as shown in Corollary 1. Once workers reach age $m*$, there will be a positive mass of workers with human capital h_{m*}, which will shift their mode of learning to external learning, accelerating the decline in the portion of internal learners. These two forces drive the portion of production workers engaging in internal learning to decline with age.

The second part of Proposition 2, namely the inverted U-shape of external learning with age follows from Proposition 1. As argued above, and due to the existence of the external learning threshold h_{m*}, the portion of individuals engaging in external learning will be zero for individuals younger than $m*$. For individuals older than age $m*$, however, this portion will be positive. This yields the first part of the inverted U-shape. As individuals continue to age, this portion may continue rising, if the portion of individuals with human capital level above the external learning threshold h_{m*}, increases, or may also begin to decline if the incentives to work become larger. At age $j \geq M$, however, this portion will begin to decline unequivocally since a positive mass of workers will reach human capital level M, where there is no learning. This yields the second part of the inverted U-shape, where the portion of individuals engaging in external learning declines.

E Appendix: Additional Information for Testable Predictions

E.1 Additional Information for Section 4.5.1

E.1.1 Definition of Trainer and Production Worker

- "Trainer": is a binary variable that takes a value of one for workers who report an occupation that involves training, teaching or instruction activities outside of school and university education.

- German BIBB Data: For the German data, we define trainers as those having occupations of "other teaching professionals" or "other teaching associate professionals", meaning workers who engage in teaching activities other than those connected with primary, pre-primary and special education school levels. The specific 3-digit ISCO 1988 codes we use to define trainers are 2359 and 0334.

- US NHES Data: For the US Data we define trainers as those having occupations of "Training and development managers", "Training and development specialists" or "Other education, training, and library workers", meaning training professionals or specialists, and teachers outside of postsecondary, preschool and kindergarten, elementary and middle school, secondary and special education. The specific ACS 2000 codes we use to define trainers are 0137, 0650 and 2550.

• "Production Worker" is defined in two ways. In the first way, "Production Worker" is a binary variable that takes a value of one for workers who report any occupation outside of being a trainer. In the second way, "Production Worker" is a binary variable that takes a value of one for workers who report any professional or technical occupation outside of being a trainer.[58]

[58]In the German data, professional and technical occupations encompass 3-digit ISCO 1988 codes in the 100s, 200s and 300s. In the US data, professional and technical occupations encompass ACS 2000 codes below 3700.

E.1.2 Quantile Regression of Potential Years of Experience for Trainers v. Production Workers

Table E.1: Quantile Regression of Potential Years of Experience for Trainers v. Production Workers

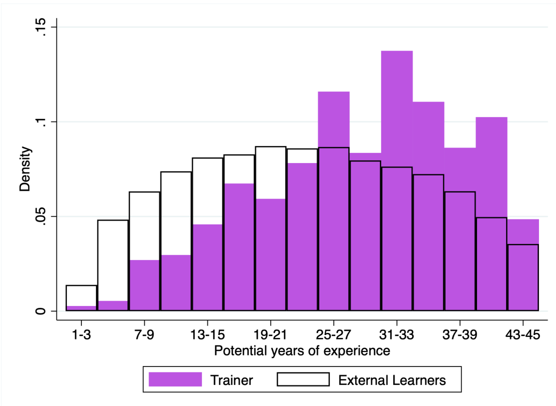
Dep. Variables	Potential Years of Experience					
	25th percentile		50th percentile		75th percentile	
Germany						
Trainer	4***	4**	3	3***	3**	2
	(1.254)	(1.708)	(2.005)	(1.055)	(1.254)	(1.269)
Constant	14***	3***	24***	14***	33***	24***
	(0.0529)	(0.287)	(0.0423)	(0.487)	(0.0529)	(0.336)
Observations	173,639	165,265	173,639	165,265	173,639	165,265
Year FE		Y		Y		Y
Demographic Controls		Y		Y		Y
Worker type FE		Y		Y		Y
Firm size FE		Y		Y		Y
USA						
Trainer	4	4	-1	4**	1	3
	(7.557)	(6.560)	(6.044)	(1.593)	(3.782)	(6.761)
Constant	10***	11***	20***	24***	31***	35***
	(0.186)	(0.375)	(0.149)	(0.496)	(0.186)	(0.410)
Observations	29,217	29,217	29,217	29,217	29,217	29,217
Demographic Controls		Y		Y		Y
Worker type FE		Y		Y		Y

Trainer (v. production worker) described in text for both countries. All regressions weighted using observation weights provided in the surveys. All regressions weighted using observation weights provided in the surveys. *Germany:* Year fixed effects correspond to year of survey fixed effects. Demographic controls include educational attainment level and gender. Worker type categories include laborer, private employee, government employee, self-employed, freelancer, or family caregiver. Firm size is a categorical variable indicating whether the firm where the worker works at has less than 4 workers, 5–9 workers, 10–49 workers, 50–99 workers, 100–499 workers, 500–999 workers and 1000 or more workers. *USA:* Demographic controls include educational attainment level, race, census region, and gender. Worker type categories include private employee, government employee, self-employed, or working without pay. We do not include wage controls, nor occupation or industry fixed effects in these regressions since trainers and production workers have inherently different wage levels, occupations and industries. We do not include age fixed effects due to collinearity between potential years of experience, education, and age. Robust standard errors in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

