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**: February 2022

Abstract

How do the sources of worker learning change over the lifecycle, and how do these changes affect on-the-job human capital accumulation? 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 engage in each type of skill acquisition 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, as well as several key findings in the literature on the effects of firm matching and coworker quality in the formation of human capital. Counterfactual analyses imply that aggregate human capital decreases by approximately 30% in the absence of either learning source, and that the two sources are highly complementary in the aggregate. We conduct a policy analysis that highlights key inefficiencies to learning investments stemming from firms' role in learning, and shows that subsidizing learning can generate sizeable increases to human capital and aggregate output.

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

We would like to thank Titan Alon, Matthias Doepke, David Lagakos, Jeremy Lise, Mark Muendler, Valerie Ramey, Esteban Rossi-Hansberg 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, and the IADB.

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 the dynamics of on-the-job skill 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 (Jarosch et al. (2019), Herkenhoff et al. (2018), Nix (2017)).[1] This literature to date has focused on studying each source of human capital accumulation individually, and thus has not yet considered how these inputs interact to jointly shape on-the-job skill acquisition. In this paper we study how different sources of worker-level skill acquisition interact and jointly shape workers' lifecycle human capital accumulation.

Studying the shared scope of different sources to influence lifetime worker learning is important for several reasons. First, in the context of schooling, researchers have found that multiple factors influence children's human capital acquisition.[2] This literature has also found that these factors are often highly complementary, and that their importance changes as children age (Cunha et al. (2010), Del Boca et al. (2014), Attanasio et al. (2020)).[3] This suggests that studying how multiple sources of learning interact and jointly shape skill acquisition could be crucial to fully understand the process of lifetime on-the-job human capital accumulation. Second, given that the sources of on-the-job skill acquisition identified in the literature are affected by fundamentally different processes, these findings also suggest that the scope of these sources to determine human capital growth will likely change as workers age. This implies, for instance, that the effectiveness of different policies aimed at on-the-job human capital accumulation, such as apprenticeship or re-training programs, may vary greatly depending on the ages or other characteristics of the workers targeted. Third, 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.[4]

[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 Hanushek (2020) for a review of the inputs considered in the education production function.

[3] For example, researchers have found that both nutrition and parental investments of time and resources are key for children's overall human capital, though the former is much more important in early childhood (Attanasio et al. (2020))

[4] 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.

Motivated by the literature and data, we focus on two categories 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. We document two novel facts. First, we show that both internal and external sources of learning are widely offered by firms, and thus largely available to firm workers using enterprise survey data from Europe. Second, we use detailed worker qualification data from Germany and the United States to 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.

We then build a benchmark model to examine 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 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 (or training) services. Production in each of these sectors is carried out by production workers and trainers respectively, 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 the random meeting resulting in a match 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 type 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 our 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 model's key learning predictions: (1) the portion of individuals that do not engage in the two types 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 two natural implications of our theory: (1) better coworkers increase the value of internal learning; and (2) 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 in order to quantify the importance of internal and external sources of learning for human capital accumulation. To this end, we embed the two-source human capital ladder mechanism formalized in the benchmark model within a Burdett-Mortensen structure where firms and workers jointly choose and pay for 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 follows a search and matching framework where firms are heterogenous and post vacancies and wages to attract both unemployed and workers from other firms. 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 type in every period, and that the probability of climbing the human capital ladder depends on the time spent on each type of learning, and the likelihood of finding a colleague or trainer with higher human capital than the own.

We calibrate the model to the United States economy, and find that the share of learning costs borne by firms is 80%, suggesting that firms play a key role in the formation of on-the-job human capital. We also find that the equilibrium of the model replicates the lifecycle results found empirically (non-targeted). In particular, the model generates that as workers' experience increases, the time spent on internal learning declines, while the time spent on external learning first increases and then declines. We also find that workers' human capital grows rapidly during the first few years after entering the labor force, and slows down in later years. This matches the evidence on the lifecycle returns to experience (Rubinstein and Weiss (2006)), and follows both from the depreciation in human capital and the reduction in the scope of learning that occurs as workers climb up the human capital ladder.

The model equilibrium also highlights the importance of firm matching and coworker quality in the formation of human capital, and matches several key findings in the literature. First, we find that workers who are matched with more productive firms spend more time on both internal and external learning. This finding is consistent with evidence found by Engbom (2017) and Arellano-Bover (2020), and stems from the fact that more productive firms exhibit both larger returns to skill acquisition and a better pool of coworkers to learn from. This latter result, in particular, reflects a positive assortative matching pattern per which more productive firms hire relatively larger shares of high-skilled workers that matches evidence on the sorting patterns between employers and employees in the US (Barth et al. (2016), Abowd et al. (2018), Song et al. (2019). In our model, this positive sorting emerges due to (1) the larger learning investments and more favorable coworker 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. 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.

To assess the importance of internal and external learning in the formation of human capital, we perform a counterfactual analysis where we subsequently shut down each of these two sources of skill acquisition in our model. 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. We also find that the distribution of human capital is more dispersed when we shut down internal learning than when we shut down external learning. Without external learning, the opportunities to meet high-skill coworkers are high for low-skill workers, but are depleted more quickly as they climb the human capital ladder. This causes a concentration of the human capital distribution around the middle of the ladder.

Finally, we assess the role of subsidies that pay for a portion of firms' overall learning cost to correct key inefficiencies to learning investments arising from the fact that firms bear the majority of learning costs, and spawn human capital accumulation and economic growth. 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. In addition, we find that the impact of jointly subsidizing both types of learning on human capital and output is much larger than the sum of the individual effects of subsidizing each type of learning due to the key complementarities between the two forms of learning described above.

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 in Section 5 we perform an empirical assessment of its testable predictions. In Section 6 we present the quantitative model, calibration, and results. We conclude in Section 7.

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. Our theory provides a unified structure to jointly understand the interaction between 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 peer learning in knowledge diffusion within coworker and production teams (Jarosch et al. (2019), Herkenhoff et al. (2018) Nix (2017)), Garicano (2000); Garicano and Rossi-Hansberg (2004, 2006), Caicedo et al. (2019)) and within the population at large (Lucas and Moll (2014), Perla and Tonetti (2014), Jovanovic (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)), and the classic labor economics literature that examines the impacts of on-the-job training on worker human capital accumulation and earnings (Hashimoto (1979), Mincer (1988), Bartel (1995), Brown (1989), Altonji and Spletzer (1991), Loewenstein and Spletzer (1999), Frazis and Loewenstein (2005), Booth (1991), Pischke (2001), Booth and Bryan (2005), Konings and Vanormelingen (2015)).

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 on earnings growth (e.g., Bunzel et al., 1999; Rubinstein and Weiss, 2006; Barlevy, 2008; Yamaguchi, 2010; Burdett et al., 2011; Bowlus and Liu, 2013; Bagger et al., 2014; Gregory, 2019). This contrasts with our focus, which is instead on disentangling the contributions of different sources of learning on human capital accumulation. 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 (e.g., 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]

Given the focus on matching between different firm and worker types in our quantitative model, 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), Bagger and Lentz (2019), De Melo (2018), Hagedorn et al. (2017)). Our theory contributes to this literature in two ways. First, sorting in our model arises from incentives to human capital accumulation. This contrasts with most of this literature, which focuses on sorting arising from worker and firm type complementarities. Second, our theory features

[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)).

two different kinds of human capital accumulation which provide asymmetric incentives to sorting and assortative matching.

Our paper also relates to the vast labor literature examining the impacts of 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 availability and lifecycle patterns of different sources of learning, using 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 several facts about on-the-job human capital accumulation process for workers throughout the lifecycle. We include further details on data sources and analysis in Appendix A.

3.1 Data

To document our first fact, namely that firm workers have access to both internal and external sources of learning in our settings studied, 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, 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 focus on enterprises with 10 or more people, providing 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 1.1.

To document our second and main set of facts regarding the patterns of learning 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, and 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). Collectively, these surveys are referred to as the BIBB/IAB and BIBB/BAuA surveys of the current working population. 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 1.2. 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 1.3.

3.2 Fact 1: Both Internal and External Sources of Learning are Available to Firm Workers

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.

Table 3.1: Proportion of Firms in which workers participate in 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	Self Directed Learning
Germany	0.436	0.532	0.44	0.52	0.08	0.15	0.33
France	0.329	0.666	0.19	0.25	0.09	0.08	0.14
UK	0.418	0.532	0.40	0.66	0.20	0.22	0.40
Italy	0.263	0.403	0.26	0.26	0.10	0.04	0.09
Spain	0.186	0.564	0.20	0.35	0.11	0.13	0.21
Poland	0.134	0.219	0.12	0.19	0.05	0.02	0.10
Romania	0.117	0.157	0.08	0.14	0.07	0.05	0.11
Belgium	0.457	0.605	0.36	0.44	0.16	0.16	0.25
Portugal	0.212	0.360	0.22	0.39	0.06	0.10	0.17
Czech Rep.	0.391	0.611	0.24	0.37	0.05	0.09	0.24
Hungary	0.171	0.307	0.21	0.20	0.03	0.06	0.15
Sweden	0.600	0.724	0.50	0.61	0.36	0.10	0.39
Bulgaria	0.163	0.177	0.13	0.23	0.06	0.09	0.09
Denmark	0.485	0.640	0.51	0.45	0.16	0.19	0.36
Slovak Rep.	0.352	0.531	0.43	0.33	0.09	0.19	0.24
Finland	0.352	0.671	0.35	0.39	0.12	0.12	0.33
Norway	0.676	0.694	0.48	0.70	0.33	0.22	0.36
Latvia	0.124	0.271	0.14	0.47	0.06	0.06	0.17
Estonia	0.326	0.552	0.28	0.44	0.16	0.10	0.26
Cyprus	0.185	0.447	0.27	0.33	0.09	0.17	0.14
Luxembourg	0.473	0.580	0.39	0.44	0.17	0.20	0.31
Malta	0.284	0.322	0.32	0.44	0.14	0.14	0.16
Total	0.275	0.433	0.24	0.33	0.10	0.09	0.19

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.

We first distinguish between internal and external CVT by relying on information on the location and instructor affiliation of CVT courses. CVT refers to education 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 require 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 this category the survey distinguishes between five learning sources: Participation in conferences and lectures, Guided-on-the job training, Job rotation, Learning or quality circles, and Self-directed learning. We provide more detailed definitions and characteristics of each source of learning in Appendix 1.1.

In Table 3.1 we show the proportion of firms for which workers participate in each type of learning activity 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. In Table 3.2, we then show that a large contingent of firms surveyed invest in both External and Internal learning activities simultaneously. To do this, we categorize firms that invest in "External CVT" 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 CVT" as those offering Internal CVT Courses, and/or the remaining categories in Other types of CVT Activities. We find in particular, that 44% of the firms surveyed offer both External and Internal CVT opportunities to their employees.[6] This suggests that in the context of Europe, for a large portion of firms, and therefore a large portion of firm workers, internal and external learning sources coexist. This motivates our investigation into how these sources of learning change throughout workers' lifecycles.