E.1.3 Additional Plots and Tables

Figure E.1: Histograms of trainers and external learners by potential experience

(a) Germany



(b) United States

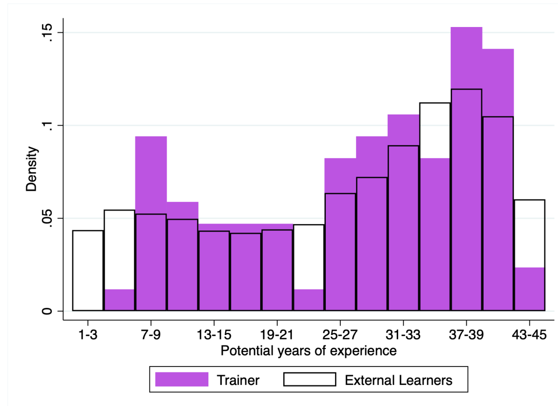
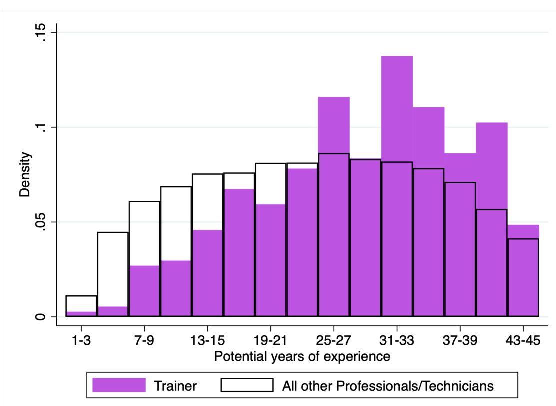


Figure E.2: Histograms of trainers and professional and technical production workers by potential experience

(a) Germany



(b) United States

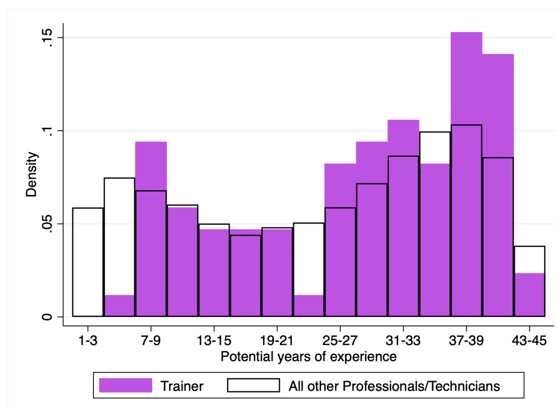


Table E.2: Quantile Regression of Potential Years of Experience for Trainers v. External Learners

Dep. Variables	Potential Years of Experience					
	25th percentile		50th percentile		75th percentile	
Germany						
Trainer	4***	3.500**	4**	3***	4***	2**
	(1.116)	(1.400)	(1.784)	(1.131)	(1.116)	(0.855)
Constant	14***	3***	23***	11.50***	32***	22.40***
	(0.0564)	(0.321)	(0.0902)	(0.363)	(0.0564)	(0.352)
Observations	121,962	116,175	121,962	116,175	121,962	116,175
Year FE		Y		Y		Y
Demographic Controls		Y		Y		Y
Worker type FE		Y		Y		Y
Firm size FE		Y		Y		Y
USA						
Trainer	2	0	-3	0	0	0
	(7.324)	(0)	(4.691)	(6.152)	(2.936)	(0)
Constant	12***	1	22***	1	32***	1
	(0.214)	(0)	(0.257)	(4.382)	(0.214)	(0)
Observations	13,613	13,613	13,613	13,613	13,613	13,613
Demographic Controls		Y		Y		Y
Worker type FE		Y		Y		Y

Trainer (v. external learner) described in text for both countries. All regressions weighted using observation weights provided in the surveys. *Germany:* Year fixed effects correspond to year of survey fixed effects. Demographic controls include educational attainment level and gender. Worker type categories include laborer, private employee, government employee, self-employed, freelancer, or family caregiver. Firm size is a categorical variable indicating whether the firm where the worker works at has less than 4 workers, 5–9 workers, 10–49 workers, 50–99 workers, 100–499 workers, 500–999 workers and 1000 or more workers. *USA:* Demographic controls include educational attainment level, race, census region, and gender. Worker type categories include private employee, government employee, self-employed, or working without pay. We do not include wage controls, nor occupation or industry fixed effects in these regressions since trainers and production workers have inherently different wage levels, occupations and industries. We do not include age fixed effects due to collinearity between potential years of experience, education, and age. Robust standard errors in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table E.3: Quantile Regression of Potential Years of Experience for Trainers v. Professional & Technical Production Workers

Dep. Variables	Potential Years of Experience					
	25th percentile		50th percentile		75th percentile	
Germany						
Trainer	5***	3***	4***	2**	3***	1.600
	(1.618)	(1.069)	(1.297)	(0.857)	(0.811)	(1.255)
Constant	13***	-3***	23***	8***	33***	20.40***
	(0.0672)	(0.548)	(0.107)	(0.606)	(0.0672)	(0.532)
Observations	46,563	45,109	46,563	45,109	46,563	45,109
Year FE		Y		Y		Y
Demographic Controls		Y		Y		Y
Worker type FE		Y		Y		Y
Firm size FE		Y		Y		Y
USA						
Trainer	4	4	0	3*	2	4
	(7.455)	(7.085)	(4.774)	(1.706)	(2.997)	(7.292)
Constant	10***	13***	19***	25***	30***	34***
	(0.209)	(0.645)	(0.250)	(0.689)	(0.313)	(0.873)
Observations	14,357	14,357	14,357	14,357	14,357	14,357
Demographic Controls		Y		Y		Y
Worker type FE		Y		Y		Y

Trainer (v. technical production worker) described in text for both countries. All regressions weighted using observation weights provided in the surveys. *Germany:* Year fixed effects correspond to year of survey fixed effects. Demographic controls include educational attainment level and gender. Worker type categories include laborer, private employee, government employee, self-employed, freelancer, or family caregiver. Firm size is a categorical variable indicating whether the firm where the worker works at has less than 4 workers, 5–9 workers, 10–49 workers, 50–99 workers, 100–499 workers, 500–999 workers and 1000 or more workers. *USA:* Demographic controls include educational attainment level, race, census region, and gender. Worker type categories include private employee, government employee, self-employed, or working without pay. We do not include wage controls, nor occupation or industry fixed effects in these regressions since trainers and production workers have inherently different wage levels, occupations and industries. We do not include age fixed effects due to collinearity between potential years of experience, education, and age. Robust standard errors in parentheses. Robust standard errors in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

E.2 Additional Information for Section 4.5.2

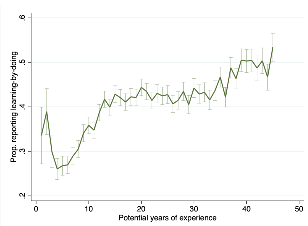
E.2.1 Definition of "Learning-by-Doing"

- "Learning-by-doing" is a binary variable that indicates whether the interviewee acquired professional skills by doing the job itself.
  - 1985/1986, 1991/1992, 1998/1999: All of the listed surveys contain questions that determine whether or not the interviewee claims to have acquired professional knowledge/skills by doing his or her job. The 1979 survey does not distinguish between learning-by-doing and internal learning, and thus is not used. [59]

E.2.2 Additional Plots and Tables[60]

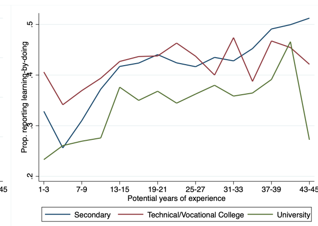
Figure E.3: Prevalence of Learning-by-Doing throughout workers' lifecycles in Germany

(a) By one-year exp. bins



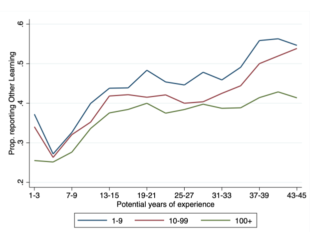
(b) By gender

(c) By educational level

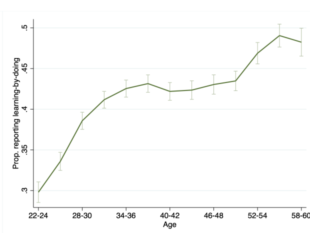


(d) By wave

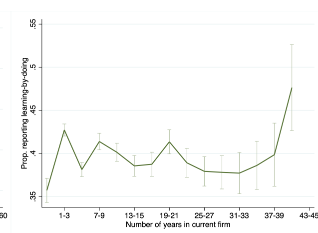
(e) By firm size



(f) By age



(g) By tenure in current firm



[59] The three most recent survey waves in 2005/2006, 2011/2012, and 2017/2018 do not contain this information.

[60] 95% confidence intervals included for some plots, but omitted in plots that consider several groups for clarity.

Table E.4: Correlations between Learning-by-Doing and potential experience

Dep. Variables	Learning-by-Doing		
Germany			
Potential Yrs. Experience	0.0091***	0.0084***	0.0091***
	(0.0007)	(0.0009)	(0.0014)
Potential Yrs. Experience^2	-0.0001***	-0.0001***	-0.0001***
	(0.0000)	(0.0000)	(0.0000)
Constant	0.2670***	0.1495***	0.1779***
	(0.0077)	(0.0216)	(0.0303)
Observations	90,536	51,455	21,378
R-squared	0.0115	0.1232	0.0812
Year FE		Y	Y
Demographic Controls		Y	Y
Worker type FE		Y	Y
Industry FE		Y	Y
Occupation FE		Y	Y
Firm size FE		Y	Y
Wage Controls			Y

Learning-by-Doing and potential years of experience described in text. All regressions weighted using observation weights provided in the surveys. *Germany:* Year fixed effects correspond to year of survey fixed effects. Demographic controls include educational attainment level and gender. Worker type categories include laborer, private employee, government employee, self-employed, freelancer, or family caregiver. Industry categories at the 1-digit level. Occupation categories at the 2-digit level (ISCO 88). Firm size is a categorical variable indicating whether the firm where the worker works at has less than 4 workers, 5–9 workers, 10–49 workers, 50–99 workers, 100–499 workers, 500–999 workers and 1000 or more workers. Wage controls include the current hourly wage for the worker. Robust standard errors in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

E.3 Additional Information for Section 4.5.3

E.3.1 Skill-Content and Tool Difference between Internal and External Learners

We provide an additional test of the human capital differences between internal and external learners by exploring differences on the skill-content of tasks performed and tools used for these tasks by each of these two groups of workers.[61]

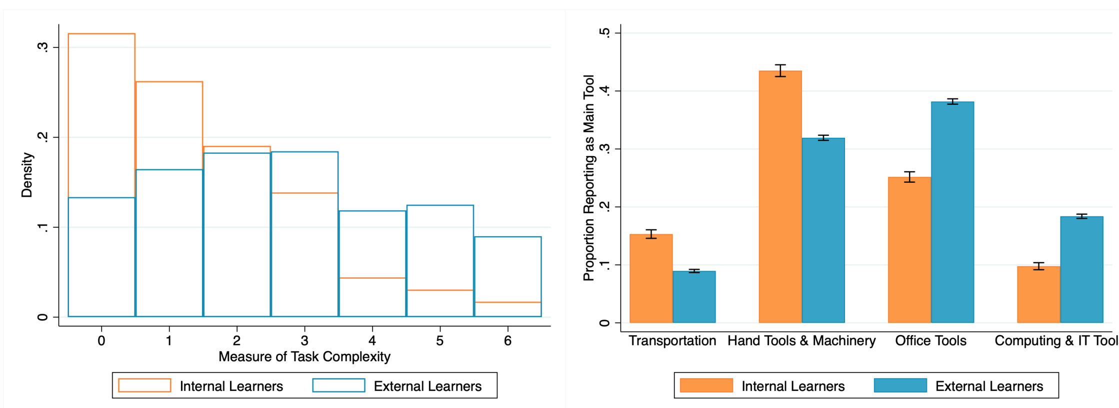
[61] We do not consider trainers here, since their tasks are qualitatively very different from those of production workers.