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

		External	
		0	1
Internal	0	0.33	0.08
	1	0.14	0.44

3.3 Fact 2: Changes in the Sources of Human Capital Accumulation across Workers' Lifecycles

Using the German BIBB and American NHES data, we document how the importance of different sources of skill acquisition change across workers' lifecycles. To conduct our analysis,

[6]We show robustness to these patterns by using different definitions to build these categories of Internal and External CVT and show the same patterns in Appendix B. Specifically, we show that (1) both the proportion of firms that engage in any CVT courses (internal or external), along with worker hours are significant (Table B.1); (2) the total proportion of workers engaging in Other CVT Activities (Table B.2); (3) distribution of internal and external CVT by type for firms (Table B.3), and (4) Distribution of internal and external CVT by firm size (Table B.4).

we rely on (1) human capital accumulation variables, and (2) potential work experience variables. We construct our main human capital accumulation variables based on the two sources of learning introduced in Section 3.2. First, we construct a measure of "internal learning" that captures workers that reported having learned some skills through colleagues or supervisors. Second, we construct a measure of "external learning" that captures workers that reported participating or having learned some skills through external training. It is important to note here that external learning as defined here focuses on the explicit role of elements outside the firm (such as teachers, instructors, specialist literature, etc) in the formation of human capital. In contrast, internal learning captures learning that draws from knowledge internal to the firm.[7]

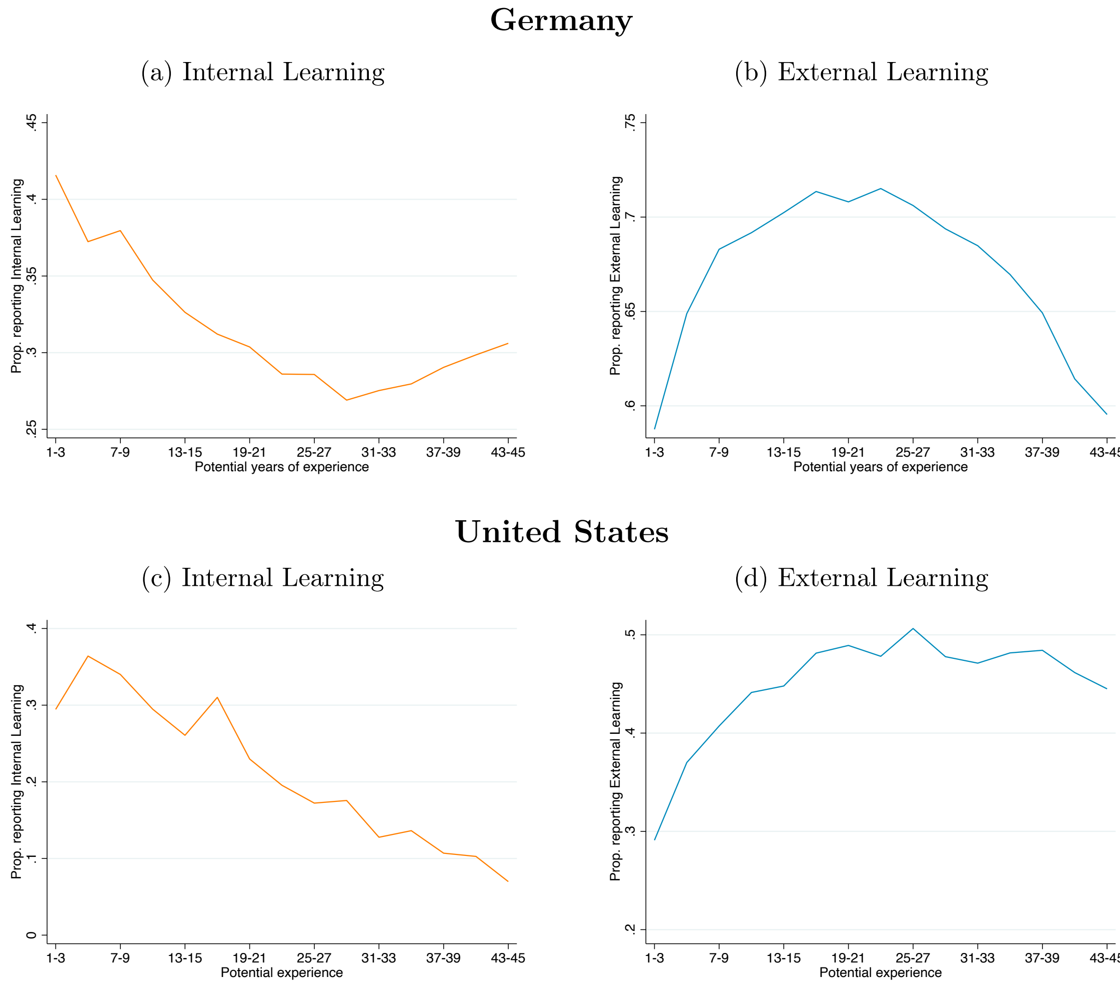
We construct our potential work experience variables using age, educational level and experience variables. Specifically, for both Germany and the US we construct potential years of experience as $Potential\ Experience = Age - Years\ of\ Schooling - 6$. In Appendix 1.2 and Appendix 1.3 we provide further details on the questions and answers used for each one of the variables in the German and American surveys. 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.

We now document differential patterns in the prevalence of each of the types of worker learning along the lifecycle. In Figure 3.1, we plot how the prevalence of workers reporting acquiring the skills necessary for their jobs or participating in each of our learning categories changes with workers' potential experience in Germany and the US. In particular, 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 worker potential experience. These patterns are robust to decomposing the data along many dimensions such as sex, educational level, firm size among others.[8]. In Table C.1, we show these correlations are

[7]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 lifecycle human capital accumulation.

[8]Specifically, the patterns are robust to decomposing by one-year experience bins (Figure C.1), sex (Figure C.2), educational level (Figure C.3), survey wave in Germany (Figure C.4), and firm size in Germany (Figure C.5). We also examine the correlation between internal and external sources of learning across workers in both settings, and find significant negative correlation in Germany, and a positive and significant correlation in the US. We present and discuss these results in detail in Appendix 3.3.

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



statistically significant, even after controlling for several demographic and firm-level variables in both settings. Moreover, in Appendix 3.5 we show that these patterns hold when 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.[9]

In addition, we also show the results are robust to considering alternate working experience variables, namely age[10], and tenure.[11] The first of these results is not surprising given the strong correlation between potential experience and age. The second result, however,

[9]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 3.5.1 for details on the Chilean data and variable construction.

[10]The age results are presented in Figure C.6.

[11]The tenure results are presented in Figure C.7.

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.

Finally, we show that the lifecycle pattern of external learning is supported when we further decompose external learning into finer categories that distinguish between different external sources of learning. Specifically, we explore the lifecycle profiles of the three components of external learning: attendance to seminars or courses, attendance to workshops or reading circles, and reading of trade journals or specialist literature. We present these results in Figure C.8. These results indicate that all three types of external learning follow an inverted U-shape like the one found previously. Interestingly, all of these categories present a very sharp rise at the beginning of worker experience, which stays stable for a significant portion of the worker's life, indicating that external learning is particularly undesirable for very young workers.

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 the 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 workers who has more human capital than them, and thus that they can learn from, shrinks. This implies that the learning from outside experts that occurs by attending seminars or courses may be relatively larger for older than younger workers. In this sense, the patterns observed indicate an intuitive pattern where younger workers tend to concentrate their human capital investments on acquiring 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.

4 Benchmark Model

In this section we develop a model which reproduces and sheds light on the above results. 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 education services by working in each of these sectors respectively, and accumulate human capital by climbing a human capital ladder through internal or external learning.[12]

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 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 start with 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.[13] Workers supply one unit of labor inelastically to the market in each period. Workers' human

[12] We assume here, thus, that 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 transferred to other tasks/firms. An alternative to this would be to distinguish between general and firm-specific human capital, and potentially allow our two sources of learning two contribute differently to them. We focus on general human capital for several reasons. First, 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. Second, we focus on differences between learning sources arising from the pool of knowledge each of them taps into. Although other differences between these sources may exist, they are not necessary to the argument, or to match our empirical findings.

[13] Please see Appendix 4.1 for details on this expected lifetime utility.

capital level h determines their labor productivity. This stock of human capital evolves throughout workers' lives in a logarithmic scale ladder fashion:

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

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. This assumption allows us to get simpler analytical results. We relax it in our quantitative model of Section 6, where we let trainers teach several individuals simultaneously. 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. 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.[14] 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

[14]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.

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.[15] 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 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 is given by[16]:

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

Where:

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

[16] Appendix 4.2 contains a full description of the worker's problem.

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

In this problem, workers choose whether to work, learn internally or externally, or to become trainers. These decisions shape human capital accumulation through time. 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:

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

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 completely trivial. In particular, 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 types of learning. In the quantitative model of Section 6 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. Workers who decide to engage in external learning randomly meet trainers, and after observing the trainer type decide to engage in training or not depending on the trainer's human capital. As such, trainers only produce and get paid if hired. The realized 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 human capital: $p_n(h)$. In short, we can write the expected payout of a trainer with human capital h_i as:

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

.

Probability $p_n(h_i)$ depends on the trainer's human capital level, and the matching probability between external learners and trainers. In particular, trainers produce and get hired and paid only if they have a higher human capital than their matched external learner so they can effectively train her. Therefore, we have that $p_n(h_m) = F_s(h_{m-1})$.

In Appendix 4.3, we define the stationary equilibrium in this model.

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

In Appendix 4.4, 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 three important issues to note here. First, there are two types of equilibria possible in this model: an equilibrium with external learning, and an equilibrium without 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. The second important issue to note is that 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 state-space. Since both of these will depend on an unique object, the price of training q, multiple equilibria arise. [17] 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 state-space. This is captured in Assumption 5, which indicates that trainers locate at all levels of human capital beyond a certain step.[18]

The final 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

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

[18] We test this formally in Section 5.1.

such policies within the context of our quantitative model in Section 6.5.

These conditions are sufficient to deliver the following results, which characterize the learning decisions and lifecycle of workers in equilibrium.[19]

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: $\forall h_m \in [h_{\underline{M}}, h_M],\ f_t(h_m) > 0$.*

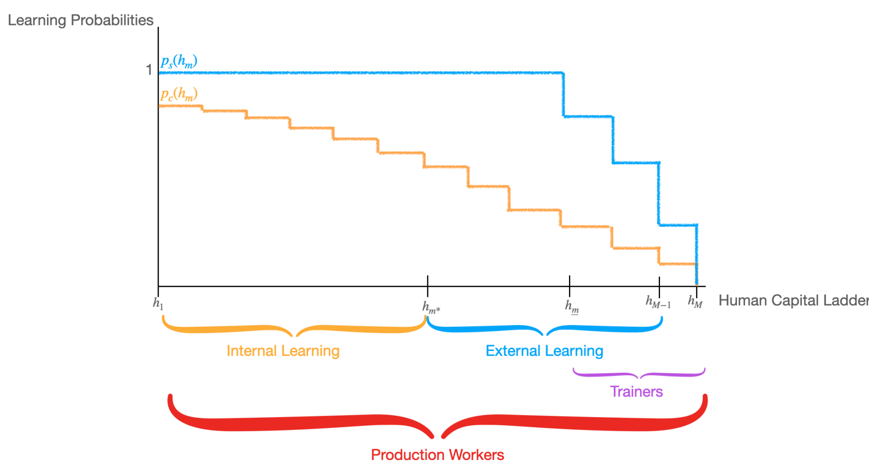
Proof: See Appendix 4.6.8.

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, because trainer wages are contingent on the fraction of external learners that a trainer can effectively train. Specifically, there are no trainers that place from h_1 to h_{m*} due to the lack of external learners and thus training production opportunities in this space. 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. This 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.

[19] In Appendix 4.5, we present additional results that characterize the equilibrium in this model.

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 4.6.9.

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 where the remaining lifetime and decisions are independent of age 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 match 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 4.6.10.

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 mentioned above, at lower levels of human capital and thus at younger ages, workers' learning mode of choice is internal learning, making the portion of external learners zero. 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.[20]

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.

[20]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.

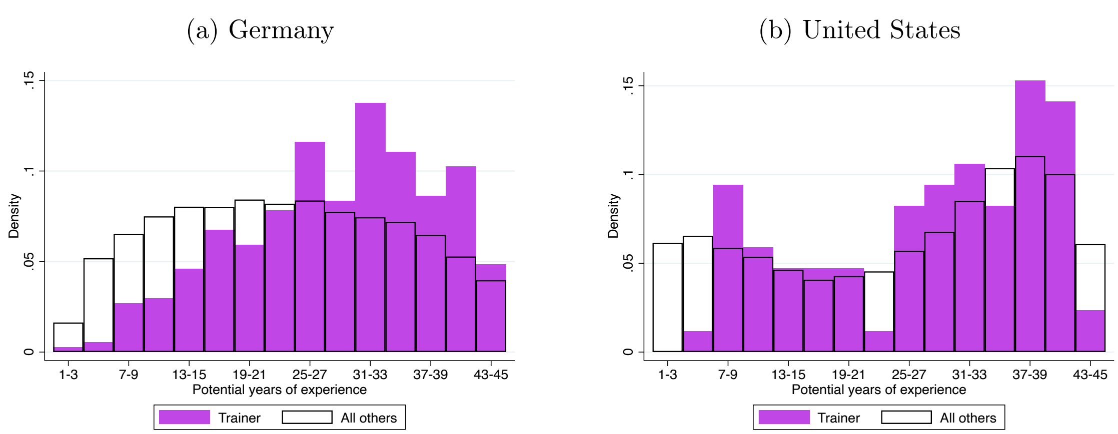
5 Evidence on Testable Predictions

Our model yields a series of testable predictions we can examine in the data. In particular, there are five 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 in our model. The last two correspond to some direct implications of our theory.

5.1 Distribution of Trainers across Human Capital Ladder is Skewed Left

Our model relies on two key results to characterize the distribution of trainers across the human capital state space. First, our model implies that the human capital distribution of trainers has a a higher starting point, median and mean relative to the distribution of production workers.[21] This follows from the fact that production workers locate at all levels of human capital due to the concavity of the production function, while trainers only place at the upper end of the human capital ladder given that their expected wages are contingent on the fraction of external learners that they can effectively train. Second, we assume in Assumption 5 that trainers locate at all levels of human capital beyond this starting point. This assumption allowed us to focus on a specific type of equilibrium.