We rely on two pieces of information from our German data, namely the skills and tools workers report using in their jobs. First, we construct a measure of task complexity by counting how many of the following skills workers' use on their jobs: Math and Stats; Foreign Language; Computing; Accounting, Purchasing, Financing and Taxes; Marketing, and Management and Organization.[62] Larger values of this measure imply a higher number of skills used in the job, and thus higher task complexity. In Panel (a) of Figure E.4 we show that the distribution of individuals concentrates more heavily among lower levels of task complexity for individuals learning internally than those learning externally. We then formally test these distribution differences through quantile regressions of the median of task complexity on the external learning variable (where the omitted category is internal learning) in Table E.6.[63] The results from these regressions indicate that the median of task complexity for external learners is larger than that of internal learners.

Figure E.4: Histogram of Task Complexity and Main Tool Use for Internal and External Learners

(a) Histogram of Task Complexity

(b) Main Tool Use



Then, we build binary variables capturing whether the main tool employed by the worker in her job corresponds to transportation equipment (such as trucks or forklifts), hand tools and machinery (such as hammers, drills or hair dryers), office equipment (such as writing materials, phones, or calculators) or computers and other IT equipment.[64] This tool information provides insights into the attributes of the worker's job, and particularly the skill-level required, as suggested by DiNardo and Pischke (1997). Specifically, the tool categories above

[62]Please see Appendix E.3.2 for details on the construction of these skill variables.
[63]We do not a quantile regression for other quantiles here, since the measure of task complexity contains only 7 values, and does not have enough variation across groups at the lower and upper ends of the distribution.
[64]Please see Appendix E.3.3 for details on the construction of these tool variables.

separate blue-collar occupations (main tools used are transportation and hand tools) from white-collar occupations (main tools used are office equipment and computers). In Panel (b) of Figure E.4 we plot the proportion of external and internal learners who report their main tool to be in each of the four categories above, along with 95% confidence intervals. The plot suggests that external learners are more likely to use "white-collar" tools than internal learners, while the opposite is true for "blue-collar" tools. We formally test the difference in "white-" versus "blue-collar" tool use in Table E.7. The results from these regressions indicate that external learners are more likely to employ "white-collar" tools than internal learners even after controls are added.

E.3.2 Construction of Job-Related Skill Variables

We construct a measure of task complexity by counting how many complex skills workers' report using on their jobs. The waves used to construct this measure encompass 1992, 1999, 2006, 2012 and 2018. Earlier waves do not contain this information. There are six categories of skills, summarized in the following variables:

- Math and Statistics is a binary variable that takes a value of one for workers who report needing math and statistics knowledge for their current job.
- Foreign Language is a binary variable that takes a value of one for workers who report needing to use a language other than German for their current job.
- Computing is a binary variable that takes a value of one for workers who report needing computing knowledge for their current job.
- Accounting, Purchasing, Financing and Taxes is a binary variable that takes a value of one for workers who report needing accounting, purchasing, financing, tax or related knowledge for their current job.
- Marketing is a binary variable that takes a value of one for workers who report needing marketing or related knowledge for their current job.
- Management and Organization is a binary variable that takes a value of one for workers who report needing management and organization knowledge for their current job.

E.3.3 Construction of Job-Related Tool Use

We construct binary variables capturing whether the main tool employed by the worker in her job corresponds to different categories. The waves used to construct these variables are 1979, 1986, 1992 and 1999. Latter waves appear to collect this information, but it is not available in the data files.

We consider four specific tool categories, summarized in the following variables:

- Transportation Equipment is a binary variable that takes a value of one for workers who report that the main tool used in their current jobs corresponds to transportation equipment such as motor vehicles, tractors, snowplows, bulldozers, forklifts, cranes, hoists, rail vehicles, handcarts, etc.
- Hand Tools is a binary variable that takes a value of one for workers who report that the main tool used in their current jobs corresponds to hand tools or machinery such as hammers, screwdrivers, gauges, welding machines, drills, hair dryers, ovens, sewing machines, elevators, etc.
- Office equipment is a binary variable that takes a value of one for workers who report that the main tool used in their current jobs corresponds to office equipment such as pencils, rulers, stamps, phones, calculators, files, books, copiers, cash registers, etc.
- Computer and Other IT Equipment is a binary variable that takes a value of one for workers who report that the main tool used in their current jobs corresponds to a computer or other IT equipment such as network devices, digital graphics systems, terminals, etc.

E.3.4 Regressions

Table E.5: Quantile Regression of Potential Years of Experience for Trainers and External Learners v. Internal Learners

Dep. Variables	Potential Years of Experience					
	25th percentile		50th percentile		75th percentile	
Germany						
External Learner	2***	1***	1***	-0	-1***	-2***
	(0.162)	(0.135)	(0.246)	(0.197)	(0.162)	(0.164)
Trainer	6***	5***	5***	2.667**	3**	0.333
	(1.172)	(1.377)	(1.874)	(1.245)	(1.172)	(0.858)
Constant	12***	2***	22***	13***	33***	25***
	(0.151)	(0.320)	(0.242)	(0.428)	(0.151)	(0.351)
Observations	138,310	131,936	138,310	131,936	138,310	131,936
Year FE		Y		Y		Y
Demographic Controls		Y		Y		Y
Worker type FE		Y		Y		Y
Firm size FE		Y		Y		Y
USA						
External Learner	6***	5***	10***	7***	11***	8***
	(0.325)	(0.359)	(0.460)	(0.498)	(0.515)	(0.625)
Trainer	8	7	7	9***	11***	11
	(7.825)	(4.742)	(5.018)	(0.698)	(3.163)	(7.339)
Constant	6***	9***	12***	19***	21***	27***
	(0.231)	(0.671)	(0.369)	(0.795)	(0.461)	(0.865)
Observations	16,600	16,600	16,600	16,600	16,600	16,600
Demographic Controls		Y		Y		Y
Worker type FE		Y		Y		Y