Figure 5.1: Histograms of potential experience for trainers and production workers



In order to provide support for these two modeling elements, we now examine whether the distribution of trainers has (1) a higher starting point, median and mean relative to the

[21] This is captured formally in Lemma 7.

distribution of production workers; and (2) that this former distribution spans across all human capital levels in the latter part of the human capital state space. To this end, in Figure 5.1, we plot the histograms of trainers and production workers in both Germany and the US by potential experience, and test whether the 25th percentile, median and 75th percentile for the distribution of the former are different from that of the latter using quantile regressions.[22]

The plots show that the distributions of trainers in both Germany and the US heavily concentrate among higher levels of potential experience relative to other workers. In particular, the distributions of trainers have very low mass at lower potential experience levels compared to production workers, but have significant levels of mass at all higher experience levels.[23] 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 larger (though not always statistically significant) than that of production workers in both settings, even after controls.

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

One key prediction of our model is that the portion of workers who do not learn from either of the two sources rises with human capital. This stems from two facts: (1) the rise in the opportunity cost of learning as human capital accumulates stemming from higher wages (captured in Corollary 1, and especially important in the lower part of the human capital ladder); and (2) the decline in the need for learning as progressively more individuals reach the last human capital level (captured in Proposition 2, and especially important in the upper part of the human capital ladder).

We provide evidence for this prediction using our German data. We construct a measure

[22] 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 5.1.1 we provide further details on the construction of the trainer and production worker variables in the German and American surveys.

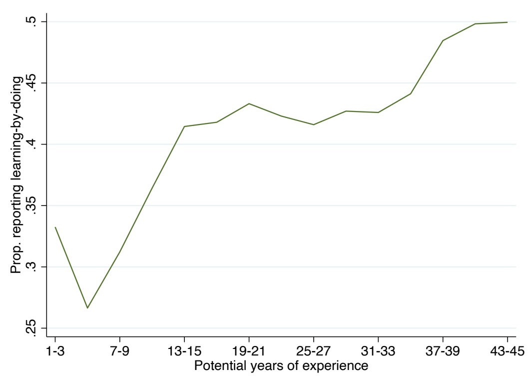
[23] In Figure E.1 and Table E.2, we compare the distribution of trainers to the distribution of external learners across these variables, showing similar patterns.

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.[24] In Figure 5.2 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. This is broadly consistent to decomposing the data by one-year experience bins (Figure E.3a), sex (Figure E.3b), educational level (Figure E.3c), survey wave (Figure E.3d), and firm size (Figure E.4a); and considering the age (Figure E.4b), or number of years with current employer as an alternate working experience variable (Figure E.4c). For the latter, however, the portion of employees with no explicit learning investments only increases sharply towards the end of the lifecycle. In addition, 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 5.2: Prevalence of Learning-by-Doing throughout workers' lifecycles in Germany

Germany

(a) Learning-by-Doing



5.3 Human Capital Ranking across Different Types of Workers

Our 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 given the high probability of matching with a more knowledgeable coworker prevalent at lower levels of human capital. In contrast, trainers concentrate in the higher part of the human capital distribution given the high probability of matching with an external learner

[24]Please see Appendix 5.2.1 to see details on the contruction of this variable.

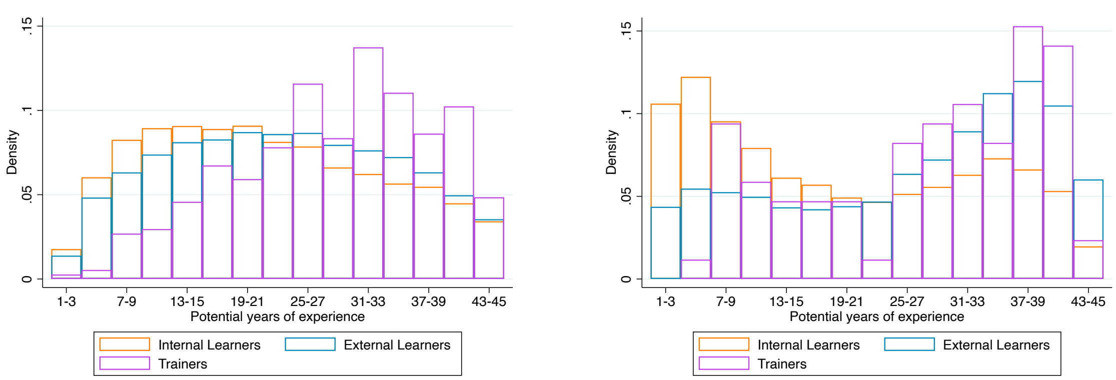
they can effectively train at these levels. External learners, on the other hand, place in between these two groups since in order to justify paying the cost of training, the probability of matching with a coworker needs to be need low enough while the probability of matching with a more knowledgeable trainer needs to be high enough.

We provide evidence for this prediction using our German and American data. We follow two approaches to do this. First, 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 5.3.[25] 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. We then formally test these distribution 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) in Table E.5. 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.

Figure 5.3: Histograms of potential experience for each type of worker

(a) Germany

(b) USA

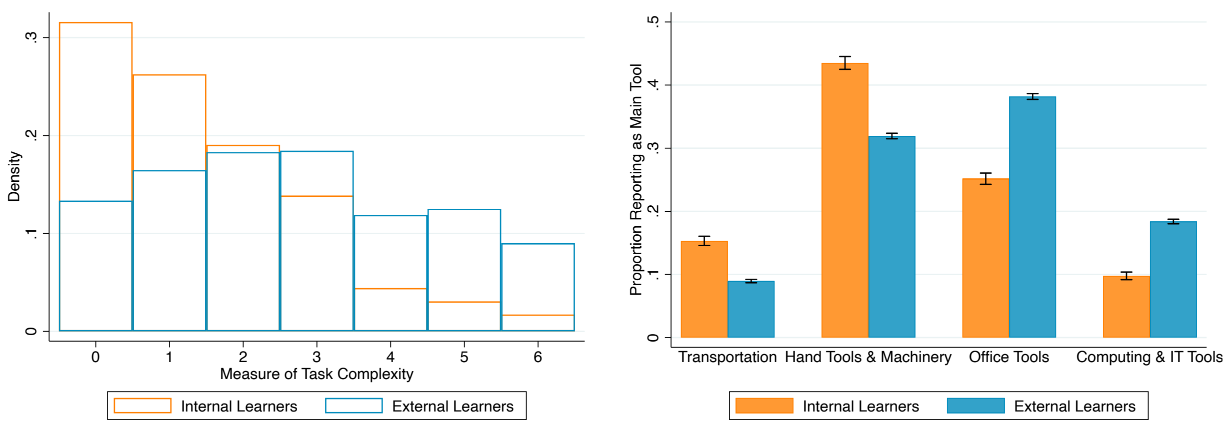


Second, we provide a test of this prediction that focuses on exploring potential differences in the skill-content of the work tasks performed by workers that report internal learning, and

[25]We define individuals who report internal and external learning as in Section 3; see Appendix 1.2 and Appendix 1.3 for details. In addition, we define trainers as in Section 5.1; see Appendix 5.1.1 for details.

those who report external learning.[26] To do this, 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.[27] 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 5.4 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.[28] The results from these regressions indicate that median of task complexity for external learners is larger than that of internal learners.

Figure 5.4: Histograms of task complexity for Internal and External Learners



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.[29] 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

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

[27] Please see Appendix 5.3.1 for details on the construction of these skill variables.

[28] 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.

[29] Please see Appendix 5.3.2 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 or computers). In Panel (b) of Figure 5.4 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.

5.4 More or Better Coworkers Increase the Value of Internal Learning

Our model predicts that the value of internal learning depends on the relative position of each worker in the human capital ladder of the production firm, and thus on the size of the mass of coworkers with higher human capital. Although our benchmark model does not feature firm heterogeneity, a natural implication of this prediction (which bears out in the quantitative model when we include firm heterogeneity) is that the value of internal learning is higher when the workforce of a firm is larger or more skilled.

A few papers have shown that coworker quality influences workers' human capital and wage growth. Herkenhoff et al. (2018) find that having more knowledgeable coworkers induces workers to catch up to them using theory and data from the US. Relatedly, Jarosch et al. (2019) show that having more highly paid coworkers is strongly associated with future wage growth in the context of Germany. Nix (2017) shows that workers with coworkers with a higher educational level experience larger increases in wages in Sweden, and that this effect is concentrated among younger workers.

In this section, we provide direct evidence linking the value of internal learning to the availability of coworkers to learn from. To do this, we rely on our German data and specifically on measures of firm size to proxy for coworker availability.[30] A larger firm size does not only increase the sheer number of coworkers a worker can possibly learn from, but also correlates with higher quality of coworkers, as proposed theoretically by Kremer (1993), and investigated empirically by Oi and Idson (1999) among others.

[30] Our US data does not contain information on firm size.

Figure 5.5: Prevalence of Internal Learning by Firm Size

In Figure 5.5 we show that the proportion of workers engaging in internal learning is higher for larger firms across all experience levels.[31] We show the existence of this correlation more formally by regressing the internal learning variable on firm size. We document our findings in Table E.7, and show that (1) workers in firms with more than 5 workers are more likely to report internal learning; (2) the size of this correlation increases with firm size.

5.5 Workers who Use Innovative Techniques Rely More on External Learning

We now present evidence supporting an additional implication of our theory. Each of the two sources of learning featured in the model relies on different pools of knowledge to encourage human capital accumulation. Internal Learning relies on internal firm knowledge in the form of coworker-sponsored learning to encourage skill acquisition, while external learning draws on external knowledge from trainers for this purpose. This distinction is key in the model, since it allows us to theoretically separate the relative costs and benefits of each of these two sources at every point of the worker's lifecycle. A natural implication of this learning distinction is that 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. In contrast, workers whose jobs do not require the use of innovative techniques or tasks are better served by engaging in internal learning.

We provide evidence for these predictions using our German data. To do this, 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

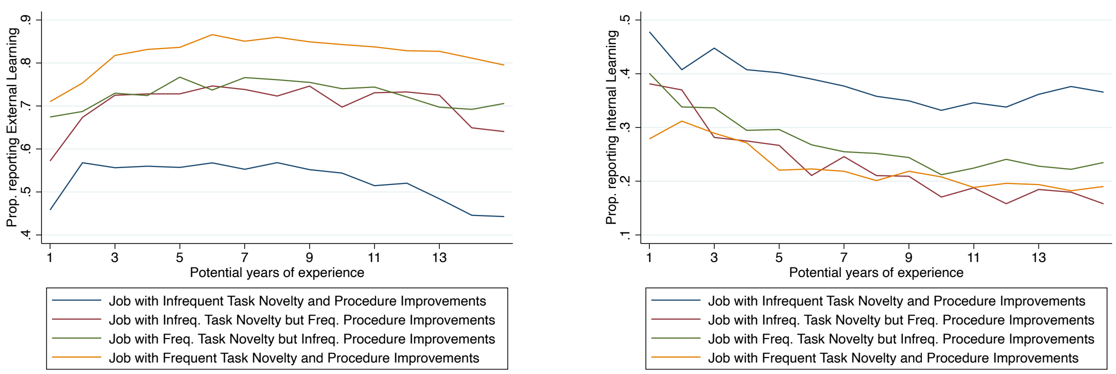
[31] In Figure E.5 we show this pattern holds when we consider more disaggregated firm sizes too.

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).[32]

In Panel (a) of Figure 5.6 we show that the proportion of workers engaging in external learning is higher for workers who report a higher degree of "work-related novelty". In particular, 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), on the other hand, we show that the opposite is true for internal learning, namely that the proportion of workers engaging in internal learning decreases with "work-related novelty".

Figure 5.6: Prevalence of Internal and External Learning by "Work-related Novelty"

(a) External Learning (b) 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 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. These results match the intuition in our model that external learning draws on knowledge that is outside the firm, and is thus better for workers with a need for novelty since it allows them to tap into knowledge that is not currently available in the firm.

[32]Please see Appendix 5.5.1 for details on the construction of these measures.

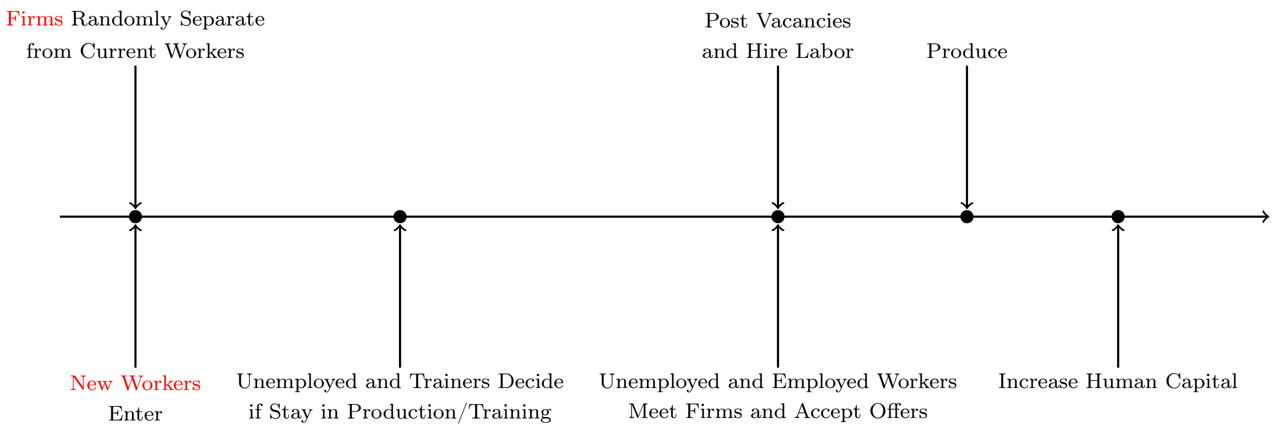
6 Quantitative Analysis

In order to quantify the importance of internal and external learning on lifecycle human capital accumulation, we now embed our two-source learning mechanism in a quantitative framework. We calibrate the model to the US economy. We examine the equilibrium properties of our model, and perform counterfactual and policy exercises that measure the importance of internal and external learning on lifecycle human capital accumulation, and evaluate the scope of learning subsidies to spawn human capital accumulation and economic growth.

6.1 Quantitative Model Setup

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 (e.g., Acemoglu, 1997; Acemoglu and Pischke, 1999; Moen and Rosén, 2004), we embed our analytical model into a Burdett-Mortensen framework with labor market frictions and joint learning decision-making by workers and firms. 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 start with a human capital level of h_1. Individuals accumulate human capital throughout their lives, and can engage either in the production sector or the training (or external learning) sector in every period. We assume that the training sector is frictionless, while the production sector is characterized by frictional labor markets with heterogeneous firms that post vacancies and wages to attract workers. 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 type in every period, and that the probability of climbing the human capital ladder depends on the time spent on each type of learning, and the likelihood of finding a colleague or trainer with higher human capital than the own. The timing of the model is presented in Figure 6.1 and will be described in more detail below.

Figure 6.1: Timing of Events in Each Period



Production Sector and Frictional Labor Market

We consider that the production sector is characterized by frictional labor markets (Burdett and Mortensen, 1998). In the production sector, we have a measure $\bar{M}$ of production firms which produce a homogeneous good and differ in their productivity level $z \sim H(z) = \exp(-z^{-\kappa})$. Firms post vacancies $v(z)$ at the start of each period, with a contract stipulating the wage rate $w(z)$. 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 $F(w(z)) = \int_{z_{min}}^{z} v(z') dH(z')/V$.[33]

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 to maximize their utility. 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 exceeding their current one. We denote the total number of job searchers as $\tilde{U}$, which includes the unemployed and 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 $q(\theta) = \frac{c_M V^{1-\phi} \tilde{U}^{\phi}}{V}$

[33]This wage offer distribution uses the result that $w(z)$ is increasing in productivity z as shown below and in Burdett and Mortensen (1998).

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, 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 wages 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 wages, combined with the min-mean wage ratio b (boundary condition). Please see Appendix 6.1 for details.

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 wage 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}: $p_n(h_i) = F_s(h_{i-1})$.

Joint Decision of Learning

Firms and workers in the production sector jointly decide learning investments. 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 moves up the human capital ladder is given by:

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

$p_c(h_i)$ and $p_s(h_i)$ denote the probabilities of matching with a higher-human-capital worker or trainer, while 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.[34] $0 < \gamma < 1$ 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, and the lower bound h_1 captures the basic skill level workers are born with.

Firms pay a fraction μ of total learning costs, while workers pay the rest. 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 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 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 with human capital level h_i in firm

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

z, the firm's value function can be recursively written as:

$$V^F(h_i,z) = \max_{g^F,g_c^F} \underbrace{(z-w(z))h_i}_{\text{labor revenue}} - \underbrace{\mu\left[\delta_s z h_i g_c^F + (\delta_s z h_i + qp_s)(1-g_c^F)\right] g^F}_{\text{learning costs borne by the firm}}$$
$$+ \underbrace{\beta(1-\delta)(1-\delta_{job})(1-\eta\theta q(\theta)\bar{F}(w))\mathbb{E}\left[p_{learn}V^F(h_{i+1},z) + (1-p_{learn})V^F(h_i,z)\right]}_{\text{the firm's future value if the worker stays}} \tag{2}$$

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 worker's value function is given by:

$$V^W(h_i,z) = \max_{g^W,g_c^W} \underbrace{w(z)h_i}_{\text{wage rate}} - \underbrace{(1-\mu)\left[\delta_s z h_i g_c^W + (\delta_s z h_i + qp_s)(1-g_c^W)\right] g^W}_{\text{learning costs borne by the worker}}$$
$$+ \underbrace{\beta(1-\delta)(1-\delta_{job})(1-\eta\theta q(\theta)\bar{F}(w))\mathbb{E}\left[p_{learn}V^W(h_{i+1},z) + (1-p_{learn})V^W(h_i,z)\right]}_{\text{the worker's future value if stays at current firm}}$$
$$+ \underbrace{\beta(1-\delta)(1-\delta_{job})\eta\theta q(\theta)\int_{w(z')>w} \mathbb{E}\left[p_{learn}V^W(h_{i+1},z') + (1-p_{learn})V^W(h_i,z')\right] dF(w(z'))}_{\text{the worker's future value for job-to-job transitions}}$$
$$+ \underbrace{\beta(1-\delta)\delta_{job}\mathbb{E}\left[p_{learn}\max\{V^U(h_{i+1}),V^{TR}(h_{i+1})\} + (1-p_{learn})\max\{V^U(h_i),V^{TR}(h_i)\}\right]}_{\text{the worker's future value for job separations}} \tag{3}$$

where the value functions of unemployment and becoming a trainer are given by:

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

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

From these value equations, and given the wage rate $w(z)$[35], 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. The proportion of learning time spent on internal learning is given by:

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

[35]The equations describing wage $w(z)$ are presented in Appendix 6.1.

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, $\frac{\delta_s zh}{\delta_s zh + qp_s}$, which prevails when workers are low in the human capital ladder, also contributes to a higher proportion of learning time spent on internal learning.

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

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

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

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 firm's and the worker's values from climbing one more step in the human capital ladder, respectively.[36] We use the term $\sum_{i'\in j, i'\neq i}\frac{\partial V^F(h_{i'},z)}{\partial h_i}$ to capture how a one-step increase in worker i's human capital benefits other workers within the same firm, as other workers now enjoy a better pool of colleagues to learn from.[37] 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.

6.2 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.

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

[37] 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.

6.2.1 Exogenously Calibrated Parameters

Table 6.1 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.

Table 6.1: 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

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).

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 surveys. 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.

6.2.2 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.

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' 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 6.2.

Table 6.2: 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).

Table 6.3 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 6.2.

Table 6.3: 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

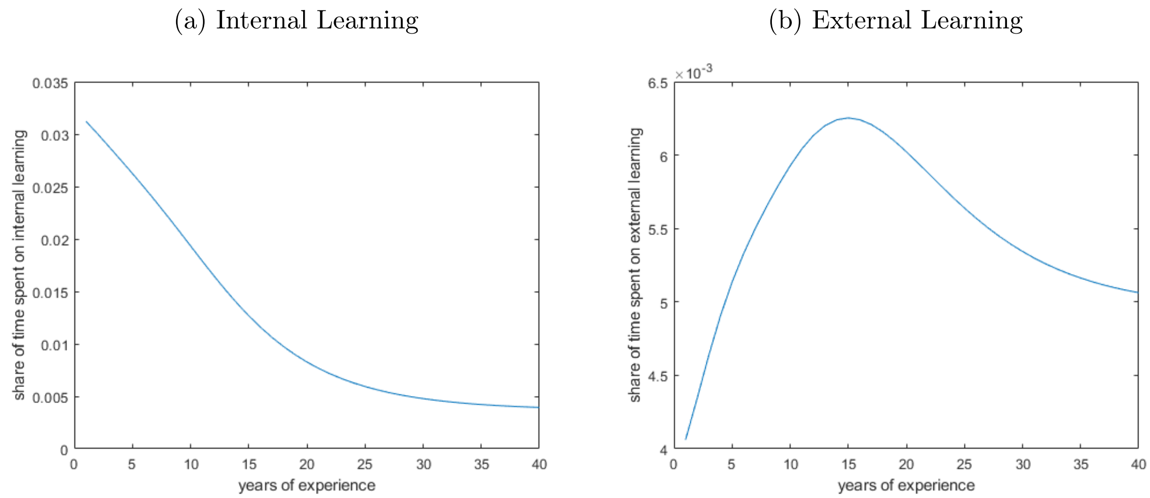
6.3 Properties of Equilibrium

6.3.1 The Lifecycle of Learning

Figure 6.2 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.

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.

Figure 6.2: Lifecycle Patterns of Internal and External Learning



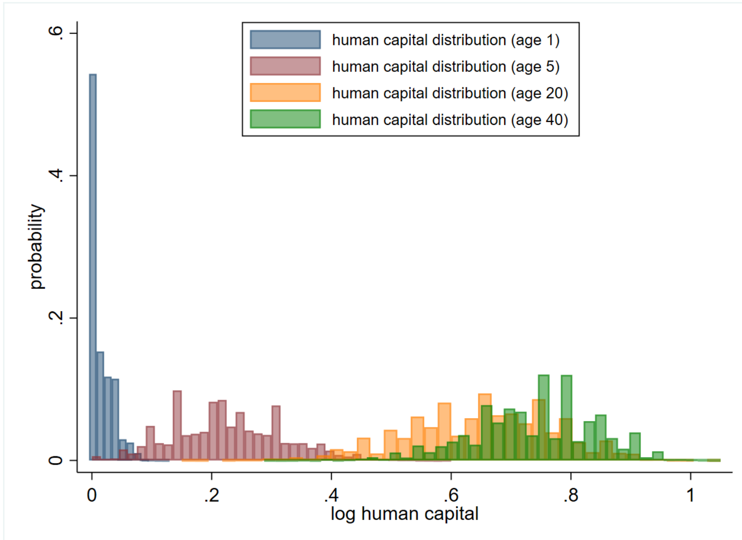
Note: These figures depict total time spent in each type of learning, namely $g_c \times g$, and $(1 - g_c) \times g$.

6.3.2 Human Capital Distribution and Assortative Matching

Now we explore how the distribution of human capital changes as workers age. Figure 6.3 plots 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 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.

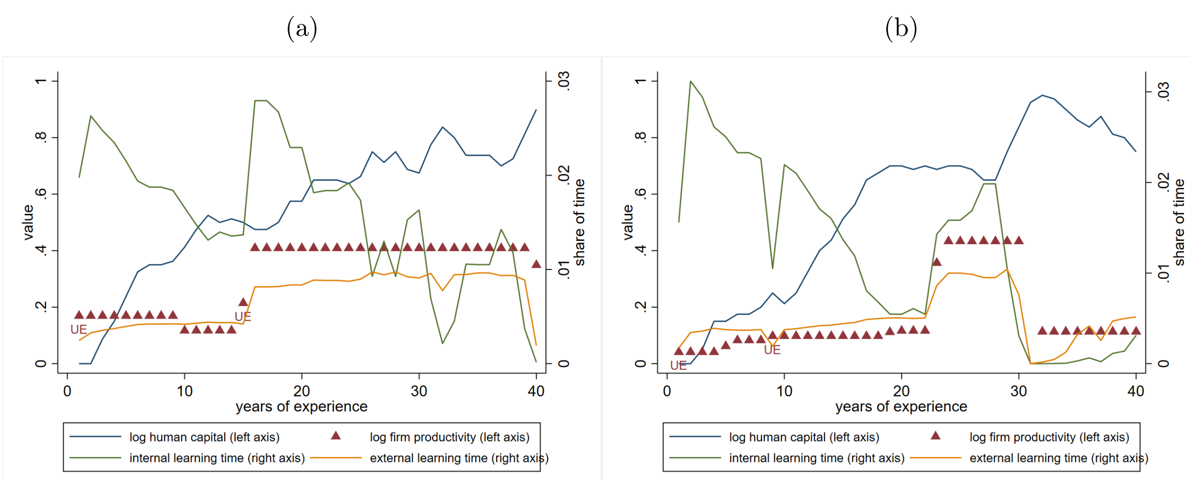
To further understand how workers' human capital is affected by their relative positions in the human capital ladder and their employers' characteristics, we plot the lifecycle human capital and learning patterns for two workers in Figure 6.4, along with their employers' productivity levels. This analysis illustrates two primary results of the model. First, young workers with a low level of human capital learn mostly from internal sources, and substitute away from these and towards external sources as they climb the human capital ladder. This is consistent with the aggregate lifecycle patterns in the model documented above and the empirical evidence. Second, workers who are matched with more productive firms (which

Figure 6.3: Distribution of Human Capital Levels



also pay higher wages) spend more time on both internal and external learning. This stems from the fact that more productive firms exhibit both larger returns to skill acquisition, and a better pool of coworkers to learn from as shown below. This finding is consistent with the evidence found by Engbom (2017) and Arellano-Bover (2020) using data from the OECD and Spain, respectively, and suggesting workers train more when employed at high-paying or large firms.