Trainer and external learner (v. internal learner) described in text for both countries. All regressions weighted using observation weights provided in the surveys. *Germany:* Year fixed effects correspond to year of survey fixed effects. Demographic controls include educational attainment level and gender. Worker type categories include laborer, private employee, government employee, self-employed, freelancer, or family caregiver. Firm size is a categorical variable indicating whether the firm where the worker works at has less than 4 workers, 5–9 workers, 10–49 workers, 50–99 workers, 100–499 workers, 500–999 workers and 1000 or more workers. *USA:* Demographic controls include educational attainment level, race, census region, and gender. Worker type categories include private employee, government employee, self-employed, or working without pay. We do not include wage controls, nor occupation, age or industry fixed effects in these regressions since trainers and production workers have inherently different wage levels, occupations and industries. We do not include age fixed effects due to collinearity between potential years of experience, education, and age. Robust standard errors in parentheses. Robust standard errors in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table E.6: Quantile Regression of Task Complexity for External Learners v. Internal Learners

Dep. Variables	Task Complexity 50th percentile	
Germany		
External Learner	2***	0.584***
	(0.0991)	(0.222)
Constant	1***	3.827***
	(0.0940)	(0.840)
Observations	84,315	29,322
Year FE		Y
Demographic Controls		Y
Age FE		Y
Worker type FE		Y
Industry FE		Y
Occupation FE		Y
Firm size FE		Y
Wage Controls		Y

External learner (v. internal learner) and task complexity described in text. All regressions weighted using observation weights provided in the surveys. Year fixed effects correspond to year of survey fixed effects. *Germany:* Year fixed effects correspond to year of survey fixed effects. Demographic controls include educational attainment level and gender. Worker type categories include laborer, private employee, government employee, self-employed, freelancer, or family caregiver. Industry categories at the 1-digit level. Occupation categories at the 2-digit level (ISCO 88). Firm size is a categorical variable indicating whether the firm where the worker works at has less than 4 workers, 5–9 workers, 10–49 workers, 50–99 workers, 100–499 workers, 500–999 workers and 1000 or more workers. Wage controls include the current hourly wage for the worker. Robust standard errors in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table E.7: Regression of White-Collar Tool Use for External Learners v. Internal Learners

Dep. Variables	White-collar tools		
Germany			
External Learner	0.209***	0.0419***	0.0433***
	(0.00508)	(0.00496)	(0.00845)
Constant	0.373***	0.363***	0.262***
	(0.00458)	(0.0183)	(0.0283)
Observations	87,047	53,163	29,162
R-squared	0.026	0.617	0.623
Year FE		Y	Y
Demographic Controls		Y	Y
Worker type FE		Y	Y
Industry FE		Y	Y
Occupation FE		Y	Y
Firm size FE		Y	Y
Wage Controls			Y

External learner (v. internal learner) and white-collar tools (v. blue-collar tools) described in text. All regressions weighted using observation weights provided in the surveys. Year fixed effects correspond to year of survey fixed effects. *Germany:* Year fixed effects correspond to year of survey fixed effects. Demographic controls include educational attainment level and gender. Worker type categories include laborer, private employee, government employee, self-employed, freelancer, or family caregiver. Industry categories at the 1-digit level. Occupation categories at the 2-digit level (ISCO 88). Firm size is a categorical variable indicating whether the firm where the worker works at has less than 4 workers, 5–9 workers, 10–49 workers, 50–99 workers, 100–499 workers, 500–999 workers and 1000 or more workers. Wage controls include the current hourly wage for the worker. Robust standard errors in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

E.4 Additional Information and Empirical Support for Section 4.5.4

E.4.1 Definition of "Work-Related Novelty" variables

- "Job with Frequent Task Novelty" is a binary variable that indicates whether the interviewee reports always or frequently being faced with new tasks she has to familiarize herself with in her job.
- "Job with Frequent Procedure Improvements" is a binary variable that indicates whether the interviewee reports always or frequently having to improve previous procedures or try something new in her job.

E.4.2 Correlations between different sources of learning and "work-related novelty"

Table E.8: Correlations between different sources of learning and "work-related novelty"

Dep. Variables	External Learning			Internal learning		
Germany						
Freq. Task Novelty	0.177***	0.0758***	0.0704***	-0.0830***	-0.0345***	-0.00773
	(0.00320)	(0.00393)	(0.00492)	(0.00373)	(0.00474)	(0.00618)
Freq. Procedure Improvements	0.124***	0.0686***	0.0623***	-0.0786***	-0.0248***	0.0171***
	(0.00312)	(0.00401)	(0.00498)	(0.00384)	(0.00484)	(0.00601)
Constant	0.540***	0.773***	0.738***	0.369***	0.225***	0.249***
	(0.00221)	(0.0301)	(0.0340)	(0.00239)	(0.0129)	(0.0178)
Observations	170,009	125,369	84,703	106,833	69,275	36,695
R-squared	0.078	0.204	0.207	0.021	0.137	0.079
Year FE		Y	Y		Y	Y
Demographic Controls		Y	Y		Y	Y
Age FE		Y	Y		Y	Y
Worker type FE		Y	Y		Y	Y
Industry FE		Y	Y		Y	Y
Occupation FE		Y	Y		Y	Y
Wage Controls			Y			Y

External learning, internal learning and "work-related novelty" variables described in text. All regressions weighted using observation weights provided in the surveys. *Germany:* Year fixed effects correspond to year of survey fixed effects. Demographic controls include educational attainment level and gender. Worker type categories include laborer, private employee, government employee, self-employed, freelancer, or family caregiver. Industry categories at the 1-digit level. Occupation categories at the 2-digit level (ISCO 88). Wage controls include the current hourly wage for the worker. Robust standard errors in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

F Appendix: Additional Results and Conditions for Simulations of Quantitative Model

F.1 Firms' Choices of Vacancies and Wage Rates

In each period, each firm with productivity z chooses the optimal amount of vacancies $v(z)$ and the associated wage rate per efficiency unit of labor $w(z)$. In the steady state, this firm solves the following problem:

$$\max_{v(z),w(z)} vq(\theta)\sum_i\left(\frac{1}{U+\eta(1-U)}m_U(h_i)+\frac{\eta}{U+\eta(1-U)}\int_{w<w(z)}m(w,h_i)dw\right)V^F(h_i,z)-\frac{c_v v(z)^{1+\gamma_v}}{1+\gamma_v}$$

where $m(w, h_i)$ is the number of workers with wage w and human capital h_i, and $m_U(h_i)$ is the number of unemployed workers, before job search happens. The first-order condition with regard to vacancies $v(z)$ is given by:

$$\underbrace{c_v v(z)^{\gamma_v}}_{\text{marginal costs of vacancies}} = \underbrace{q(\theta)\sum_i\left(\frac{1}{U+\eta(1-U)}m_U(h_i)+\frac{\eta}{U+\eta(1-U)}\int_{w<w(z)}m(w,h_i)dw\right)V^F(h_i,z)}_{\text{benefits of posting a vacancy}}$$

The first-order condition with regard to wage $w(z)$ yields:

$$\underbrace{\sum_i\left(m_U(h_i)+\eta\int_{w<w(z)}m(w,h_i)dw\right)DV_1^F(z,h_i)}_{\text{costs of higher wages—reduced profits per unit of labor}}$$
$$=\underbrace{\sum_i\left(m_U(h_i)+\eta\int_{w<w(z)}m(w,h_i)dw\right)DV_2^F(z,h_i)f(w(z))}_{\text{benefits of higher wages—lower leaving rates of workers}}+\underbrace{\sum_i \eta m(w(z),h_i)V^F(h_i,z)}_{\text{benefits of higher wages—poaching more workers}}$$

where $DV_1^F(z, h_i)$ and $DV_2^F(z, h_i)$ represent changes in the firm's value for a worker with human capital h_i due to reduced wages per labor and due to higher chances of keeping workers, which can be solved recursively (in absolute values).[65] This equation, combined with the min-mean wage ratio b (boundary condition), can solve the wage (similar as in Burdett and Mortensen (1998)):

[65]In particular, we can write $DV_1^F(z, h_i)$ and $DV_2^F(z, h_i)$ in the recursive form,

$$DV_1^F(z,h_i)=h_i+\beta(1-\delta)(1-\delta_{job})(1-\eta\theta q(\theta)\bar{F}(w))\left[p_{learn}(h_i,z)\mathbb{E}DV_1^F(h_{i+1},z)+(1-p_{learn}(h_i,z))\mathbb{E}DV_2^F(h_i,z)\right]$$

$$DV_2^F(z,h_i)=\beta(1-\delta)(1-\delta_{job})\eta\theta q(\theta)\left[p_{learn}(h_i,z)\mathbb{E}V_1^F(h_{i+1},z)+(1-p_{learn}(h_i,z))\mathbb{E}V_2^F(h_i,z)\right]+\beta(1-\delta)(1-\delta_{job})(1-\eta\theta q(\theta)\bar{F}(w))\left[p_{learn}(h_i,z)\mathbb{E}DV_2^F(h_{i+1},z)+(1-p_{learn}(h_i,z))\mathbb{E}DV_2^F(h_i,z)\right].$$

We do not take into account the differential of g with regard to $w(z)$ as in the numerical analysis, firms bear most of the training costs and therefore in most cases $g = g^F$, which implies that we can apply the envelope theorem.

$$w(z) = b\bar{w} + \int_{z_{min}}^{z} \frac{\sum_i \left[\left(m_U(h_i) + \eta \int_{w<w(z')} m(w(z'), h_i) dw \right) DV_2^F(z', h_i) + \eta \frac{m(w(z'),h_i) V^F(h_i,z')}{f(w(z'))} \right]}{\sum_i \left(m_U(h_i) + \eta \int_{w<w(z')} m(w(z'), h_i) dw \right) DV_1^F(z', h_i)} dF(w(z')).$$

G Additional Quantitative Results

G.1 Properties of Equilibrium

Figure G.1: Distribution of Human Capital Levels at Different Ages

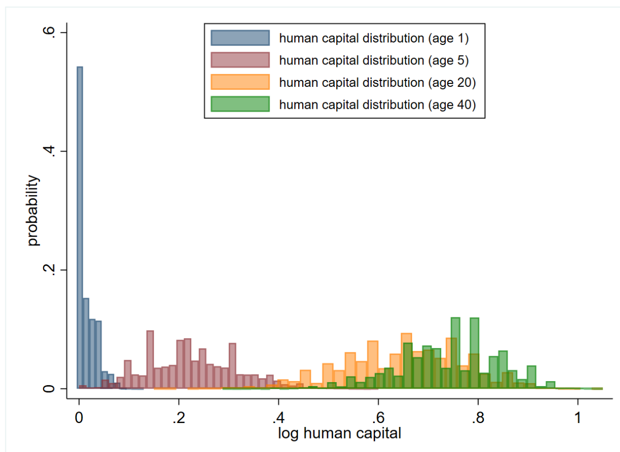
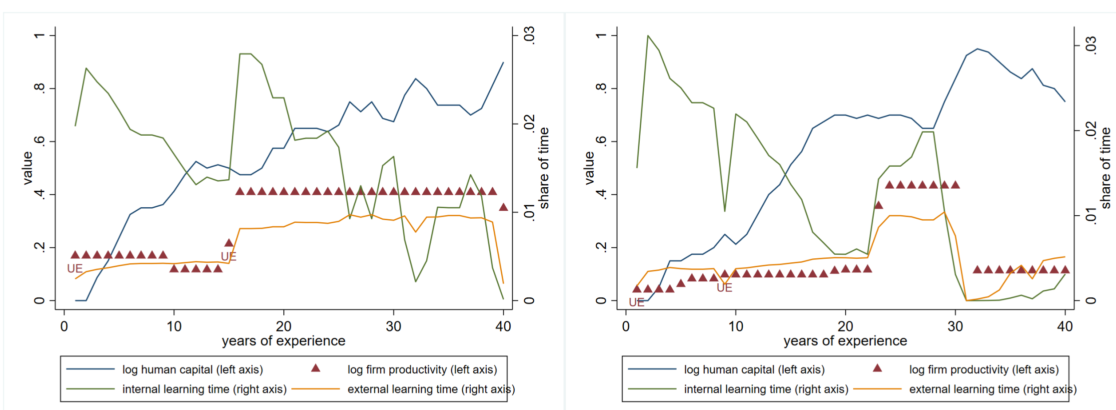