Figure 6.4: Examples of Workers' Lifecycles

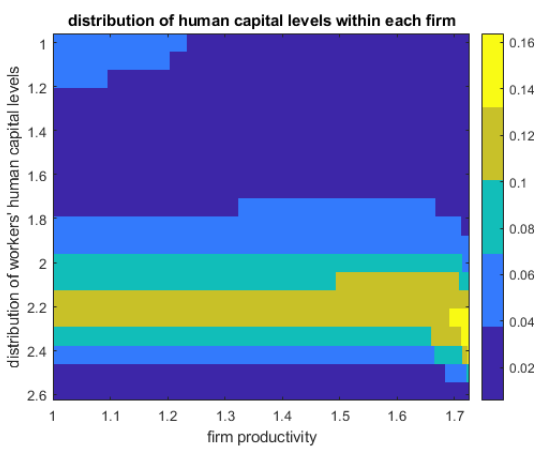


Note: "UE" represents that the worker partly experiences unemployment during the current year.

We now consider the equilibrium pattern of sorting in the model, both between worker and firm types, and worker and coworker types. Figure 6.5a 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 (e.g., 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 6.5: Assortative Matching and Workers' Sorting

(a) Distribution of Workers' Skills Across Firms



(b) Distribution of Coworkers' Skills Across Workers

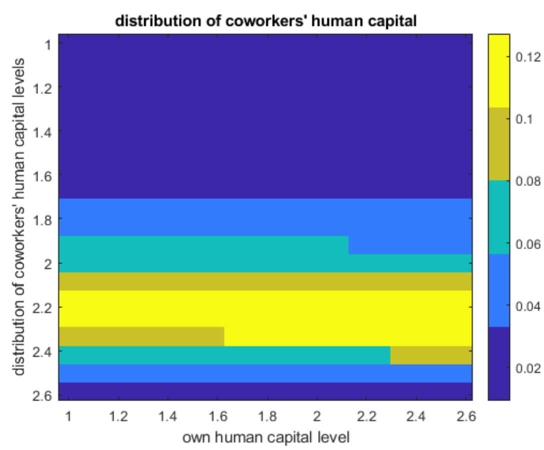


Figure 6.5b illustrates the distribution of coworker types by worker type by presenting the share of coworkers of each human capital level (vertical axis) given worker human capital level (horizontal axis). We find that workers with higher human capital levels tend to have coworkers of higher human capital. This is consistent with the sorting of high-skill workers into more productive firms shown in Figure 6.5a. 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.

6.3.3 Comparison with Evidence on Peer Effects

To further validate our model, 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 6.4: 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%.

Columns (1)–(2) report the regression results from our model generated data, and Columns (3)–(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 a higher average wage of coworkers implies a better pool of coworkers from whom the worker can learn, thus leading to faster human capital growth and higher wages in the next job.

6.4 Counterfactual Analysis: Role of Each Source of Learning

To assess the role of each source of learning in explaining aggregate human capital accumulation, we now perform counterfactual exercises which subsequently shut down each of the two sources of learning, and examine the prevailing human capital level in each of these scenarios. 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.

Table 6.5: 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 Types of Learning	0	0	1.10

Table 6.5 summarizes the shares of time spent on each source of learning and workers' average human capital levels in our baseline model, and 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 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.

Figure 6.6: Human Capital Distribution in Baseline and Counterfactual Exercises

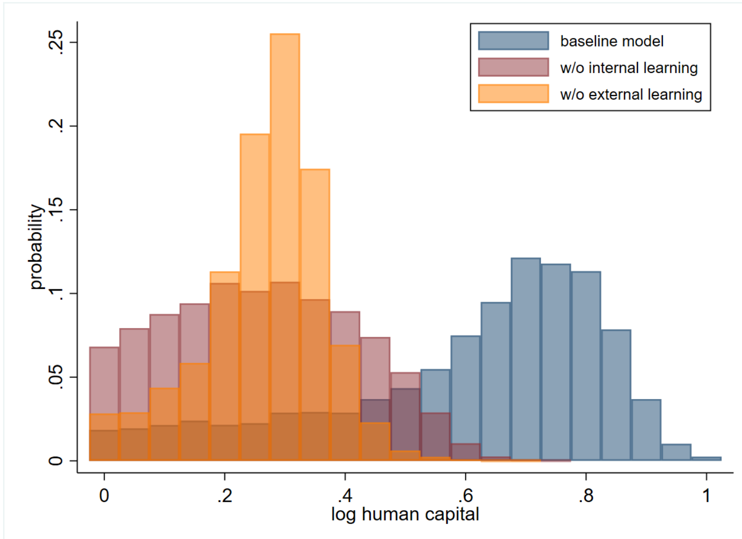
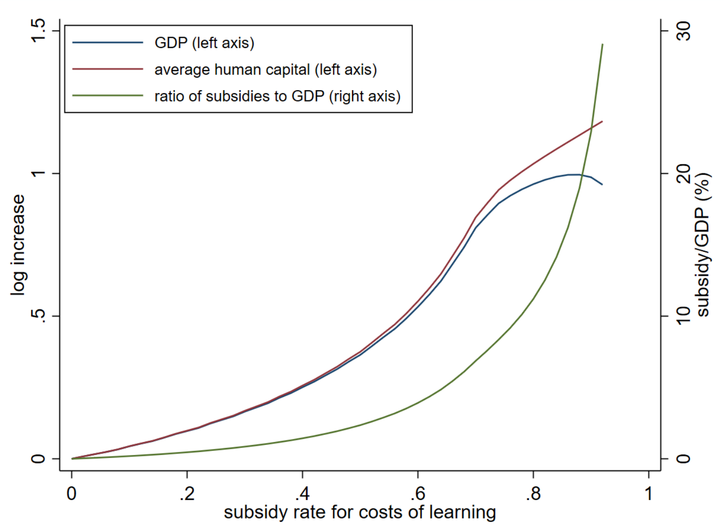


Figure 6.6 plots the distribution of human capital across workers in our benchmark model, and the counterfactual scenarios. Unsurprisingly, we find that shutting down either source of learning leads to lower human capital levels relative to the baseline case. Interestingly, we also find that the distribution of human capital is more dispersed when we shut down internal learning than when we shut down external learning. This is because without external learning, the opportunities to meet high-skill coworkers are still high (and cheap) for low-skill workers, but are depleted more quickly as they climb the human capital ladder. This causes the distribution of human capital to heavily concentrate around the middle of the ladder. Without internal learning, skill acquisition is more expensive for low-skill workers, but learning opportunities are depleted less quickly since trainers are positively selected. This causes the distribution of human capital to be more dispersed.

6.5 Subsidies to Learning

Firms bear the majority of learning costs both empirically and in our quantitative setup. 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.[38]

Figure 6.7: Subsidies to All Learning Cost



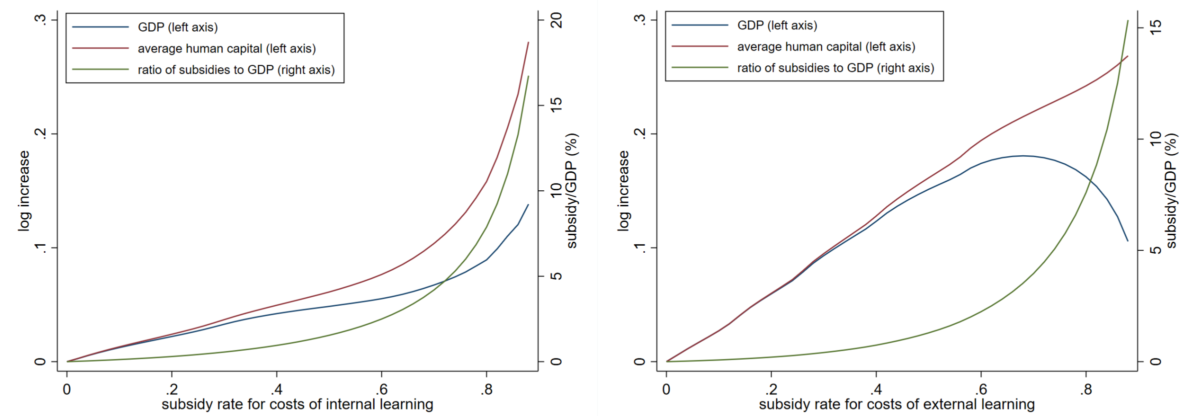
We now assess the role of subsidies in correcting these inefficiencies and spawning human capital accumulation and economic growth. In particular, we consider government-sponsored subsidies to learning that pay for a portion of firms' overall learning costs (including both production losses and trainers' fees). These subsidies are financed by lump-sum taxes. Figure 6.7 plots the gains in firms' GDP[39] and average workers' human capital against the government's subsidy rates. 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%.

[38]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.

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

Figure 6.8: Subsidies to Each Type of Learning Costs

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



Panels (a) and (b) of Figure 6.8 report the results when only internal and external learning are subsidized, respectively. We have two main findings. First, when the government only subsidizes one type 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 types of learning on human capital and GDP is much larger than the sum of the individual effects of subsidizing each type of learning. For example, Figure 6.7 indicates that a subsidy rate of 10% of GDP leads to a 50% gain in GDP and human capital, whereas Figure 6.8 indicates that subsidy rates of 5% of GDP to internal and external learning lead only to 10% and 7% gains in GDP and human capital, respectively. This is consistent with the complementary effects between the two types of learning described above.

7 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. First, we use enterprise survey data from Europe to show that both internal (learning through colleagues) and external (learning through external training) learning are widely offered by firms, and thus largely available to firm workers. Second, we use detailed worker qualification data from Germany and the US to document that the prevalence of internal learning decreases with worker experience; and that the prevalence of external learning has an inverted U-shape in worker experience. We build a model where the incentives to engage in each type 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 and with joint firm and worker learning decisions. We use this quantitative framework to assess the importance of each source in explaining lifecycle human capital growth, and examine the impact of subsidies to learning costs on human capital accumulation and economic growth.

Our results have several implications for understanding economic growth and conducting policy. 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, it is important to note 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.

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.

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

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

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

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

Altonji, J. G. and Spletzer, J. R. (1991). Worker Characteristics, Job Characteristics, and the Receipt of On-the-Job Training. *Industrial and Labor Relations Review*, 45(1):58.

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

Attanasio, O., Meghir, C., and Nix, E. (2020). Human Capital Development and Parental Investment in India. *Review of Economic Studies*, 87:2511–2541.

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

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

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.

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

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

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

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.

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

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

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

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.

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.

Booth, A. L. (1991). Job-Related Formal Training: Who Receives It and What Is It Worth? *Oxford Bulletin of Economics and Statistics*, 53(3):281–294.

Booth, A. L. and Bryan, M. L. (2005). Testing Some Predictions of Human Capital Theory :. *The Review of Economics and Statistics*, 87(2):391–394.

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.

Brown, J. N. (1989). Why Do Wages Increase with Tenure? On-the-Job Training and Life-Cycle Wage Growth Observed within Firms. *The American Economic Review*, 79(5):971–991.

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.

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

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

Burstein, A. and Monge-Naranjo, A. (2007). Foreign Know-How, Firm Control, and the Income of Developing Countries. *NBER Working Paper Series*.

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.

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.

Cunha, F., Heckman, J. J., and Schennach, S. (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*, 78(3):883–931.

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

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

Del Boca, D., Flinn, C., and Wiswall, M. (2014). Household Choices and Child Development. *Review of Economic Studies*, 81:137–185.

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.

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

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

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

Engbom, N. (2017). Worker Flows and Wage Growth over the Life-Cycle: A Cross-Country Analysis.

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*.

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

Frazis, H. and Loewenstein, M. A. (2005). Reexamining the returns to training: Functional form, magnitude, and interpretation. *Journal of Human Resources*, 40(2):453–476.

Garicano, L. (2000). Hierarchies and the organization of knowledge in production. *Journal of Political Economy*, 108(5):874–904.

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

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

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

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

Hanushek, E. (2020). Education production functions. In Bradley, S. and Green, C., editors, *The Economics of Education*, pages 161–170. Academic Press, Cambridge, Massachusetts, second edi edition.

Hashimoto, M. (1979). Bonus Payments, on-the-Job Training, and Lifetime Employment in Japan. *Journal of Political Economy*, 87(5):1086–1104.

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

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.

Herkenhoff, K., Lise, J., Menzio, G., and Phillips, G. M. (2018). Production and Learning in Teams.

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.

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

Jarosch, G., Oberfield, E., and Rossi-hansberg, E. (2019). Learning from Coworkers.

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

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

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

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

Kremer, M. (1993). The O-Ring Theory of Economic Development. *The Quarterly Journal of Economics*, 108(3):551–575.

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

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

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

Loewenstein, M. A. and Spletzer, J. R. (1999). General and specific training evidence and implications. *Journal of Human Resources*, 34(4):731–733.

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

Lucas, R. E. and Moll, B. (2014). Knowledge Growth and the Allocation of Time. *Journal of Political Economy*, 122(1):1–51.

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

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

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.

Mincer, J. (1988). Job training, wage growth and labor turnover.

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

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

Oi, W. Y. and Idson, T. L. (1999). Firm size and wages. *Handbook of Labor Economics*, 3rd Part(2):2165–2214.

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

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

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

Rubinstein, Y. and Weiss, Y. (2006). Chapter 1 Post Schooling Wage Growth: Investment, Search and Learning.

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

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

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

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

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

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

A Appendix: Data Construction

1.1 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 waves focus on enterprises with 10 or more people, providing a sample of 78,000, 101,000 and 111,000 enterprises, respectively, from across all EU member states and Norway.

1.1.1 Firms' Investments in Skills

To build our measures of internal and external learning availability, 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:

- 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 for participating is training/learning." This is categorized as external learning.

1.2 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 a few important limitations. One of the key limitations, is the slight variation on questions asked across survey waves. This change partially compromises the comparability of our skill acquisition measures across waves, and thus the longitudinal nature of the data. Another issue is the response rate to the survey, which is as low as 44.0%. We address the first issue by showing that even though the mean level of the types of learning studies differ across waves, the pattern of change across workers' lifecycles is symmetric across these (see Figure C.4 for these results). To address the second 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.

1.2.1 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 a professional task 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; (2) the 2006 survey, which asks about having receiving professional development through coaching from superiors, and (2) the 2011/2012 and 2017/2018 surveys, when no related question was asked.
- "External Learning": is a binary variable that indicates whether an individual received external on-the-job training, or acquired necessary skills to complete a professional task 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 stem from 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 sources of professional knowledge/skills for the job stem

from 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 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: $Birth\ Year - Year\ of\ Graduation - 6$.
- 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.

1.2.2 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.

1.3 American 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.

Something to note with regards to the US data is that some of the learning questions were asked to workers who reported participating in a "work experience program", which is defined as a job with learning attributes. About 0.25 of the surveyed sample reported having been part of such a program. In Figure C.9 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 not. Please see below for details on the construction of the variables and these additional exercises.

1.3.1 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 type 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.[40]

- "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 employement-related training certificate are assumed to not have this type 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.[41]

- 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.

1.3.2 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.

[40] In Panel (a) of Figure C.9, we show that the results are robust to limiting the internal learning variable only to individuals reporting participating in a work-experience program.

[41] In Panels (b) and (c) of Figure C.9, we show that the results are robust to decomposing across the two learning components mentioned before.

1.4 Summary Statistics of Learning Data in Germany and the US

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

VARIABLES	(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

Table B.1: Proportion of Firms and workers hours participation in CVT

Country	% of Firms which workers participate in CVT courses			% of working hours spent training		
	CVT Courses	Internal CVT Courses	External CVT Courses	CVT Courses	Internal CVT Courses	External CVT Courses
Germany	0.592	0.436	0.532	0.007	0.004	0.003
France	0.721	0.329	0.666	0.009	0.003	0.006
United Kingdom	0.645	0.418	0.532	0.007	0.004	0.003
Italy	0.495	0.263	0.403	0.008	0.004	0.004
Spain	0.605	0.186	0.564	0.008	0.004	0.004
Poland	0.245	0.134	0.219	0.006	0.002	0.003
Romania	0.217	0.117	0.157	0.007	0.005	0.003
Belgium	0.682	0.457	0.605	0.013	0.008	0.006
Portugal	0.427	0.212	0.360	0.012	0.007	0.006
Czech Republic	0.720	0.391	0.611	0.008	0.004	0.004
Hungary	0.346	0.171	0.307	0.004	0.001	0.003
Sweden	0.787	0.600	0.724	0.009	0.005	0.004
Bulgaria	0.249	0.163	0.177	0.009	0.006	0.003
Denmark	0.748	0.485	0.640	0.011	0.005	0.006
Slovak Republic	0.587	0.352	0.531	0.009	0.004	0.005
Finland	0.716	0.352	0.671	0.007	0.003	0.004
Norway	0.844	0.676	0.694	0.010	0.006	0.004
Latvia	0.304	0.124	0.271	0.007	0.004	0.003
Estonia	0.589	0.326	0.552	0.008	0.004	0.004
Cyprus	0.492	0.185	0.447	0.008	0.005	0.004
Luxembourg	0.662	0.473	0.580	0.014	0.008	0.005
Malta	0.407	0.284	0.322	.	.	.
Total	0.498	0.275	0.433	0.008	0.004	0.004

Notes: This table shows the proportion of firms and working hours 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.

Table B.2: Proportion of Workers taking part in Other CVT activities.

Country	Conferences, workshops or lectures	Guided on the Job Training	Job Rotation	Learning and quality circles	Self Directed Learning
Germany	0.21	0.39	0.08	0.19	0.19
France	0.15	0.33	0.23	0.20	0.15
United Kingdom	0.18	0.40	0.15	0.21	0.23
Spain	0.22	0.42	0.21	0.20	0.18
Poland	0.17	0.39	0.11	0.09	0.13
Romania	0.14	0.38	0.22	0.15	0.16
Belgium	0.25	0.33	0.13	0.15	0.14
Portugal	0.14	0.33	0.18	0.27	0.15
Czech Republic	0.21	0.52	0.14	0.16	0.17
Hungary	0.13	0.34	0.12	0.16	0.11
Sweden	0.34	0.40	0.24	0.18	0.27
Bulgaria	0.13	0.49	0.16	0.24	0.14
Denmark	0.24	0.33	0.13	0.28	0.47
Slovak Republic	0.17	0.43	0.10	0.39	0.25
Finland	0.18	0.23	0.11	0.13	0.20
Norway	0.24	0.32	0.20	0.27	0.25
Latvia	0.26	0.36	0.10	0.19	0.12
Estonia	0.16	0.32	0.14	0.20	0.17
Cyprus	0.23	0.31	0.16	0.21	0.18
Luxembourg	0.27	0.35	0.14	0.25	0.23
Total	0.18	0.38	0.17	0.19	0.18

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.3: Share of firms Providing Internal & External Learning Activities of Different Kinds

		External CVT 0	External CVT 1
Internal CVT	0	0.41	0.21
	1	0.06	0.31

		External CVT 0	External CVT 1
Internal Other	0	0.37	0.18
	1	0.11	0.34

		External Other 0	External Other 1
Internal Other	0	0.48	0.07
	1	0.212	0.24

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.

Table B.4: Share of firms Providing Internal & External Learning Activities by Firm Size

Small firms, 1–9 Workers

		External CVT 0	External CVT 1
Internal CVT	0	0.61	0.19
	1	0.07	0.13

Medium firms, 10–99 Workers

		External CVT 0	External CVT 1
Internal CVT	0	0.46	0.24
	1	0.06	0.24

Large firms, 100+ Workers

		External CVT 0	External CVT 1
Internal CVT	0	0.19	0.20
	1	0.05	0.55

Notes: These tables show the proportion of firms of each size category which reported having persons employed participating in Internal and External CVT activities. The data Data from CVTS3, CVT4 and CVT5 surveys.

C Appendix: Robustness of Fact 2

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

Dep. Variables	Internal Learning			External Learning		
Germany						
Potential Yrs. Experience	-0.0104***	-0.00836***	-0.00626***	0.0105***	0.00779***	0.00462***
	(0.000672)	(0.000669)	(0.000856)	(0.000552)	(0.000541)	(0.000648)
Potential Yrs. Experience^2	0.000167***	0.000130***	0.000112***	-0.000247***	-0.000188***	-0.000139***
	(1.37e-05)	(1.37e-05)	(1.76e-05)	(1.14e-05)	(1.11e-05)	(1.33e-05)
Constant	0.439***	0.395***	0.397***	0.602***	0.666***	0.680***
	(0.00734)	(0.0121)	(0.0148)	(0.00592)	(0.00978)	(0.0119)
Observations	109,478	103,651	61,003	173,391	165,049	113,704
R-squared	0.006	0.085	0.083	0.005	0.137	0.133
Year FE		Y	Y		Y	Y
Firm size FE		Y	Y		Y	Y
Demographic Controls		Y	Y		Y	Y
Wage Controls			Y			Y
USA						
Potential Yrs. Experience	-0.00489***	-0.00820***	-0.00886***	0.0153***	0.0147***	0.0150***
	(0.00120)	(0.00109)	(0.00110)	(0.00141)	(0.00142)	(0.00145)
Potential Yrs. Experience^2	-4.13e-05*	7.09e-05***	8.58e-05***	-0.000277***	-0.000252***	-0.000261***
	(2.49e-05)	(2.27e-05)	(2.31e-05)	(3.05e-05)	(3.08e-05)	(3.15e-05)
Constant	0.352***	0.209***	0.204***	0.289***	0.242***	0.235***
	(0.0119)	(0.0140)	(0.0143)	(0.0129)	(0.0175)	(0.0179)
Observations	29,217	29,217	27,585	29,217	29,217	27,585
R-squared	0.040	0.162	0.165	0.013	0.025	0.026
Demographic Controls		Y	Y		Y	Y
Wage Controls			Y			Y

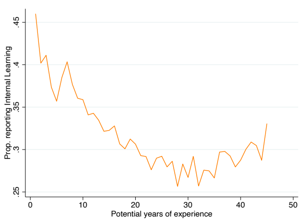
Internal learning, external learning and potential years of experience described in text for both countries. All regressions weighted using observation weights provided in the surveys. *Germany:* Year fixed effects correspond to year of survey fixed effects. Firm size is a categorical variable indicating 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. Demographic controls include worker type (laborer, private employee, government employee, self-employed, freelancer, or family caregiver), educational attainment level, and sex. Wage controls include the current hourly wage for the worker. *USA:* Demographic controls include worker type (private employee, government employee, self-employed, or working without pay), educational attainment level, race, census region, and sex. Wage controls include the current hourly wage for the worker. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

3.1 Decomposition by Different Subgroups

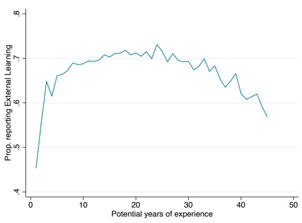
Figure C.1: Types of learning throughout workers' lifecycles by one-year experience bins

Germany

(a) Internal Learning

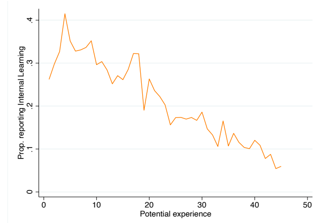


(b) External Learning



USA

(c) Internal Learning



(d) External Learning

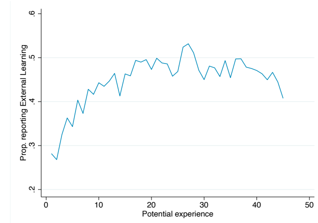
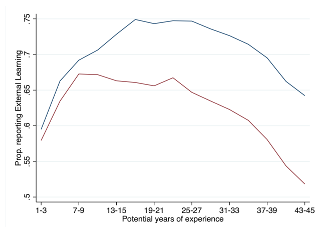


Figure C.2: Types of learning throughout workers' lifecycles by sex

Germany

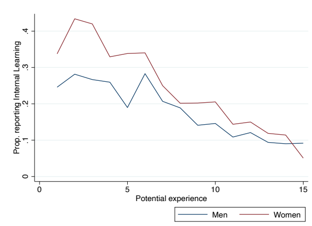
(a) Internal Learning

(b) External Learning



USA

(c) Internal Learning



(d) External Learning

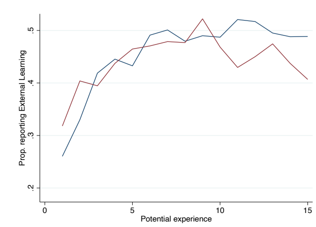
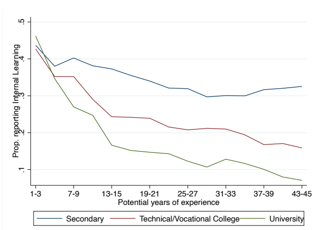


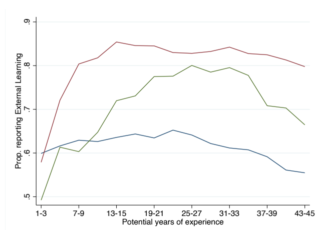
Figure C.3: Types of learning throughout workers' lifecycles by educational level