Figure G.2: Examples of Workers' Lifecycles

(a) (b)



Note: "UE" represents that the worker partly experiences unemployment during the current year.

G.1.1 Assortative Matching

In this section, we consider the equilibrium pattern of sorting in the model, both between worker and firm types, and worker and coworker types. Figure G.3a illustrates the distribution of worker types by firm type by presenting the share of workers of each human capital level (vertical axis) given firm productivity level (horizontal axis). We find that firms with higher productivity levels hire relatively larger shares of high-skill workers, consistent with the positive assortative matching patterns between employers and employees documented in the US (Barth et al. (2016), Abowd et al. (2018), Song et al. (2019)). This result is driven by two phenomena in our model. First, the larger learning investments and more favorable coworker learning environments prevalent in more productive firms allow workers to climb the human capital ladder faster. Second, on-the-job search helps these more productive firms poach employed workers who tend to be more skilled than the unemployed.

Figure G.3: Assortative Matching and Workers' Sorting

(a) Distribution of Workers' Skills Across Firms

(b) Distribution of Coworkers' Skills Across Workers

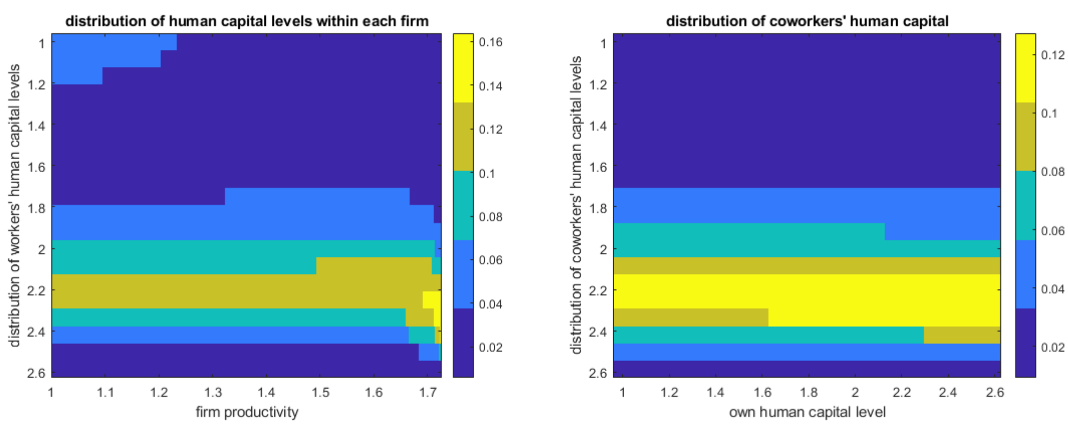


Figure G.3b illustrates the distribution of coworker types by worker type by presenting the share of coworkers of each human capital level (vertical axis) given worker human capital level (horizontal axis). We find that workers with higher human capital levels tend to have coworkers of higher human capital. This is consistent with the sorting of high-skill workers into more productive firms shown in Figure G.3a. Compared with random assignment of workers to different firms, workers' sorting benefits high-skill workers as they now enjoy a better pool of coworkers to learn from.

G.1.2 Comparison with Evidence on Peer Effects

In this section, we compare our quantitative results with Herkenhoff et al. (2018) who use the employer-employee data of US firms and workers to show that a worker's future wage is affected by the average wages of its coworkers in the current firm. Using our calibrated model, we simulate a panel of 10,000 workers for 40 years from the beginning of their career. To ensure comparability with their results, we use a sample of workers from our simulated data who experience an EUE transition: a transition from employment at a firm in year t into unemployment in $t+1$ and then back into employment at a different firm in $t+2$. We replicate their regression by regressing the wage of the worker in $t+2$ on average wages of coworkers in t, controlling for the own wage in t.

Table G.1: EUE Sample: Wage Regressions

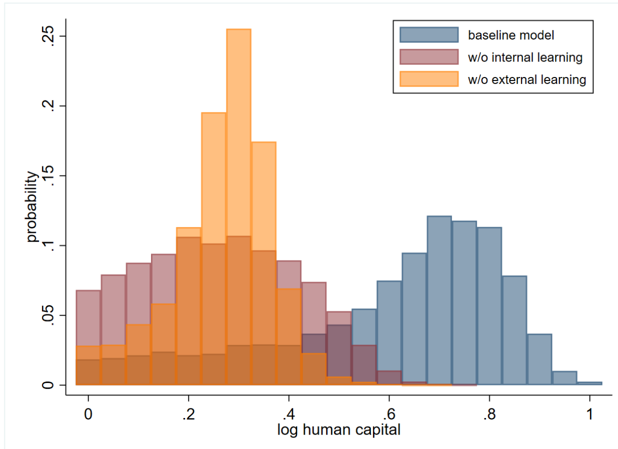
	(1)	(2)	(3)	(4)
Dependent variable	Wage t+2	Wage t+2	Wage t+2	Wage t+2
	Model generated data		Employer-employee data (HLMP)	
Sample	$w_{it} < w_{-i,j,t}$	$w_{it} > w_{-i,j,t}$	$w_{it} < w_{-i,j,t}$	$w_{it} > w_{-i,j,t}$
Coworker Wage, t	0.145***	0.066	0.145***	0.041***
	(0.038)	(0.061)	(0.024)	(0.012)
R-squared	0.963	0.957	0.317	0.488