Germany

(a) Internal Learning

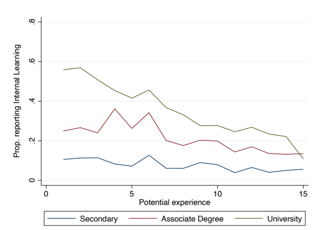


(b) External Learning



USA

(c) Internal Learning



(d) External Learning

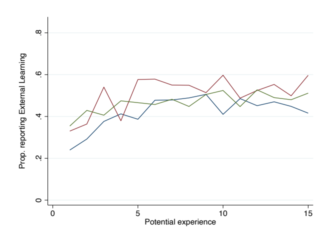
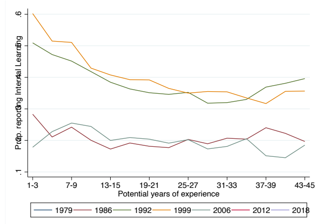
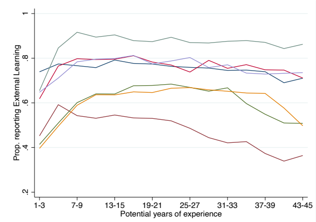


Figure C.4: Types of learning throughout workers' lifecycles by wave in Germany

(a) Internal Learning



(b) External Learning



3.2 Alternate Experience Variables

Figure C.5: Types of learning throughout workers' lifecycles by firm size in Germany

(a) Internal Learning

(b) External Learning

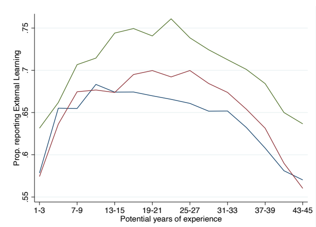
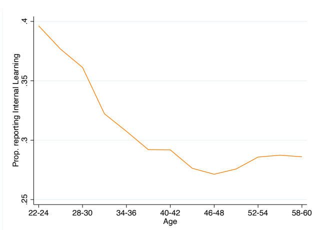


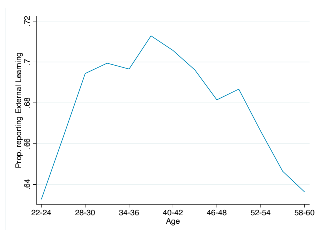
Figure C.6: Types of learning by age

Germany

(a) Internal Learning

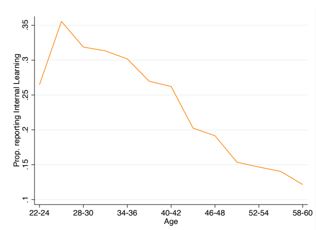


(b) External Learning

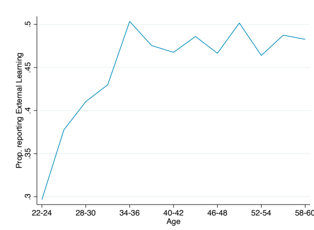


USA

(c) Internal Learning



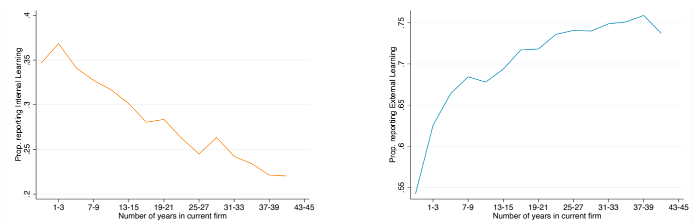
(d) External Learning



Notes: We exclude individuals who are 21 years of age or younger, or older than 60, since there are very few such observations, especially given our 45 years cut for potential experience.

Figure C.7: Types 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 with the same employer, 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).

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

Dep. Variables	Internal Learning				External Learning			
Germany								
Yrs. w/ Current Emp.	-0.00591***	-0.00591***	-0.00465***	-0.00340***	0.00891***	0.00590***	0.00477***	0.00753***
	(0.000560)	(0.000557)	(0.000721)	(0.000765)	(0.000473)	(0.000452)	(0.000551)	(0.000580)
Yrs. w/ Current Emp.^{2}	5.63e-05***	6.30e-05***	6.46e-05***	2.48e-05	-0.000145***	-0.000106***	-8.38e-05***	-8.02e-05***
	(1.67e-05)	(1.65e-05)	(2.18e-05)	(2.30e-05)	(1.39e-05)	(1.31e-05)	(1.60e-05)	(1.67e-05)
Constant	0.372***	0.324***	0.357***	1.070***	0.609***	0.702***	0.690***	0.975***
	(0.00354)	(0.0108)	(0.0132)	(0.0145)	(0.00304)	(0.00870)	(0.0110)	(0.00752)
Observations	102,761	100,565	59,683	59,683	165,275	160,629	111,280	111,280
R-squared	0.007	0.085	0.086	0.088	0.009	0.139	0.134	0.142
Year FE		Y	Y	Y		Y	Y	Y
Firm size FE		Y	Y	Y		Y	Y	Y
Demog. Controls		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. 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. Demographic controls include worker type (laborer, private employee, government employee, self-employed, freelancer, or family caregiver), educational attainment level, and sex. Wage controls include the current hourly wage for the worker. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

3.3 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. 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.

Table C.3: Correlations between different types of learning

Dep. Variables	Internal Learning		
Germany			
External Learning	-0.203***	-0.182***	-0.179***
	(0.00354)	(0.00369)	(0.00466)
Constant	0.438***	1.112***	1.221***
	(0.00301)	(0.119)	(0.0146)
Observations	109,478	103,651	61,003
R-squared	0.045	0.116	0.115
Year FE		Y	Y
Firm size FE		Y	Y
Age FE		Y	Y
Demographic Controls		Y	Y
Wage Controls			Y
USA			
External Learning	0.0949***	0.0941***	0.0987***
	(0.00773)	(0.00719)	(0.00732)
Constant	0.185***	-0.0543***	-0.0617***
	(0.00474)	(0.0133)	(0.0139)
Observations	29,217	7,991	7,401
R-squared	0.012	0.079	0.089
Age FE		Y	Y
Demographic Controls		Y	Y
Wage Controls			Y

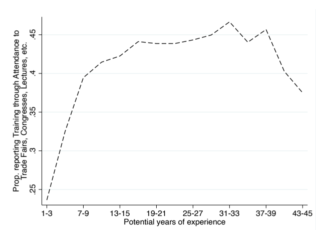
Internal learning and external learning described in text for both countries. All regressions weighted using observation weights provided in the surveys. *Germany:* Year fixed effects correspond to year of survey fixed effects. Firm size is a categorical variable indicating whether the firm where the 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. Demographic controls include worker type (laborer, private employee, government employee, self-employed, freelancer, or family caregiver), educational attainment level, and sex. Wage controls include the current hourly wage for the worker. *USA:* Demographic controls include worker type (private employee, government employee, self-employed, or working without pay), educational attainment level, race, census region, and sex. Wage controls include the current hourly wage for the worker. Robust standard errors in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

We find a significant negative correlation external and internal learning which is robust to controlling for several demographic and firm-level variables in Germany. These results are suggestive that internal and external sources of learning are more substitutable rather than complementary in the human capital accumulation process, consistent with our model. 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 types of learning are potentially more likely to coexist.

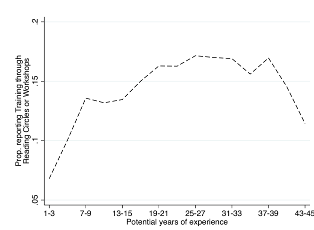
3.4 Decomposition by Different Categories of Learning

Figure C.8: Decomposing types of learning in external learning by dependence on outside experts throughout workers' lifecycles in Germany

(a) Attendance to Seminars or Courses



(b) Attendance to work shops or reading circles



(c) Reading of Trade Journals or Specialist Literature

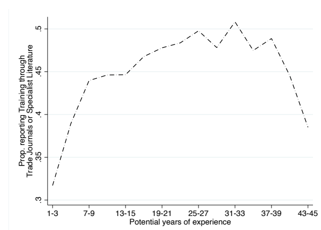
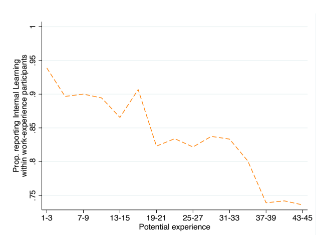
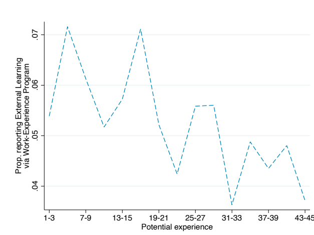


Figure C.9: Decomposing types of learning in the US

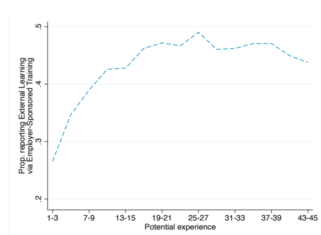
(a) Prevalence of internal learning within individuals reporting participating in a work-experience program



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



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

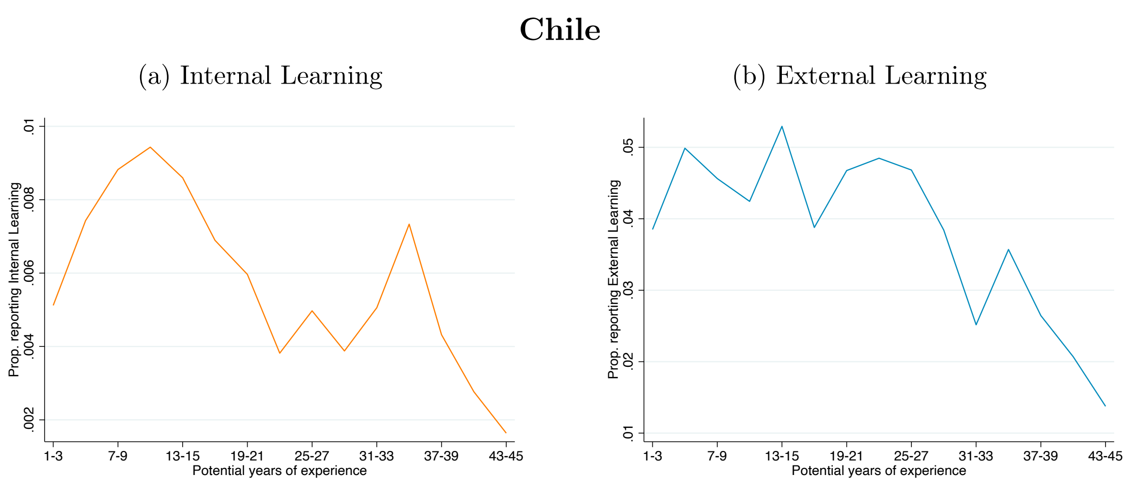


3.5 Evidence from Chile

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 3.5.2 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 3.5.1.

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



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.

3.5.1 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.

3.5.2 Summary Statistics and Additional Results

Table C.4: Summary Statistics in Chile

VARIABLES	(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 types 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
Firm size FE		Y	Y	Y		Y	Y	Y
Demographic Controls		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. Year fixed effects correspond to year of survey fixed effects. 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. Demographic controls include worker type (employer, own-account, private employee, government employee, live-in domestic worker, live-out domestic worker, unpaid family worker or military employee), industry (agriculture and fishing, mining, manufacturing, electricity and gas, construction, wholesale and retail trade, transportation, financial services, community and social services, and other) educational attainment level, and sex. 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

4.1 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.

4.2 Worker's lifetime problem

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

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

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 of any wage with human capital h_m (meaning step m in the human capital ladder) therefore has the following Bellman equation:

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

Where:

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

4.3 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:

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

- 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).[42]

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

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

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

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

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

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

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

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

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

4.4 Assumptions needed 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 $\epsilon < \frac{\delta}{1-\delta}$.*

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

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

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: $f_t(h_m) > 0 \; \forall h_m > h_{\underline{m}}$.*

4.5 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 4.6.1.

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 4.6.2.

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 4.6.3.

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 4.6.4.

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 4.6.5.

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 4.6.6.

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 4.6.7.

4.6 Proofs

4.6.1 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 4.2, 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.

4.6.2 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 type 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.

4.6.3 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.

4.6.4 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.

4.6.5 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.

4.6.6 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 4.2, 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.

4.6.7 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 4.2 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:

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

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 4.2 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:

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

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 $h_{\underline{m}-1}$: $f_s(h_m) > 0 \; \forall h_m \in [h_m, h_{\underline{m}-1}]$.

4.6.8 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 $h_{\underline{m}} > h_{m*}$.

4.6.9 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 4.2, 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[43]

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

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:

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

[43]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.

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 $m* = M-2$.

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

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

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 $\frac{k_l(h_{m+1})}{k_l(h_m)} > 1\ \forall h_m \in [h_1, h_{m*-2}]$.

4.6.10 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

5.1 Additional Information for Section 5.1

5.1.1 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.
  - American NHES Data: For the American 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.[44]

5.1.2 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
Firm size FE		Y		Y		Y
Demographic Controls		Y		Y		Y
USA						
Trainer	4	4	-1	4**	1	3
	(7.557)	(2.448)	(6.044)	(5.191)	(3.782)	(0.932)
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

Trainer (v. 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. 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. Demographic controls include worker type (laborer, private employee, government employee, self-employed, freelancer, or family caregiver), educational attainment level, and sex. *USA:* Demographic controls include worker type (private employee, government employee, self-employed, or working without pay), educational attainment level, race, census region, and sex. We do not include wage controls in these regressions since trainers and production workers have inherently different wage levels. Robust standard errors in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

[44] 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.