Notes: HLMP is short for Herkenhoff et al. (2018). The table regresses the log wage of the worker in $t+2$ on the average log wage of coworkers in t, controlling for their own log wage in t. Columns (1)–(2) use our model generated data. To be consistent with controlling for workers' and firms' demographics in Herkenhoff et al. (2018), we control for workers' age and firms' wage per efficiency unit, which are the main demographic variables in the model generated data. Columns (3)–(4) present the regression results from Table 1 in Herkenhoff et al. (2018). Robust standard errors are in parentheses. Significance levels: * 10%, ** 5%, *** 1%.

We present these results in Table G.1. Columns (1) and (2) report the regression results from our model generated data, while Columns (3) and (4) present the corresponding results drawn from Herkenhoff et al. (2018), focusing on two subsamples in which the worker's wage in year t is higher or lower than the average wage of his coworkers. Consistent with Herkenhoff et al. (2018), our quantitative model predicts that for workers paid less than their coworkers in t, their coworkers' wage has a larger positive impact on their wages in the next job in $t+2$ than for workers paid more than their coworkers in t. In our model, this result follows because having coworkers with a higher average wage translates into a more knowledgeable pool of coworkers from whom the worker can learn, which is particularly important for workers who are still young, low human capital, and engaging in internal learning.

G.2 Counterfactual Exercises

Figure G.4: Human Capital Distribution in Baseline and Counterfactual Exercises



G.3 Subsidies to Learning

We now assess the role of subsidies in correcting the inefficiently low levels of learning in our setup arising from the fact that firms determine learning investments but fail to internalize workers', future employers, and economy-wide gains from learning. To this end, we consider government-sponsored subsidies to learning that pay for a portion of firms' overall learning costs (including both production losses and trainers' fees). These subsidies are financed by lump-sum taxes.

Figure G.5: Subsidies to All Learning Costs

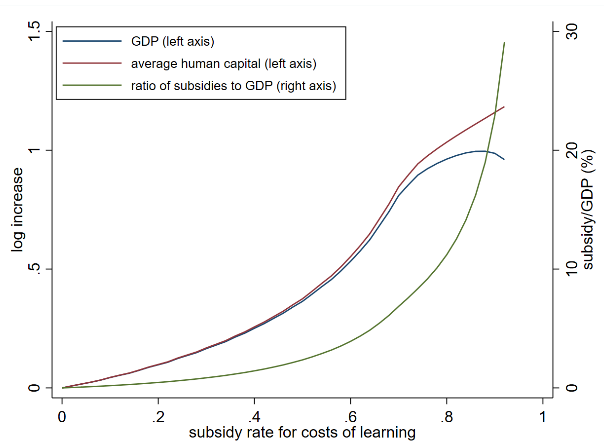
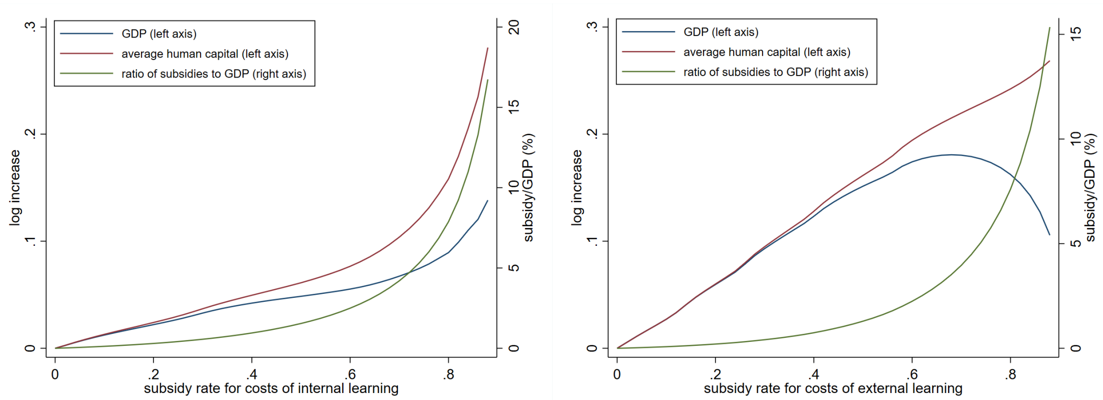


Figure G.5 plots the gains in firms' GDP and average workers' human capital against the government's subsidy rates.[66] We find that with a 40% subsidy rate (corresponding to 1.5% GDP used for learning subsidies), average human capital and GDP increase by 25% in the steady state, indicating very sizeable potential gains from government-sponsored learning policies. We also find that the effect of subsidy rate on GDP is maximized when the subsidy rate is around 85%.

Figure G.6: Subsidies to Each Source of Learning

(a) Subsidies to Internal Learning (b) Subsidies to External Learning



Panels (a) and (b) of Figure G.6 report the results when only internal and external learning are subsidized, respectively. We have two main findings. First, when the government only subsidizes one source of learning, the impact on GDP peaks at lower subsidy rate levels for external learning than for internal learning, reflecting that external learning is more costly than internal learning. Second, the impact of jointly subsidizing both sources of learning on human capital and GDP is much larger than subsidizing each source of learning. For example, Figure G.5 indicates that a subsidy rate of 1% of GDP leads to a 18% gain in GDP and human capital, whereas Figure G.6 indicates that subsidy rates of 1% of GDP to internal and external learning lead only to 4.2% and 14.4% gains in GDP and human capital, respectively. This is consistent with the complementary effects between the two sources of learning described above.

[66] In the model, we consider GDP to be the total output in the production sector net of learning and hiring costs.