5.1.3 Additional Plots and Tables

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

(a) Germany (b) United States

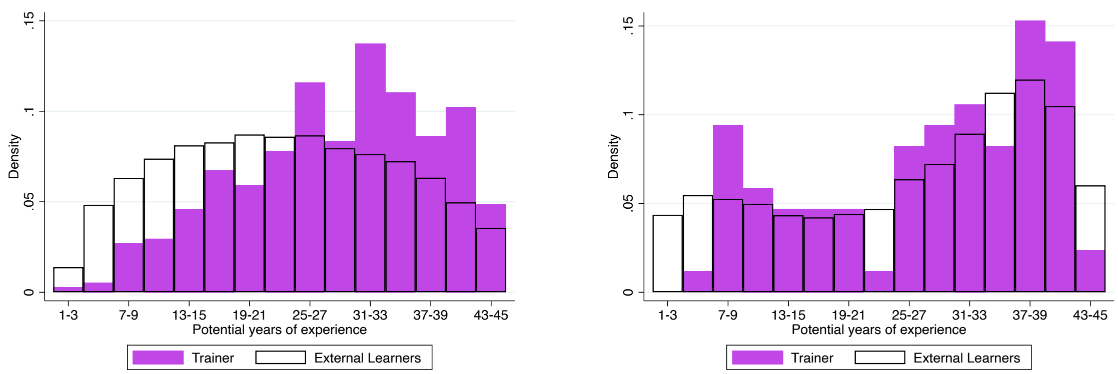


Figure E.2: Histograms of trainers and professional and technical production workers by potential experience and educational attainment

(a) Germany (b) United States

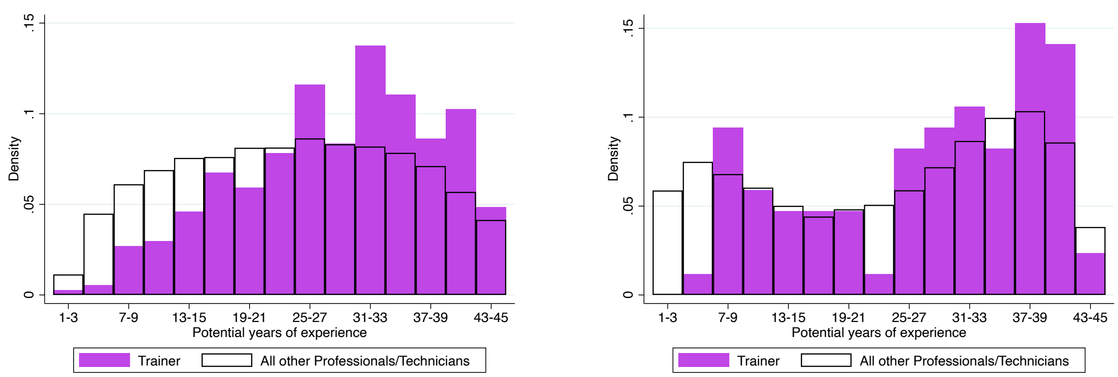


Table E.2: Quantile Regression of Potential Years of Experience for Trainers v. External Learners

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
Firm size FE		Y		Y		Y
Demographic Controls		Y		Y		Y
USA						
Trainer	2	3.333	-3	2**	0	3
	(7.324)	(5.450)	(4.691)	(1.000)	(2.936)	(6.315)
Constant	12***	15***	22***	27.50***	32***	36***
	(0.214)	(0.586)	(0.257)	(0.699)	(0.214)	(0.671)
Observations	13,613	13,613	13,613	13,613	13,613	13,613
Demographic Controls		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. 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. Demographic controls include worker type (laborer, private employee, government employee, self-employed, freelancer, or family caregiver), educational attainment level, and sex. *USA:* Demographic controls include worker type (private employee, government employee, self-employed, or working without pay), educational attainment level, race, census region, and sex. We do not include wage controls in these regressions since trainers and production workers have inherently different wage levels. 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
Firm size FE		Y		Y		Y
Demographic Controls		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

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. 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. Demographic controls include worker type (laborer, private employee, government employee, self-employed, freelancer, or family caregiver), educational attainment level, and sex. *USA:* Demographic controls include worker type (private employee, government employee, self-employed, or working without pay), educational attainment level, race, census region, and sex. We do not include wage controls in these regressions since trainers and production workers have inherently different wage level. Robust standard errors in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

5.2 Additional Information for Section 5.2

5.2.1 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. [45]

5.2.2 Additional Plots and Tables

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

Dep. Variables		Learning-by-Doing	
Germany			
Potential Yrs. Experience	0.0091***	0.0092***	0.0104***
	(0.0007)	(0.0007)	(0.0010)
Potential Yrs. Experience^2	-0.0001***	-0.0001***	-0.0001***
	(0.0000)	(0.0000)	(0.0000)
Constant	0.2670***	0.2296***	0.2526***
	(0.0077)	(0.0174)	(0.0223)
Observations	90,536	85,489	45,464
R-squared	0.0115	0.0784	0.0791
Year FE		Y	Y
Firm size FE		Y	Y
Demographic Controls		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. 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. Demographic controls include worker type (laborer, private employee, government employee, self-employed, freelancer, or family caregiver), educational attainment level, and sex. Wage controls include the current hourly wage for the worker. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

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

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

(a) By one-year exp. bins (b) By sex (c) By educational level (d) By wave

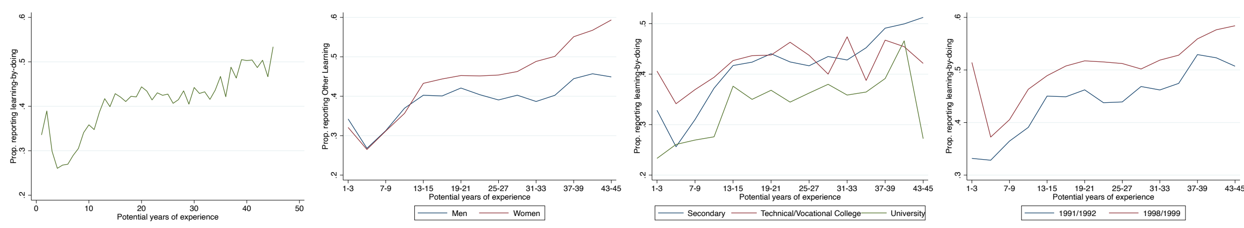
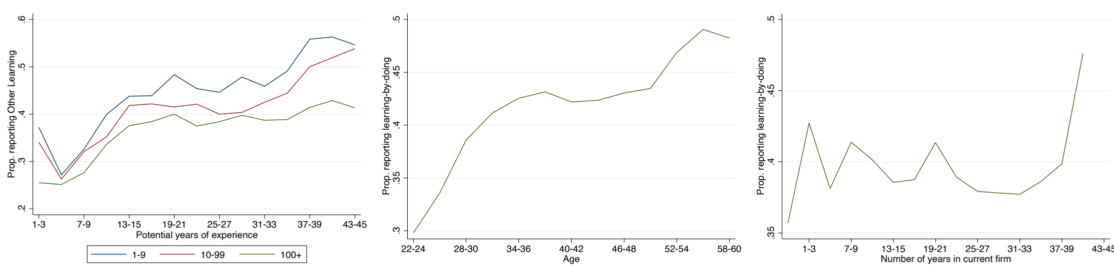


Figure E.4: Prevalence of Learning-by-Doing throughout workers' lifecycles in Germany

(a) By firm size (b) By age (c) By tenure in current firm



5.3 Additional Information for Section 5.3

5.3.1 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.

5.3.2 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.

5.3.3 Quantile 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
Firm size FE		Y		Y		Y
Demographic Controls		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

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. 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. Demographic controls include worker type (laborer, private employee, government employee, self-employed, freelancer, or family caregiver), educational attainment level, and sex. *USA:* Demographic controls include worker type (private employee, government employee, self-employed, or working without pay), educational attainment level, race, census region, and sex. We do not include wage controls in these regressions since trainers and production workers have inherently different wage levels. Robust standard errors in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

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

Dep. Variables	Task Complexity	
	50th percentile	
Germany		
External Learner	2***	1***
	(0.0990)	(0.0395)
Constant	1***	-1.28e-15
	(0.0940)	(0.06968)
Observations	84,315	46,992
Year FE		Y
Firm size FE		Y
Demographic Controls		Y

Trainer and external learner (v. internal learner) described in text for both countries. All regressions weighted using observation weights provided in the surveys. Year fixed effects correspond to year of survey fixed effects. 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. Demographic controls include worker type (laborer, private employee, government employee, self-employed, freelancer, or family caregiver), educational attainment level, and sex. Wage controls include the current hourly wage for the worker. Robust standard errors in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

5.4 Additional Information and Empirical Support for Section 5.4

5.4.1 Correlation between Internal Learning and Firm Size

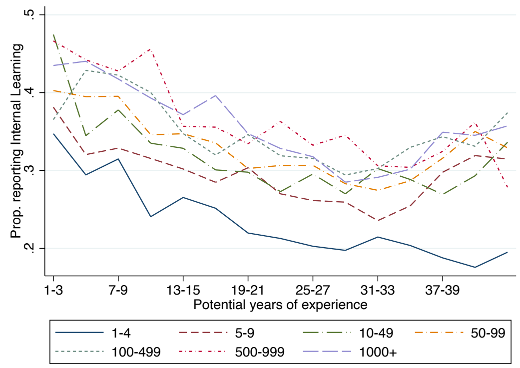
Table E.7: Correlation between internal learning and firm size

Dep. Variable	Internal Learning		
Germany			
5-9 Workers	0.0640***	0.00435	0.0144
	(0.00660)	(0.00685)	(0.00873)
10-49 Workers	0.0807***	0.0322***	0.0454***
	(0.00560)	(0.00633)	(0.00811)
50-99 Workers	0.0959***	0.0557***	0.0716***
	(0.00672)	(0.00737)	(0.00943)
100-499 Workers	0.116***	0.0730***	0.0937***
	(0.00598)	(0.00675)	(0.00862)
500-999 Workers	0.129***	0.0880***	0.120***
	(0.00845)	(0.00891)	(0.0116)
1000+ Workers	0.124***	0.0847***	0.118***
	(0.00667)	(0.00739)	(0.00958)
Constant	0.228***	0.395***	0.397***
	(0.00452)	(0.0121)	(0.0148)
Observations	103,959	103,651	61,003
R-squared	0.006	0.085	0.083
Year FE		Y	Y
Experience Controls		Y	Y
Demographic Controls		Y	Y
Wage Controls			Y

Internal Learning described in text. All regressions weighted using observation weights provided in the surveys. Firm size is a categorical variable, and the omitted category is a firm with fewer than 5 workers. Year fixed effects correspond to year of survey fixed effects. Demographic controls include worker type (laborer, private employee, government employee, self-employed, freelancer, or family caregiver), educational attainment level, and sex. Wage controls include the current hourly wage for the worker. Robust standard errors in parentheses.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

5.4.2 Additional Plots

Figure E.5: Prevalence of Internal Learning by Firm Size (disaggregated categories)



5.5 Additional Information and Empirical Support for Section 5.5

5.5.1 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.

5.5.2 Correlations between different types of learning and "work-related novelty"

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

Dep. Variables	External Learning			Internal learning		
Germany						
Freq. Task Novelty	0.177***	0.105***	0.101***	-0.0830***	-0.0642***	-0.0512***
	(0.00320)	(0.00334)	(0.00410)	(0.00373)	(0.00387)	(0.00492)
Freq. Procedure Improvements	0.124***	0.0785***	0.0739***	-0.0786***	-0.0468***	-0.0201***
	(0.00312)	(0.00344)	(0.00425)	(0.00384)	(0.00400)	(0.00505)
Constant	0.540***	0.586***	0.613***	0.369***	0.417***	0.409***
	(0.00221)	(0.00992)	(0.0120)	(0.00239)	(0.0122)	(0.0148)
Observations	170,009	161,771	112,882	106,833	101,077	60,771
R-squared	0.078	0.156	0.151	0.021	0.092	0.087
Year FE		Y	Y		Y	Y
Experience Controls		Y	Y		Y	Y
Demographic Controls		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. Year fixed effects correspond to year of survey fixed effects. Demographic controls include worker type (laborer, private employee, government employee, self-employed, freelancer, or family caregiver), educational attainment level, and sex. 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

6.1 Firms' Choices of Vacancies and Wages

In each period, each firm with productivity z chooses the optimal amount of vacancies $v(z)$ and the associated wage $w(z)$. 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).[46] This equation, combined with the min-mean wage ratio b (boundary condition), can solve the wage (similar as in Burdett and Mortensen (1998)):

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

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

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

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

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.